White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/129247/

Version: Accepted Version

Article:
Matthews, D., Biney, H. and 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). pp. 186-223. ISSN: 1547-5441

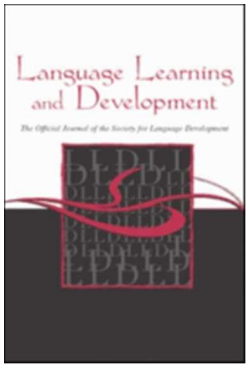
https://doi.org/10.1080/15475441.2018.1455584

© 2018 Taylor & Francis Group. This is an author produced version of a paper subsequently published in Language Learning and Development. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York



Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition and executive functions

Journal:	*Language Learning and Development*
Manuscript ID	HLLD-2017-0019.R2
Manuscript Type:	Original Articles
Date Submitted by the Author:	n/a
Complete List of Authors:	Matthews, Danielle; University of Sheffield, Department of Psychology Biney, Hannah; University of Sheffield Abbot-Smith, Kirsten; University of Kent
Keywords:	Pragmatics, executive function, Social communication, formal language, Theory of mind

For Peer Review Only

SCHOLARONE™
Manuscripts

Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition and executive functions.

Danielle Matthews[1], Hannah Biney[1], and Kirsten Abbot-Smith[2]

[1] Department of Psychology, University of Sheffield

[2]School of Psychology, University of Kent

Correspondence concerning this article should be addressed to Danielle Matthews, Department of Psychology, University of Sheffield, 1 Vicar Lane, Sheffield S1 2LT, UK

Contact: danielle.matthews@sheffield.ac.uk

Acknowledgements : We would like to thank Colin Bannard for helpful comments on the manuscript and Claudia von Bastian and Emma Blakey for recommending reading on executive functions.

For Peer Review Only

Abstract

Children vary in their ability to use language in social contexts and this has important consequences for wellbeing. We review studies that test whether individual differences in pragmatic skill are associated with formal language ability, mentalising and executive functions in both typical and atypical development. The strongest and most consistent associations found were between pragmatic and formal language. Additional associations with mentalising were observed, particularly with discourse contingency and irony understanding. Fewer studies considered executive function and evidence is mixed. To make progress, high-quality studies of specific pragmatic skills are needed to test mechanistic models of development. We propose 6 goals for future research: 1) developing an empirically-based taxonomy of pragmatic skills; 2) establishing which skills matter most for everyday functioning; 3) testing specific hypotheses about information processing; 4) augmenting measures of individual differences; 5) considering a broader set of psychological associates; 6) employing statistical tools that model the nested structure of pragmatics and cognition.

For Peer Review Only

Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition and executive functions.

People vary enormously in their pragmatic proficiency - the extent to which they can use language in context to engage with others. 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 understanding non-literal language including irony. While non-verbal communication (e.g., making eye contact, smiling and nodding during conversation) is often included in this family of skills, for the current review we define pragmatics as the linguistic component of social communication.

Individual differences in pragmatic ability have profound consequences for all arenas of social life. Within the typical population, pragmatic proficiency is positively correlated with peer popularity and the ability to engage in collaborative-based learning, and negatively correlated with social-emotional and behavioural difficulties, and mental health problems (e.g., Gottman, Gonso & Rasmussen, 1975; Helland, Lundervold, Heimann & Posserud, 2014; Kemple, Speranza & Hazen, 1992; Murphy, Faulkner & Farley, 2014). For people with Autism Spectrum Disorder (ASD), Social (Pragmatic) Communication Disorder (SCD), and Developmental Language Disorder (DLD) impairments in pragmatic skills have a long-term impact on relationship formation (e.g., Whitehouse, Watt, Line & Bishop, 2009) employability (e.g., Lewis, Woodyatt & Murdoch, 2008; Eaves & Ho, 2008) and behavioural, social and emotional problems (e.g., St Clair, Pickles, Durkin & Conti-Ramsden, 2011) respectively. Pragmatic language impairments are also strongly associated with other developmental disorders including Attention Deficit/ Hyperactivity Disorder (ADHD.

For Peer Review Only

Camarata & Gibson, 1999), Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD. Gilmour, Hill, Place & Skuse, 2004) as well as hearing loss (e.g., O'Reilly, Peterson, & Wellman, 2014). There is thus a clear need to explain why individual differences in pragmatic ability exist so that we can find the best means of supporting development and function.

One way forward is to explore the social and cognitive skills that theoretically underpin pragmatic skills. This research should ultimately contribute to a mechanistic model of pragmatic development by identifying potential bottlenecks in function and growth. For example, if performance on an inhibition task is shown to be associated with communicative perspective taking, then we can build a model of communication whereby inhibiting one's own perspective is a key sub-step. Of course, this approach has its limitations, most notably because it relies on finding measures with the requisite variance for correlational work to be carried out. It is possible that some steps in deploying a pragmatic skill are so readily achieved by all speakers as to show no variance (and we would not be able to distinguish this from a case where the step was not required at all). However, the assumption is that a correlational approach will be informative often enough to offer insight into the architecture of the developing language system. With a model of this system in hand, we will be better able to predict developmental outcomes, identify children at risk, create supportive interventions, and match these interventions to individuals who stand to benefit from them the most.

Of all the cognitive domains that could be important for pragmatic functioning, three have received the most attention: *formal language* proficiency (vocabulary and grammar), *mentalising* (Theory of Mind) and *executive functions* (including inhibition and working memory amongst others). The goal of this review paper was to establish what evidence there is of broad-brush associations between each of these domains and pragmatics and of more

For Peer Review Only

specific links. Testing for the latter is a challenge because it requires 1) a good information processing account of why a specific social or cognitive process would be implicated in a specific pragmatic function, 2) good measures for each domain, and 3) control measures to rule out the possibility that associations reflect more domain general ability. Even when these considerable challenges are met, theoretically anticipated collinearity between variables can make null findings hard to interpret. One further question to consider is therefore how often informative studies result from taking an individual differences approach.

For a paper to be included in this review, it needed to report both a measure of at least one of the three cognitive domains (formal language, mentalising, executive function) and the relation to a measure of pragmatic skill (see Appendix A for key words, search strategy and inclusion criteria). Complete findings of all papers identified in a systematic search are reported in table 1, Appendix B, with effect sizes reported in terms of Cohen's d, where we consider d=0.2 a 'small' effect size, 0.5 a 'medium' effect size and 0.8 a 'large' effect size (Cohen, 1992). While we wanted to include a broad range of literature, setting a wide scope for the search process means there will inevitably be gaps and missed papers. Furthermore, the review reports only published studies and does not address the possibility that the literature is biased in favour of reporting positive results over negative ones (something that is known to be a problem in the cognitive literature more broadly. E.g., De Bruin, Treccani & Della Sala, 2015). We therefore provide a representative review of research on this broad topic rather than an exhaustive summary.

Notwithstanding these caveats, having analysed over 50 papers, it would be fair to say that there is evidence that formal language, mentalising and executive functions are all broadly implicated in pragmatic function but evidence for specific associations with well-defined pragmatic skills is currently limited. There appear to be two main reasons for this. First, the information processing rationale for an expected association between a given

For Peer Review Only

pragmatic measure and a given social or cognitive measure is often underspecified. Second, methodological problems often limit the conclusions that can be drawn due to sample size, measurement quality, lack of variance, task specific demands or lack of control for likely covariates. Nonetheless, in some cases, there is good evidence of specific links and we highlight these in the main text of the review before sketching proposals for future research.

In what follows, we first introduce pragmatic measures commonly reported in the reviewed studies. We then present the main body of the review, which is organised according to the three domains: formal language, mentalising and executive function. At the start of the three sections, a summary of findings is given followed by more detailed results organised by type of pragmatic measure (global assessments, naturalistic conversation, referential communication, narrative, irony comprehension). A reader seeking the gist of the argument could follow the summaries before heading to the discussion section.

Measures of pragmatic skills

Measures of pragmatic skills vary in terms of focus, coverage and quality (see Adams, 2002; O'Neill, 2014; Russell & Grizzle, 2008 for reviews). In order to aid interpretation of the following correlational research, we briefly describe the measures most commonly reported in the individual differences literature.

Global measures. The most frequently used global tests of pragmatics are the subscales from the Comprehensive Assessment of Spoken Language, (CASL, Carrow-Woolfolk, 1999). For the *pragmatic judgment subscale* children respond to a scenario with the appropriate thing to say or do (e.g., "Suppose the telephone rings. You pick it up. What do you say?"). On the *nonliteral language* subscale, children explain the nonliteral meaning of statements (e.g., "When 5-year-old Jimmy started pulling his sister's hair, Dad said, "Jim, you're not a puppy anymore." What did he mean?"). Other standardised tests (the ELI battery (Saborit & Julian, 2005) and the Test of Pragmatic Language (TOPL; Phelps-Terasaki &

For Peer Review Only

Phelps-Gunn, 1992) also involve generating pragmatically appropriate responses to presented scenarios and are therefore meta-cognitive in nature.

One commonly used global clinical assessment is the Autism Diagnostic Observation Schedule (ADOS, Lord et al., 2000). This contains many measures of the pragmatic skills considered in this review (such as the ability to initiate or respond appropriately in verbal interaction) alongside measures of non-verbal social interaction (e.g. appropriate eye-gaze and use of gestures) and some measures that do not specifically relate to communication at all (e.g. imagination, understanding of romantic relationships).

Other measures of global pragmatic ability rely on parent/teacher reports, which have the advantage of gauging a broad range of abilities outside of a test situation – i.e., they measure pragmatic function in a range of real social contexts. Questionnaires in reviewed studies include the Children's Communication Checklist (in either its original version or in its current version as the CCC2, Bishop, 2003), the Language Use Inventory (LUI, O'Neill, 2009) and the Mindful Conversational Difficulties Scale (MCDS; Peterson, Garnett Kelly & Attwood, 2009). The CCC2 was designed to screen for potential communication disorders and sub-classify children from four years of age. It includes four subscales that can be considered pragmatic (E – initiation; F – stereotyped language; G – use of context and H – non-verbal communication). Many researchers now use it to study individual differences although studies differ in terms of which subscales are included because a pragmatic composite is no longer available (Norbury, Nash, Baird & Bishop, 2004). The LUI is a relatively new standardized parent report for assessing how younger children (18 months to 4 years) use language in everyday situations and also has the advantage of covering a broad range of skills. It is intended to capture functions of language that develop in tandem with children's growing social-cognitive understanding (O'Neill, 2007). The MDCS consists of eight questions, five of which tap language use requiring some level of perspective taking

For Peer Review Only

(e.g., 'Does the child frequently switch or omit topics in a conversation so that others become confused?'). However, it also includes two items that directly tap mentalising.

Conversation. Measures of conversational skill rely on a researcher engaging in semi-structured conversation with a child and then later analysing what is said with detailed coding schemes. They have the advantage of being direct measures of a skill that requires many pragmatic functions to come together in concert and have high ecological validity. However, this very advantage can lead to challenges with quantitative assessment as no two conversational turns are the same. Common measures include the production of conversational turns that are related to what one's interlocutor has just said and that provide relevant or new information. Given the importance of assessing conversation for clinical diagnosis (using DSM-V), future development of measures is a priority (Norbury, 2014).

Referential communication. Tests of referential communication tend to forgo ecological validity for precision of assessment. They generally tap a very specific skill, either producing or comprehending expressions that refer to objects by taking into account what an interlocutor can see or has previously experienced. They can be useful for isolating the use of specific social and cognitive processes.

Narrative. Tests of narrative generally involve children retelling a story (e.g., the standardized Renfrew Bus Story Test, Renfrew, 2010) or narrating a wordless picture book (with studies differing according to whether the book can be seen at the time of narration) although they sometimes also include a comprehension element (e.g., Multilingual Assessment Instrument for Narratives: LITMUS-MAIN, Gagarina et al., 2012). Like conversation, producing narratives is culturally universal but cognitively challenging –its complexity is what makes for an ecologically valid test of advanced, real world language use. The properties of narrative that are coded vary widely from one study to another and picking out specifically pragmatic aspects of narrative production is a known challenge (e.g.,

For Peer Review Only

Ketelaars, Jansonius, Cuperus, and Verhoeven, 2012). Commonly coded properties are management of common ground (including referring to characters so they are accessible to the listener), inclusion of (ir)relevant information, cohesion and coherence (although the latter is not always considered pragmatic). Any correlations observed should thus be interpreted with respect to a specific coding scheme. When doing so, it is worth considering what succeeding on a given measure would require of the speaker. For example, many aspects of narrative production may not necessitate perspective-taking (e.g., Arnold, Bennetto & Diehl, 2009).

Irony. Tests of irony generally require the comprehension of short stories in which one character, the ironist, directs a sarcastic comment to another. Test questions vary but generally tap understanding of: 1) whether the comment was meant literally (*meaning*), 2) whether the ironist believed the literal content of his comment and whether the ironist thought the other character would believe that he thought the literal content to be true (*belief*) and 3) why the ironist would have said what they did (*motivation*). Again, these tests generally require meta-linguistic insight.

Associations between pragmatic measures and formal language ability

For want of better expressions, researchers often distinguish *formal language skills* from *pragmatic skills*. This distinction is somewhat artificial for a few reasons. While it makes sense to talk about linguistic forms (speech sounds and grammatical structures) and their functions (semantic or pragmatic), measures of 'formal language' are at best tests of the semantic functions of language forms: they test whether children understand the meanings of words and sentences. To the extent that there is no clear division between semantics and pragmatics, it is inevitably difficult to construct tests that tap separable domains. When understanding words and sentences, we often engage in reasoning that would be considered pragmatic (indeed for social cognitive theories of language development such processes are

For Peer Review Only

fundamental, see Clark, 2005). Likewise, nominally pragmatic tests call on lexical and grammatical knowledge. Nonetheless, the distinction has been found to be useful in the sense that some children appear to have difficulties that are most noticeable when there is a need to use language in a social context (as is the case for children with ASD: Adams, Green, Gilchrist & Cox, 2002; Jones & Schwartz, 2009; Loukusa & Moilanen, 2009; Tager-Flusberg, Paul & Lord, 2005; Volden, 2002). Therefore, when seeking to explain individual differences in pragmatic skills, it makes sense to consider them potentially separable (at least partially) from formal language and then to test the extent to which in practice they are related.

With a few notable exceptions, most studies we review below find evidence of a medium to large correlation between pragmatic and formal language measures. This is consistently the case for studies with a global measure of pragmatic ability (direct or parent reported), naturalistic conversation or irony comprehension. Studies that focus on perspective taking in referential communication paradigms or on narrative production report more mixed results. However, it is not always clear what associations would be predicted for these measures. Overall, there is sufficient evidence of an association with formal language ability that, later in the review, it will be necessary to look for controls for formal language when exploring evidence for specific links between pragmatics, mentalising and executive functioning.

Global measures of pragmatics. Due to their wide-ranging nature, there are any number of reasons why global pragmatic measures may be associated with formal language. At best such measures allow us to establish whether these two domains are linked such that more specific measures can unpack why. There is consistent evidence of a medium to large association with formal language for typically developing children (Bernard & Deleau, 2007; De Rosnay, Fink, Begeer, Slaughter, & Peterson, 2014), children with ASD (Volden,

For Peer Review Only

Coolican, Garon, White & Bryson, 2009; Akbar, Loomis, & Paul, 2013; Whyte & Nelson, 2015), children with Developmental Language Disorder (DLD, Andrés-Roqueta, Adrian, Clemente, & Katsos, 2013), deaf children (Rinaldi, Baruffaldi, Burdo, & Caselli, 2013) and a group of children covering the full range of pragmatic abilities on the normed LUI measure (Pesco & O'Neill, 2012).

Naturalistic conversation. Engaging in fluid conversation calls on formal language skills that allow the rapid processing of incoming speech and planning of speaker turns. Two studies of children with ASD found large correlations between conversational ability and formal language (Capps, Kehres & Sigman, 1998; Hale & Tager-Flusberg, 2005), providing consistent evidence of an association.

Referential communication. Studies focused on referential communication report more mixed results. This might be because, for some studies, measuring individual differences was not the focus of the study and/or sample sizes were small. Furthermore, referential communication tasks are often designed to have limited formal language demands, with simple instructions that are within the grasp of young participants. For example, if requested to "pick up the big cup", a 3-year-old might not struggle to understand the instruction (but may struggle to take into account their partner's visual perspective when selecting a cup). Thus, the language demands of the test may not be the most important bottleneck in successful performance. Nonetheless, we might still predict an association between the domains if we assume that more experience of linguistic interactions (and more facility with learning from them) would result in both better vocabulary comprehension and better ability to accommodate an interlocutor's perspective (e.g., Matthews, Butcher, Lieven & Tomasello, 2012).

Looking at typical development, Nilsen and Graham (2009, 2012) found mixed evidence of an association between vocabulary and egocentric errors of different kinds

For Peer Review Only

whereas Gillis and Nilsen (2014) found a medium-sized correlation between formal language and children's ratings of the helpfulness of ambiguous messages. Research on atypical populations also finds for children with ASD both mixed results (Fukumura, 2016), on the one hand, as well as, on the other hand, evidence of an association (Dahlgren & Dahlgren Sandberg, 2008; Nadig, Vivante & Ozonoff, 2009) and DLD (Davies, Andrés-Roqueta, & Norbury, 2016).

Narrative. The production of narrative necessarily calls on vocabulary, grammar and knowledge of language structure at the supra-sentential level. Separating these skills from more pragmatic aspects of narrative production (e.g., managing information flow for a listener) is a challenge and findings are mixed. Fernández (2013) observed a medium sized correlation between receptive vocabulary and overall coherence of narratives but not other measures. Blom and Boerma (2016) found medium sized correlations between formal language measures and concurrent measures of narrative comprehension and production (macrostructure) for children with DLD. For their typically developing group, however, very few such associations were observed. De Marchena and Eigsti (2016) found no association between receptive vocabulary and narrative adaptation to common ground in children with ASD (although the authors note that participants were selected to be in the typical range for vocabulary). Losh and Capps (2003) found no correlations between verbal IQ and narrative measures including evaluation in very high functioning children with ASD. Norbury, Gemmel and Paul (2014) found that while formal language ability (assessed by the CELF, Semel, Wiig & Secord, 1995) did not correlate with pragmatic error for a DLD group, it was negatively correlated with pragmatic errors for their ASD group. However, it was also negatively correlated with relevant propositions (-1.35), suggesting that more verbally able children with ASD may be more verbose but are then prone to more irrelevance.

For Peer Review Only

Irony. Since tests of irony comprehension usually require understanding short stories, they also necessarily call on receptive vocabulary and grammar. Performance on corresponding tests could be related for this reason alone. Understanding the ironic element of these stories specifically could also plausibly be linked to language ability through language experience. That is, children with more experience of language will have larger vocabularies and more incidental practice with this aspect of non-literal language use (see, e.g., O'Reilly, Peterson & Wellman, 2014, for evidence of delay in deaf children of hearing parents). The relatively large literature on this topic reports consistent evidence of medium-large associations with formal language ability for typically developing children (Filippova & Astington, 2008; Angeleri & Airenti, 2014; Massaro, Valle & Marchetti, 2014; Mewhort-Buist & Nilsen, 2013; Nilsen, Glenwright & Huyder, 2011). Studies of children with ASD (Huang, Oi & Taguchi, 2015; Nicholson, Whalen & Pexman, 2013) and ADHD (Adachi et al., 2004; Caillies, Bertot, Motte, Raynaud & Abely, 2014) did not find relationships with the particular formal language measure used although parent report measures may not have been sensitive enough. All things considered, there tends to be an association with receptive vocabulary, certainly in typical development.

Summary. Many tests of pragmatics have clear formal language demands and this is reflected in a consistent pattern of medium-large associations for all measures except tests that deliberately seek to minimise demands on vocabulary and grammar. This would suggest that formal language and pragmatic language are not entirely separable, a conclusion consistent with studies of atypical development that find children with DLD to be impaired relative to typical controls on many measures of pragmatic language (e.g. Norbury, Nash, Baird & Bishop, 2004; Norbury, Gemmel & Paul, 2014) even if children with ASD show greater pragmatic impairments (e.g., Colozzo, Morris & Mirenda, 2015; Norbury & Bishop, 2003; see also Miller et al., 2015). It is therefore clear that future studies should control for

For Peer Review Only

formal language ability if they are to look for specific links between pragmatics and other domains. This said, from the current research it is not possible to rule out Matthew effects (Merton, 1968; Stanovich, 1986) whereby children with high scores on pragmatic and formal language tests would have had high scores on many other tests, including non-verbal IQ. An important step for the field will be to do the necessary factor analytic work to gauge the extent to which pragmatic and formal language skills reflect a single underlying factor, and the extent to which they are separable from each other and from other social and cognitive dimensions, or indeed reflect a domain general construct (e.g., 'g', Gustafsson, 1984) at certain points in development.

Associations between pragmatic measures and mentalising ability

The term mentalising is used here to encompass children's understanding of themselves and others as mental beings who are guided by their attentional states, beliefs, desires, intentions, emotions, interests and perspectives. Mentalising (or Theory of Mind, ToM) has traditionally been seen as the most important cognitive underpinning of pragmatics (e.g., Baron-Cohen, 1988; Geurts, Broeders & Nieuwland, 2010; Perner, Frith, Leslie & Leekam, 1989). Under O'Neill's (2012) pragmatic taxonomy, one of the three main sets of pragmatic skills is labelled 'mindful pragmatics', grouping together instances of language use that require taking the perspective of a specific interlocutor (as opposed to having a model of interlocutors in general, or a routine social situation). On the face of it, having difficulties recognising such perspectives should cause communication problems. However, mentalising may not be a unitary construct, particularly in terms of the distinction between understanding the emotions of others versus understanding their knowledge states. Therefore, part of the challenge of research is to identify which types of mentalising are necessary in specific communicative situations. To date, the vast majority of research has focused on false belief understanding. Yet, for many pragmatic tasks it is not always clear why this would be called

For Peer Review Only

upon and so an important question is whether it can be taken as a proxy for mentalising ability more generally or not. Another aspect of the challenge is to unpack the evolving association between mentalising and pragmatic development given that each is likely to influence the other. Indeed, there is a large body of research highlighting the importance of language exposure for typical development of theory of mind (e.g., Astington & Wild, 2006; de Rosnay & Hughes, 2005; Dunn, Brown, Slomkowski, Tesla & Youngblade, 1991; Hughes, 2011; Hughes, et al., 2005; Milligan, Astington & Dack, 2007).

Overall, where studies have managed to run tests of mentalising that yield substantial variance, this tends to be associated with individual differences in pragmatic function, especially for global pragmatic measures, conversational ability and irony comprehension. For many of these studies, few control variables are taken into account and so the specificity of the association is unclear. For others, when language is controlled for, the association remains providing strong evidence for a specific link between mentalising (particularly 2nd order ToM – knowing that person A thinks that person B thinks something) and both conversation skills and irony comprehension, although notably no studies have included a control measure of non-verbal IQ. Occasionally, once controls are accounted for, no association between mentalising and pragmatic function remains (e.g., Pellicano, 2013; Whyte & Nelson, 2015). In these cases, it is not always easy to know whether this is because mentalising is not the limiting factor for the type of pragmatic ability assessed (something that is hard to establish when global measures are used) or whether all the variance in mentalising was already lost in the soup of control variables that are known to co-vary with it. It might be possible to solve this problem but it is also possible that we are trying to hone in on dimensions of cognitive function that cannot exist in isolation from each other – raising questions about the limits of individual differences research. Finally, for studies of referential communication and narrative production, the evidence of an association with

For Peer Review Only

mentalising, often assessed as 1st order false belief, is mixed. This could sometimes be due to lack of variance in measures and/or a mismatch between the mentalising skill tested and the mentalising demands of the pragmatic task. In sum, while there is enough evidence of a broad-brush link between the two domains, we now need to pin down specifically how this link is substantiated for individual pragmatic functions.

Global measures of pragmatics. Studies of typical development that have explored how mentalising relates to a broad measure of pragmatic ability consistently find a medium to large association. Bernard and Deleau (2007) found medium to large correlations between a composite false belief measure and a composite measure of communicative perspective taking at all three time points (3;8, 4;2, 4;8) in their longitudinal study. Likewise, De Rosnay, Fink, Begeer, Slaughter, and Peterson (2014) found a large correlation between mindful conversational competence and ToM, a relationship that remained after covariates including age, receptive vocabulary, emotion understanding and shyness were controlled for. This suggests that measures taken from traditional assessments of false belief understanding can predict real-world use of mentalising for conversation. However, one quarter of the items on de Rosnay et al.'s (2014) measure of conversational competence were measures of mentalising so the specificity of association is not clear.

Looking at atypical development, one study that stands out from the point of view of tackling the methodological challenges of indvidual differences research is reported by Losh, Martin, Klusek, Hogan-Brown, and Sideris (2012). ToM was assessed using one of two batteries (made up of tasks from published studies) depending on the child's developmental level. Both batteries included the same test of perspective taking and this allowed the development of a single theory of mind scale onto which all children could be mapped. This is a good solution to the problem of needing measures with sufficient variance. Large correlations were observed between this ToM score and performance on the pragmatic

For Peer Review Only

judgment subtest of the CASL for children with ASD, Fragile X Syndrome, Down Syndrome and neuro-typical children. Two further studies provide support for this association between pragmatic ability and concurrent mentalising (Whyte & Nelson, 2015, for TD children and children with ASD; Andrés□-Roqueta et al., 2013, for children with DLD).

One problem with research on the association between mentalising and pragmatic skills is that other co-variates are not always controlled for. When they are, the picture regarding associations becomes mixed (e.g., Whyte & Nelson, 2015). One study of high functioning children with ASD found no role for mentalising in explaining variance on the ADOS-G once a range of other factors had been controlled for. Pellicano (2013) found a large negative association between ToM at time 1 and communication problems at time 2 (i.e, better ToM, fewer problems). However, when age, verbal ability, and non-verbal ability were partialled out, this association did not hold (only early differences in executive function scores remained related to later ADOS-G scores – see EF section below). It is difficult to interpret this absence of correlation (since absence of evidence is not evidence of absence), an issue we return to in the discussion.

Overall, there is some evidence of an association between mentalising and global assessments of pragmatic function but the specificity of this association is currently unclear.

Conversational skills. While extent of engagement in a conversation can vary, one needs, at a minimum, to attend to an interlocutor's conversational turn, understand it and respond, ideally by taking into account common ground, the question under discussion, and interlocutor interests. In a study of peer interaction in typically developing 4-year-olds, Slomskowski and Dunn (1996) found a large correlation between connectedness of peer conversation and performance on 1st order false belief tasks. An initial study of children with ASD (Capps et al., 1998) found that, once formal language had been controlled for, there was no evidence of a specific association between mentalising ability and provision of new and

For Peer Review Only

relevant information. However, when Hale and Tager-Flusberg (2005) followed up on this by using a more extensive battery of ToM tests (to avoid the possibility that lack of variance was responsible for null findings), they observed that once age, IQ and vocabulary score were controlled for, ToM explained additional variance (8%) in the amount of speech that was contingent on what their partner had said. This study provided good evidence for a role of mentalising. However, it is worth noting that the ToM scale included items (e.g., about lies and jokes) that directly tap nominally pragmatic abilities and it is not clear which aspects of the ToM tasks related to conversational proficiency. Unpacking this association is an important challenge for future research.

Referential communication. Studies of referential communication require children to take common ground into account – i.e., to assess whether a partner can see something that is being referred to or whether they have prior experience with it. There is tentative evidence that referential communication ability is related to mentalising. For example, Resches and Perez-Pereira (2007) found evidence of large correlations between mentalising (knowledge/ignorance and false belief combined) and some (but not all) measures of performance on a highly motivating task that required children to describe to a peer the location of hidden treasure (a location that the director child had previously experienced but the hunter child had not). Similar mixed results are reported by Maridaki-Kassotaki and Antonopoulou (2011) and Dahlgren and Dahlgren Sandberg (2008). Overall, studies on referential communication and mentalising have not resulted in a consistent picture, sometimes because of a lack of variance on key measures. Furthermore, studies do not always control for formal language or assess the specific mentalising skill that is presumed to be required for the referential communication task. That is, if the task requires adjusting language production according to what an interlocutor can see, we would expect success to

For Peer Review Only

correlate with a measure of level 1 visual perspective taking but not necessarily with tests of false belief understanding.

Narrative. Producing a narrative potentially calls on mentalising in that one could model the listener's (or at least a generic listener's) epistemic state in order to manage information flow for them (e.g., introducing characters appropriately, maintaining accessibility of story elements as the narrative progresses, building up tension and resolution). However, evidence regarding the link between mentalising and narrative production is mixed. Fernández (2013) found no correlation between any of four measures of narrative quality and 1st order ToM scores (perhaps due to a ceiling effect on the latter) but did observe a medium correlation with 2nd order ToM and narrative coherence. Ketelaars, Jansonius, Cuperus, and Verhoeven (2012) found no evidence of an association between 1st order ToM and narrative organization (amount of relevant content) or cohesion (use of anaphora and deixis as cohesive devices) for children with SCD or typically developing children. Capps, Losh and Thurber (2000) found, for a group of children with ASD (but not a group with developmental delay), a correlation between 1st order false belief understanding and a range of narrative properties including evaluatuative statements and mention of mental state terms, although only the latter remained once language ability was controlled for. Losh and Capps (2003) found no correlations between advanced ToM (strange stories) and narrative measures including evaluation in high functioning children with ASD, although associations with emotion understanding were observed. Kuijper, Hartman and Hendriks (2015) found that 1st and 2nd order ToM were predictors of appropriate referent reintroduction (using a noun instead of a pronoun), with 2nd order ToM remaining predictive in a multivariate model. In sum, it is hard to pinpoint which measures of narrative we should consider 'pragmatic' (see related discussion in O'Neill, 2014 on anaphora) and, given the range of different options and mixed evidence, it is hard to come to any solid conclusions.

For Peer Review Only

However, given the importance of narrative production as a human activity (Bruner, 1990), future work could aim at a consensus on measures of interest and then establish the extent to which these depend on formal language skills or social cognition.

Irony. The understanding of irony should theoretically be related to higher order theory of mind in that it requires understanding that a speaker *thinks* their addressee will *know* they are not being literal, and understanding the speaker's attitude in producing the statement. The literature tends to confirm that this is the case. In a particularly well-controlled study, Filippova and Astington (2008) assessed 2nd order false belief, receptive vocabulary, forwards and backwards digit span and the ability to detect emotions from prosodic cues. There was a large correlation between 2nd order ToM and irony comprehension. Regression models showed that even once age, memory, attunement to prosody and receptive vocabulary were controlled for, ToM was a significant predictor of irony comprehension (explaining an additional 4% of variance). This study was notable in clearly spelling out why a specific type of mentalising should be associated with a specific pragmatic skill and then testing for the association while carefully controlling for relevant covariates.

Several more studies support this association. Nilsen et al. (2011) pinpointed a medium-large relationship between 2nd order false belief and a measure of children's sensitivity to the effect of a listener's knowledge state on their understanding of a speaker's ironic intentions. A further five studies of irony comprehension in typically developing children also found a relationship with ToM although they did not control for formal language ability (Massaro, et al., 2014; Banasik, 2013(mixed results); Mewhort-Buist & Nilsen; 2013 (mixed results); Angeleri & Airenti, 2014; Nicholson, Whalen & Pexman 2013). Massaro, Valle and Marchetti (2014) is the only study of typically developing children that did not find evidence of a relationship between ToM and irony interpretation. However,

For Peer Review Only

this study used few items for both measures. Finally, in a study of children with ADHD, Caillies et al. (2014) found a large correlation between 2nd order false belief understanding and irony comprehension. Taken together, there is good evidence of an association between 2nd order mental state reasoning and irony comprehension. An important question for future research is the extent to which this relation is observed because many 2nd order ToM tests tap essentially pragmatic skills.

Associations between pragmatic measures and Executive Functions

Executive Functions (EFs) are a set of higher order skills that allow individuals to think and behave in a flexible, controlled and goal-directed way (Diamond, 2013). Working memory, inhibition and cognitive flexibility are the EFs most commonly reported in the literature on pragmatic development, although organisation, planning, self-control and generativity are also considered. It is intuitively plausible that any of these would be called upon for pragmatic function. Language use regularly requires us to hold in mind and update linguistic and contextual information (working memory), suppress one's own perspective (inhibition), flexibly respond as a discourse unfolds (cognitive flexibility), think ahead to what will be communicated (planning), order information to form cohesive and coherent narratives (organisation) and develop new topics of conversation (generativity). Limits on any of these abilities are likely to shape the course of typical pragmatic development and potentially explain departure from the normal trajectory for many children with developmental disorders.

Deficits in executive functions have been implicated in a number of neuro-developmental disorders (e.g., Kingdon et al., 2016; Lai et al., 2017; Landry & Al-Taie, 2016; Sjöwal et al., 2013). While many assume that executive problems cause pragmatic problems, some have flipped the question and asked whether language impairments can explain deficits in executive function (e.g., Akbar, Loomis, & Paul, 2013), for example,

For Peer Review Only

because inner speech is needed to regulate non-routine behaviours (Joseph, McGrath, & Tager-Flusberg, 2005; Ren, Wang & Jarrold, 2016). Thus any link between the two domains is at least theoretically bi-directional.

Perhaps the most important point to bear in mind while interpreting reported associations below is that there is currently little consensus in the field of executive function about what the different types of EF are, how they relate to each other and how reliable measures of them are. Wiebe et al. (2011) suggest that early on in development EF is a unitary, domain general construct. For adults, Miyake et al.'s (2008; 2012) factor analytic work found separable but correlated factors. We are therefore in the tricky business of looking for associations between two domains (EF and pragmatics) for which the dimensional structure is unclear and likely to change over developmental time. The adult psycho-linguistic literature further suggests that evidence of correlations will be patchy because the types of cognitive support called on, for example, in communicative perspective-taking, are highly task specific (Ryksin et al. 2015), anticipated effect sizes are small (and sample sizes are often not large enough to detect them), and measurement reliability for both domains is suboptimal (Brown-Schmidt & Fraundorf, 2015; Hedge, Powell & Sumner, 2017).

With these substantial caveats in mind, overall, the research reviewed below does suggest that executive functions are globally associated with pragmatic functions, as we would expect. Beyond this, however, a consistent picture is yet to emerge. There is evidence of inhibitory control affecting performance on some tests of communicative perspective taking, and of a relationship between working memory and the ability to respond contingently in conversation. However, the latter needs to be followed up since there is some debate about the order of dependence between working memory and language development (e.g., Jones et al., 2014). There has been less work on other EFs but there is some evidence

For Peer Review Only

that cognitive flexibility may be an important bottleneck early on, particularly in the process of repair. Less well-studied abilities that are arguably executive in nature, like generativity, also emerge as potentially important for language use in social contexts. Certainly, research on atypical development suggests that poor executive function is often associated with pragmatic difficulties. Precisely why this is the case is not yet clear.

Global measures of pragmatic development. When global measures of executive function and pragmatics are used, there is evidence of a broad-brush association between the two domains. Thus, in a study of high functioning children with ASD, Pellicano (2013) observed a large negative association between an aggregate EF measure at time 1 (assessing planning, cognitive flexibility and inhibition) and social communication problems (ADOS-G) at time 2. This association remained when age, verbal ability, and non-verbal ability at time 1 were partialled out, leading Pellicano to argue that EFs are likely to place direct constraints on the development of social communicative behaviour (i.e., they do not only exert an influence via mentalising). While the composite task used here had the advantage of factoring out task specific variance and reducing risk of type I error (because it avoids running many correlations, one for each EF measure), the next step is to break down this broad-brush association to look at specific executive functions. There are relatively few studies that have looked for links between specific EFs and global pragmatic ability but some have investigated inhibition, working memory and generativity.

Regarding inhibition, Rints, McAuley, and Nilsen (2015) found that children whose parents rated them as more inattentive or hyperactive-impulsive also rated them as having poorer pragmatic skills on the LUI. They also found a large correlation between inhibition (movement errors on a statue task) and pragmatic scores on the Pragmatic Judgment subtest of the CASL. In contrast, Akbar, Loomis, and Paul (2013) found no correlation between this subtest of the CASL and inhibition (a colour-word interference test) in a study of children

For Peer Review Only

with ASD. Finally, in a study of children with DLD, Andrés-Roqueta, Adrian, Clemente, and Katsos (2013) found that inhibition was associated with pragmatic proficiency. Thus, evidence is currently mixed.

Regarding working memory, links to global pragmatic function in children with ASD have been observed, although the direction of causation is assumed to be from language to executive function. Akbar, Loomis, and Paul (2013) found a large correlation between performance on the pragmatic judgement subtest of the CASL and both working memory and organisation (but not with cognitive flexibility or inhibition).

Regarding generativity, there is evidence that the ability to fluently generate novel ideas (i.e., thinking of all the possible uses of a pencil or all the possible interpretations of an abstract line drawing) is associated with general pragmatic function. Bishop and Norbury (2005) derived a composite pragmatic language score (from the CCC pragmatic composite, the SCQ communication scale, and ADOS–G) for children with a range of related development disorders (DLD, SCD, ASD) and found a large correlation with generativity. An association held when structural language and age were controlled for. The authors proposed that generation of relevant ideas is needed to consider multiple possible meanings of an utterance and to avoid restricting conversation to specific topics or have it depart on unexpected tangents. This study was also interesting in that it took a dimensional approach and included children with different diagnostic labels in the same analyses.

Conversation skills. Although intuitively conversation should call on executive functions (e.g., updating the record of conversation or, as just noted, inhibiting irrelevant or tangential turns) evidence regarding a relation is still sparse. One comprehensive study of typical conversational development points to a role for both inhibition and working memory. Blain-Brière, Bouchard, and Bigras (2014) coded semi-structured conversations and found that *talkativeness* was negatively correlated with inhibition and *responsiveness* was positively

For Peer Review Only

associated with working memory. In general, self-control, flexibility and planning showed very little association with pragmatic function. This confirms the general picture in this section whereby inhibition and working memory appear to be the most consistently important factors for pragmatic development.

Referential communication. Succeeding on referential communication tasks in principle requires inhibiting one's own perspective to consider the interlocutor's and potentially switching back and forth between the two. In a particularly meticulous test of this hypothesis, Nilsen and Graham (2009) reported two studies assessing comprehension and production of referring expressions alongside tasks tapping working memory, inhibition and cognitive flexibility. In a first study, executive function was not associated with performance on a production task. However, there was a medium-sized correlation between egocentric errors on the comprehension task and inhibitory control (which withstood controlling for age and verbal ability). Of particular interest was the fact that children's performance on control trials (where the speaker could see all relevant objects and so inhibiting conflicting information about one's own perspective to avoid egocentric responses was not required) was not correlated with any executive function measure. This provides a clear demonstration that inhibition is involved in preventing the selection of referents that, while plausible from the child's point of view, are not from the speaker's. A second study found that better conflict inhibition specifically (i.e., inhibition where a specific response is suppressed while an alternative response is generated as opposed to delay inhibition where a response is suppressed for a given amount of time) was related to reduced egocentric looking during communicative perspective taking. Taken together, these studies provide a clear demonstration of the potential importance of inhibition in explaining why some children find communicative perspective taking easier than others.

One challenge with individual differences research is finding measures with sufficient

For Peer Review Only

variance to test hypotheses. Nilsen and Graham, (2009) noted that their measure of cognitive flexibility was negatively skewed, which may have explained null findings. However, in a follow-up, Gillis and Nilsen (2014) found a large negative correlation between cognitive flexibility (sorting objects according to different dimensions) and 5-year-olds' rating of how helpful an ambiguous description was (a relationship that remained after controlling for age, receptive language and baseline ratings of unambiguous cues). No significant correlations were found for older children, illustrating how EF bottlenecks in language processing are likely to change over developmental time

Looking at production of unambiguous referring expressions and children's ability to repair misunderstandings, Bacso and Nilsen (2017) found that cognitive flexibility and working memory (but not inhibition) were correlated with the quality of children's descriptions even when expressive vocabulary was controlled. The authors suggest working memory may be called upon to identify which features best distinguish a target from distractors and/or to update a discourse model. Cognitive flexibility was also correlated with children's ability to repair initially under-informative descriptions, which the authors suggest is because flexibility allows consideration of the referent from a different angle in order to generate new descriptions.

Overall, there is evidence that executive capacity explains variance on referential communication tasks although why specific relations hold and how these vary from comprehension to production tasks is not yet clear. Inhibition seems to be important for comprehension and future studies could establish if this is generally the case or whether this relation depends on the set up of referential communication tasks (where inhibiting a prepotent reach to respond is important). Future research might also assess whether working memory is less important in tasks that manipulate what is visually available (which is given at the time of testing and thus has a low memory demand) compared to what is available in

For Peer Review Only

the discourse or in social common ground (which needs to be remembered from prior interaction – episodic memory - and updated as the discourse unfolds – working memory).

Narrative. Narrative production requires organising a large amount of linguistic information and producing it in a given order such that a listener may be able to follow it. The need to manage common ground and update one's model of what has been said over an extended period of time suggests an important role for executive functions. This may explain why children with SCD struggle to convey as much plot content as typically developing peers (often omitting initiating events and the story outcome). Ketelaars, Jansonius, Cuperus, and Verhoeven (2012) derived a composite EF measure (covering planning, inhibition, cognitive flexibility and working memory) and found that, once language ability was controlled for, EF was predictive of narrative productivity (i.e., length) in children with Pragmatic Language Impairment, explaining an additional 9% of variance (NB children with PLI – now referred to as SCD – were defined as those with a pragmatic composite below the cutoff score of 132 on the CCC). In contrast, no specific links with EF held for typically developing children once formal language ability was controlled for. Narrative length is not a measure of pragmatic ability, however, so future research could clarify these findings.

With a large group of children with DLD and a group of typically developing children, Blom and Boerma (2016) measured narrative comprehension and production (macrostructure measure). For the DLD group, they found working memory was associated with concurrent narrative comprehension and production as well as production measured a year later. For the typically developing group, only concurrent narrative comprehension was significantly associated with working memory (although other correlations approached significance). This study also measure sustained attention and, while this was not included as an executive function in this review, it is worth noting that mediation analyses suggested an important role for sustained attention in story generation.

For Peer Review Only

Looking more specifically at referring expression production in narrative, Kuijper, Hartman and Hendriks (2015) found that working memory and inhibition were predictive of appropriate referent reintroduction (using a noun instead of a pronoun), although formal language ability was not controlled for.

Irony. Just as irony comprehension should require mentalising skill, it should also require holding in mind different perspectives and potentially inhibiting one protagonist's point of view in order to understand the other's. Filippova and Astington (2008) observed a medium correlation between working memory and the performance of typically developing children on irony comprehension tasks, although the degree to which it explained unique variance is unknown since it was used only as a control measure. Godbee and Porter (2013) found a similar association for a typically developing group but no role for working memory in children and adults with Williams Syndrome (possibly due to floor effects). Finally, Caillies, et al. (2014) measured working memory and inhibitory control with multiple tasks and found a large correlation between inhibitory control and irony comprehension for typically developing children but surprisingly not for their ADHD group, although the authors note that the small sample size limits conclusions. Overall, the picture is mixed but it would be worth exploring the role of inhibition and working memory further.

Discussion

Of the three domains considered in this review, formal language was the one most consistently associated with pragmatic ability. Once this had been controlled for, a number of studies found that a measure of mentalising explained further variance in pragmatic skills ranging from contingent conversation to irony comprehension. When no association with mentalising was observed, it was not clear whether this was because the specific type of mentalising measured was not required for the pragmatic task in question or whether collinearity with control variables could explain null findings. While there were fewer studies

For Peer Review Only

on executive functions and these covered a broad range of skills, there was evidence that inhibitory control is associated with communicative perspective taking, and that working memory and generativity are important for some tasks. Overall, however, few studies examining the relationship between pragmatics, on the one hand, and either mentalising or EF on the other, had: a) a sufficient sample size for correlational analyses; b) good quality measures yielding sufficient variance; c) a clear information processing rationale and d) controls for theoretically important covariates (e.g., formal language, non-verbal IQ).

Given the overall conclusion that all three domains are likely to be important for pragmatic function but in currently unspecified ways, we need to consider how future research would be best directed. One obvious next step would be to unpack why the broad associations that are observed to hold do so for specific pragmatic functions (or not). To date, generic measures of pragmatic ability and associated variables have often been employed and, while these have been helpful for establishing global links and providing cognitive profiles for developmental disorders, these do not allow for mechanistic explanations of the psycholinguistic processes that constitute pragmatic language use or the developmental processes that explain pragmatic growth. Pinning down the specific associates of every possible pragmatic skill (and every candidate social and cognitive underpinning) individually is a quite an undertaking and has the potential of generating an uninterpretable mass of data. To make progress on this front, then, a number of things need to happen.

First, we need to empirically test potential taxonomies of pragmatics skills (e.g., as put forward by O'Neill, 2012). Factor analytic work in pragmatics is rare. Given the extreme diversity of abilities that make their way into pragmatic tests (Russel & Grizzle, 2008) it is important to get a better grip of the dimensional structure of this heterogeneous set of skills – and to consider stability over developmental time. Unfortunately, the same problem holds for the other social and cognitive domains reviewed. Ideally, then, we would consider a set of

For Peer Review Only

these related domains in order to establish whether or not pragmatic skills can sensibly be thought of as a natural kind, separate from formal language skills and other aspects of cognition. Once we have an understanding of the dimensional structure of cognition and a clearly articulated taxonomy of pragmatic skills, systematically studying the development of different branches will be more manageable.

Second, we need to establish which pragmatic skills matter the most for everyday functioning and wellbeing. Doing this will help direct research at those skills that will be most consequential. For example, being able to maintain a conversation by responding to a partner is presumably very important for peer relations at every stage in development (Slomkowski & Dunn, 1996), whereas the ability to understand sarcasm might be less essential, especially for pre-adolescents. Ideally, we would be able to develop lab-based measures that are highly related to measures of everyday functioning and parent/teacher reports. While it has traditionally been held that lab tasks necessarily bleach out the very context-specific challenges that everyday conversation involves, there is recent work to suggest, for example, that parent reports of conversation skills (measured by the LUI) correlate with some carefully designed lab tasks (e.g., Abbot-Smith, Nurmsoo, Croll, Ferguson & Forrester, 2016).

Third, for each important pragmatic function, we need to aim for a mechanistic account of its use, taking an information processing approach to specify which social and cognitive sub-processes are called upon and therefore may explain atypical development. For example, given an individual who has difficulties with discourse contingency, we need to establish whether this is because they have difficulty with, say, switching from their own topic of interest to that of a conversational partner, quickly accessing lexical and grammatical content, understanding the informational needs of the conversational partner, attending to the same things, generating new ideas in order to elaborate on the conversational partner's

For Peer Review Only

comment and/or integrating all of these things in time. Without clearly specifying expected sub-processes, studies run the risk of measuring social and cognitive associates that actually do not have that much in common with the kinds of things we expect a child would need to do for the given pragmatic task in question. This is particularly important given the task specificity of many cognitive-pragmatic associations found in the adult psycho-linguistic literature (e.g., Ryskin et al., 2015).

Fourth, once we have specific hypotheses about mechanisms, we need to develop tests of pragmatic and related skills that lend themselves to individual differences research (Cronbach, 1957; Hedge, Powell & Sumner, 2017). The availability of suitable measures was perhaps the most important limiting factor, alongside sample size, for studies in this review (see also Brown-Schmidt & Fraundorf, 2015 for similar conclusions in the adult literature). Many studies could not test for an association because one measure used was at floor or was otherwise problematic (sometimes because it was directly imported from experimental research and not ideal for correlational research). There are relatively few tests of individual differences in pragmatic skills available that have good psychometric properties in terms of reliability, validity and distributional properties. Indeed, those that do exist tend to be very generic in nature, leaving us with the problem of identifying specific links. Some studies took the approach of having a large number of items on a test of a single skill that was well adapted to developmental level so as to yield the requisite variance (something that demands substantial piloting). A few studies used multiple measures and collapsed them onto one, for example, using principle components analysis. This has the advantage of removing task-specific variance and avoiding running a large number of correlations (with the risk of type I errors). However, while in principle this could yield a specific measure, in practice it often resulted in a general measure (i.e., measurement quality often trades-off with specificity). While a global measure was the goal for the reported research, we will want more specific

For Peer Review Only

measures in the future. Other work has employed scales, such as the one developed by Wellman and colleagues to measure ToM (Wellman & Liu, 2004). Often these have been shown to form Guttman scales so we know that children are likely to pass through steps on them in a given order. The assumption is that the individual skills on the scale are progressive expressions of, for example, a single underlying mentalising construct. Whether this assumption is valid for all scales and what relative success on them tells us about performance on a given pragmatic task is not always clear, however. It could be that a single item on a scale taps an ability that is central to performance on a given pragmatic task and only this item is of relevance. Ideally, we would be able to analyse a given communicative task for the specific type of, say, mentalising we think is required and then test whether children who master this type of mentalising did well. The inescapable fact is that it is very challenging indeed, particularly for cross-sectional studies, to find a set of measures (for the domains of interest plus controls) that tap specific constructs (avoiding problematic levels of incidental task demands) and yield sufficient variance.

It would be particularly helpful to have non-verbal tests of covariates including mentalising. These are extremely difficult to develop for 2nd order false belief and more advanced forms of mentalising (see e.g., Freed et al., 2015). The most frequently used measures either require an advanced vocabulary (RMET, Baron-Cohen et al., 2001) or involve comprehension of vignettes of around 100 words each (e.g., Astington et al., 2002; Happé, 1994). The Animations task (Abel, Happé & Frith, 2000) does not burden language comprehension but participants are required to formulate descriptions of moving shapes that include mental state vocabulary. Film versions of Happé's (1994) Strange Stories task, as developed by Devine and Hughes (2013), are an improvement but still require participants to formulate fairly complex responses and are still related to verbal ability. While less word-heavy measures of 2nd order false belief exist (Grueneisen, Wyman & Tomasello, 2014),

For Peer Review Only

these are time-consuming to administer and not easily scalable. Without such measures, it is possible that some associations are found between 2nd order ToM and some pragmatic tests simply because they essentially measure the same verbal abilities.

The problem of cross-over of measures from one domain to another is quite frequently observed. Thus, the ability to understand jokes formed parts of a mentalising scale for at least one study reviewed here (Hale & Tager-Flusberg, 2005; see also Happé, 1994) whereas in another both jokes and irony were considered aspects of pragmatic language and were thus conflated in the analyses (Angeleri & Airenti, 2014). Similarly, the Mindful Conversational Difficulties scale used by de Rosnay et al. (2014) included items on the ability to understand others' thoughts (i.e., mentalising). Elements of the CCC2 such as the scales on social relations and interests are included in measures of pragmatic ability for some studies. Similar measures of narrative (e.g., coherence) are sometimes considered pragmatic, sometimes not. Referential communication tasks are considered by some researchers to measure pragmatics (e.g., Nilsen & Graham, 2009) and by others to measure dynamic theory of mind (e.g., Begeer, Malle, Nieuwland & Keysar, 2010). Such blurring is inevitable for a field that has no clearly developed taxonomies or dimensional structure.

Fifth, assuming the above methodological challenges can be overcome, we will need to cast the net wider when considering the psychological variables that explain differences in pragmatic function. Some of the studies in the current review hinted at socio-emotional and personality traits that would be associated with pragmatics. We would also argue that lower level cognitive factors such as attentional biases and statistical learning (Bannard, Rosner & Matthews, 2017) are likely to be important. And at the other end of the spectrum, meta-cognition would deserve attention, at least in so far as poor meta-cognition might prevent an individual from improving their pragmatic proficiency (e.g., Collins, Lockton & Adams, 2014). To fully understand how these factors come to have their effects, it will be necessary

For Peer Review Only

to determine their genetic and environmental bases (see Losh, et al, 2012a; Losh et al., 2012b for an example of work seeking to uncover genetic bases of social communication).

Sixth, we need to adopt statistical and modelling tools that will enable us to make best use of a dimensional approach to the study of pragmatic development. By a dimensional approach, we mean one where an individual child's relative strengths and weaknesses are represented as a vector in multidimensional space. As mentioned above, through factor analysis, datasets with multiple measures from multiple children can help to give a sense of the dimensional structure of cognition. They can be used to derive clinical groups - children with particular clusters of strengths and weaknesses may be diagnosed with a given disorder in order to best target support. Datasets collected over longitudinal time can also be helpful in predicting outcomes (e.g., Pellicano, 2013). However, to really explain pragmatic development, we would argue that individual differences data sets should also be used for building and testing mechanistic models of development.

One challenge to overcome with using individual differences data to build models is that the dimensions of interest rarely stand in isolation from another. When looking at higher-order cognition, we assume this is because performance on tests of, say, pragmatics, mentalising and inhibition, all call on related sub-processes. That is, different skills are often nested in others or are related in an interconnected hierarchy of cognitive functions. Basic correlation and regression will not yield such a hierarchy or account for this nested structure and thus we are often left at a dead end. For example, if when predicting a general measure of pragmatic ability we control for vocabulary, grammar, IQ and some measure of executive function and we then fail to see an association with mentalising, is this because we do not take the other's perspective during this task, is it because we do but everyone manages to do so equally, or is it because real variance in this ability was already accounted for by control measures that were each somewhat correlated with mentalising because they were drawing

For Peer Review Only

on the same cognitive sub-processes? This kind of problem seemed to be quite common in this review. And of course the better controlled the study, the more likely the problem is. To make progress we need to make sense of the collinearity we see, rather than simply analysing it away. That is, we could use multidimensional data sets to generate (and/or test theoretical proposals about) an underlying hierarchy of cognitive functions and thereby identify plausible cognitive sub-processes that are called upon to deploy a range of higher order functions. These sub-processes could in turn be verified by looking for their neural signatures. For this process to be rigorous and to avoid likely problems with publication bias, theory-driven analyses will ideally be preregistered.

The argument being made here is not a reductionist one, but rather a call to integrate the results of modelling indiviudal differences data and complementary experimental work (which would appear on the basis of the current review to be essential) into psycholinguistic models of language processing. Pragmatic phenomena have often caused problems for the construction of language processing models. For example, Hagoort and van Berkum (2007) review evidence that contextual information rapidly influences utterance interpretation (contra original assumptions that the language system first computes the meaning of an utterance and then engages in pragmatic processes). If we adopt Levinson's (1983) Artificial Intelligence definition of pragmatics, where the contextual information that needs to be considered covers the entirety of interlocutors' knowledge of the world, then it becomes clear why integrating pragmatic processes into models of language processing is very rarely attempted. Allowing for individual differerences in model architecture only compounds the difficulty. We might therefore wonder whether it is naïve to aim for a full working model of language use (which necessitates a model of cognition generally) in the same way as we might have a model of, say, how the heart works. While we certainly need to be careful to articulate the level of explanation we are looking for (Rosenblueth & Wiener, 1945),

For Peer Review Only

ultimately, a mechanistic model of language processing systems is the only satisfying explanation worth having. Since we have seen how cognitive bottlenecks to pragmatic function are not persistent over development (e.g., Gillis & Nilsen, 2014), a valuable model will be one that changes over time. Certainly it seems worth directing research on pragmatics to explicitly aim for this goal.

In sum, there is considerable potential to make progress in understanding how we achieve the communicative feats we do, why some children struggle to achieve them, and what can be done to help those who face difficulties. The cultural differences in many aspects of pragmatic language (e.g., Filippova, 2014; Küntay, Nakamura & Ateş-Şen, 2014) mean that it will not always make sense to say that someone is 'good' at pragmatics (nor to only seek to change children's skills to the norm if they develop atypically). However, there is significant potential to help children for whom pragmatic difficulities impair wellbeing (Adams et al., 2014; Pickles et al., 2016). Individual differences research has the potential to contribute to this progress but we need to overcome a few challenges in order to take a programmatic approach to the problem.

For Peer Review Only

References

Note: **References preceded by an asterisk indicate articles included in the table of reviewed studies (Appendix B).**

Abel, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development,* 15(1), 1-16.

Abbot-Smith, K., Nurmsoo, E., Croll, R., Ferguson, H & Forrester, M. (2016). Awareness of 2½-year-olds to levels of listener informational need when making requests. *Journal of Child Language, 43(6)*: 1277-1291.

* Adachi, T., Koeda, T., Hirabayashi, S., Maeoka, Y., Shiota, M., Wright, C., & Wada, A. (2004). The metaphor and sarcasm scenario test: A new instrument to help differentiate high functioning pervasive developmental disorder from attention defecit/hyperactivity disorder. *Brain and Development, 26*(5), 301-306.

Adams, C. (2002). Practitioner review: The assessment of language pragmatics. *Journal of Child Psychology and Psychiatry*, 43(8), 973-987.

Adams, C., Green, J., Gilchrist, A., & Cox, A. (2002). Conversational behaviour of children with Asperger syndrome and conduct disorder. *Journal of Child Psychology and Psychiatry*, *43*(5), 679-690.

* Akbar, M., Loomis, R., & Paul, R. (2013). The interplay of language on executive functions in children with ASD. *Research in Autism Spectrum Disorders, 7*(3), 494-501.

* Andrés-Roqueta, C., Adrian, J., Clemente, R., & Katsos, N. (2013). Which are the best predictors of theory of mind delay in children with specific language impairment? *International Journal of Language & Communication Disorders, 48*(6), 726-737.

*Angeleri, R., & Airenti, G. (2014). The development of joke and irony understanding: A study with 3-to 6-year-old children. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *68*(2), 133.

Ariel, M. (2010). *Defining pragmatics*. Cambridge University Press.

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition, 110*(2), 131-146. doi:10.1016/j.cognition.2008.10.016

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's 2nd order false-belief understanding and their ability to reason about evidence. *New ideas in Psychology*, *20*(2), 131-144.

For Peer Review Only

Astington, J. W., & Baird, J. A. (Eds.). (2005). *Why language matters for theory of mind*. Oxford University Press.

*Banasik, N. (2013). Non-literal speech comprehension in preschool children–An example *Language and Communication*, *17*(3), 309-324from a study on verbal irony. *Psychology of Language and Communication*, *17*(3), 309-324.

Banerjee, R. (2000). The development of an understanding of modesty. *British Journal of Developmental Psychology*, *18*(4), 499-517.

Bannard, C., Rosner, M., & Matthews, D. (2017). What's worth talking about? Information theory reveals how children balance informativeness and ease of production. *Psychological Science*, *28*(7), 954-966

*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.*

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: cognitive or affective? *Journal of Autism and Developmental Disorders, 18*(3), 379-402.

Bauer, P., & Burch, M. (2004). Developments in early memory: multiple mediators of foundational processes. In J. Lucariello, J. Hudson, R. Fivush, & P. Bauer, *The development of the mediated mind: Culture and cognitive development* (pp. 101-125). Mahwah: NJ : Lawrence Erlbaum Associates.

Begeer, S., Malle, B., Nieuwland, M. & Keysar, B. (2010). Using Theory of Mind to represent and take part in social interactions: Comparing individuals with high-functioning autism and typically developing controls. *European Journal of Developmental Psychology, 7*(1), 104–122

For Peer Review Only

* Bernard, S., & Deleau, M. (2007). Conversational perspective-taking and false belief attribution: A longitudinal study. *British Journal of Developmental Psychology , 25*(3), 443-460.

Berument, S., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: diagnostic validity. *The British Journal of Psychiatry , 175*(5), 444-451.

Bishop, D. (1993). Annotation: autism, executive functions and theory of mind: A neuropsychological perspective. *Journal of Child Psychology and Psychiatry , 34*, 279-293.

Bishop, D. (2003). *The Children's Communication Checklist: CCC-2.* London: Harcourt Assessment.

* Bishop, D., & Norbury, C. (2005). Executive functions in children with communication impairments, in relation to autistic symptomatology. 2: Response inhibition. *Autism: The International Journal of Research and Practice, 9*(1), 29-43.

* Blain-Brière, B., Bouchard, C., & Bigras, N. (2014). The role of executive functions in the pragmatic skills of children aged 4-5 . *Frontiers in Psychology*, 5.

Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure?. *Research in Developmental Disabilities*, *55*, 301-311.

Bonifacio, S., Girolametto, L., Bulligan, M., Callegari, M., Vignola, S., & Zocconi, E. (2007). Assertive and responsive conversational skills of Italian□speaking late talkers. *International Journal of Language & Communication Disorders*, *42*(5), 607-623.

Boudreau, D. (2005). Use of a parent questionnaire in emergent and early literacy assessment of preschool children . *Language, Speech and Hearing Services in Schools, 36*(1), 33-47.

For Peer Review Only

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893-900.

Brown-Schmidt, S., & Fraundorf, S. H. (2015). Interpretation of informational questions modulated by joint knowledge and intonational contours. *Journal of Memory and Language*, *84*, 49-74.

Bruner, J. (1990). *Acts of Meaning.* Cambridge, MA. Harvard University Press. Developmental Psychology, 13(4), 377-388.

*Caillies, S., Bertot, V., Motte, J., Raynaud, C., & Abely, M. (2014). Social cognition in ADHD: Irony understanding and recursive theory of mind. *Research in Developmental Disabilities*, *35*(11), 3191-3198.

Camarata, S. M., & Gibson, T. (1999). Pragmatic language deficits in attention□deficit hyperactivity disorder (ADHD). *Developmental Disabilities Research Reviews*, *5*(3), 207-214.

*Capps, L., Losh, M., & Thurber, C. (2000). "The frog ate the bug and made his mouth sad": Narrative competence in children with autism. *Journal of Abnormal Child Psychology*, *28*(2), 193-204.

* Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism, 2*(4), 325-344.

Carrow-Woolfolk, E. (1995). *OWLS, Oral and Written Language Scales*. NCS Pearson Incorporated.

Carrow-Woolfolk, E. (1999). *CASL: Comprehensive Assessment of Spoken Language* . American Guidance Service.

Carter, J., Lees, J., Murira, G., Gona, J., Neville, B., & Newton, C. (2005). Issues in the development of cross-cultural assessments of speech and language for children. *International Journal of Language & Communication Disorders, 40*(4), 385-401.

For Peer Review Only

Clark, E. (2005). Pragmatics and language acquisition. In Horn, L. & Ward, G. (Eds.). *The Handbook of Pragmatics*. Blackwell

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Cohen, M. (1997). *Children's Memory Scale*. San Antonio, TX: Psychological Corporation.

Collins, A., Lockton, E., & Adams, C. (2014). Metapragmatic explicitation ability in children with typical language development: Development and validation of a novel clinical assessment. *Journal of Communication Disorders*, *52*, 31-43.

Colozzo, P., Morris, H., & Mirenda, P. (2015). Narrative production in children with autism spectrum disorder and specific language impairment. *Canadian Journal of Speech-Language Pathology & Audiology, 39*(4).

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12(11), 671.

* Dahlgren, S., & Dahlgren Sandberg, A. (2008). Referential communication in children with autism spectrum disorder. *Autism, 12*(4), 335-348.

* Davies, C., Andrés-Roqueta, C., & Norbury, C. (2016). Referring expressions and structural language abilities in children with Specific Language Impairment: A pragmatic tolerance account. *Journa of Experimental Child Psychology, 144*, 98-113.

De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*(1), 99-107.

*De Marchena, A., & Eigsti, I. M. (2016). The art of common ground: Emergence of a complex pragmatic language skill in adolescents with autism spectrum disorders. *Journal of Child Language*, *43*(1), 43-80.

* 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*(05), 1179-1193.

For Peer Review Only

De Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio□cognitive understanding?. *British Journal of Developmental Psychology*, *24*(1), 7-37.

Deleau, M. (1999). Assessing links between conversational awareness and cognitive development. *Developmental Science , 2*(1), 16-18.

Devine, R. & Hughes, C. (2013). Silent films and strange stories: theory of mind, gender and social experiences in middle childhood. *Child Development,* 84(3): 989-1003.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*, 135-168.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development*, *62*(6), 1352-1366.

Dunn, L., & Dunn, L. (1981). PPVT-R Manual. *Circle Pines, MN: American Guidance Service*.

Dunn, D. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test: Manual*. Pearson.

Durkin, K. (1987). Minds and Language: Social cognition, social interaction and the development of language. *Mind & Language, 2*, 105-140.

Eaves, L. C., & Ho, H. H. (2008). Young adult outcome of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 739-747.

- Farrant, B., Maybery, M., & Fletcher, J. (2010). Socio-emotional engagement, joint attention, imitation and conversational skill: Analysis in typical development and specific language impairment . *First Language, 31*(1), 23-46.

Fenson, L., Dale, P., Reznick, S., Thai, D., Bates, E., Hartung, J., . . . Reilly, J. (1994). *MacArthur Communicative Development Inventories user's guide and technical manual.* San Diego: Thomson Learning.

For Peer Review Only

* Fernández, C. (2013). Mindful storytellers: Emerging pragmatics and theory of mind development. *First Language, 33*(1), 20-46.

Filippova, E. (2014). Developing appreciation of irony in Canadian and Czech discourse. *Journal of Pragmatics, 74*: 209-223.

* Filippova, E., & Astington, J. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79*(1), 126-138.

Freed, J., McBean, K., Adams, C., Lockton, E., Nash, M. & Law, J. (2015). Performance of children with social communication disorder on the Happé Strange Stories: physical and mental state responses and relationship to language ability. *Journal of Communication Disorders, 55*: 1-14.

*Fukumura, K. (2016). Development of audience design in children with and without ASD. *Developmental Psychology*, *52*(1), 71.

Gallagher, T. (1991). A retrospective look at clinical pragmatics. In T. Gallagher, *Pragmatics of Language: Clinical Practice Issues* (Vols. 1-9). New York: Chapman and Hall.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., ... & Walters, J. (2012). *MAIN: Multilingual assessment instrument for narratives*. Zentrum für Allgemeine Sprachwissenschaft.

Garon, N., Bryson, S., & Smith, I. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological Bulletin, 134*(1), 31.

Geurts, H., Broeders, M., & Nieuwland, M. (2010). Thinking outside the executive functions box: Theory of mind and pragmatic abilities in attention deficit/hyperactivity disorder . *European Journal of Developmental Psychology, 7*(1), 131-151.

* Gillespie-Lynch, K., Khalulyan, A., del Rosario, M., McCarthy, B., Gomez, L., Sigman, M., & Hutman, T. (2015). Is early joint attention associated with school-age pragmatic language? *Autism, 19*(2), 168-177.

* Gillis, R., & Nilsen, E. (2014). Cognitive flexibility supports preschoolers' detection of communicative ambiguity. *First Language, 34*(1), 58-71.

For Peer Review Only

Gilmour, J., Hill, B., Place, M., & Skuse, D. H. (2004). Social communication deficits in conduct disorder: a clinical and community survey. *Journal of Child Psychology and Psychiatry*, *45*(5), 967-978.

Girolametto, L. (1997). Development of a parent report measure for profiling the conversational skills of preschool children. *American Journal of Speech-Language Pathology, 6*(4), 25-33.

Glucksberg, S., & Krauss, R. (1967). What do people say after they have learned to talk? Studies of development of referential communication . *Merrill-Palmer Quarterly, 13*, 309-316.

*Godbee, K., & Porter, M. (2013). Comprehension of sarcasm, metaphor and simile in Williams syndrome. *International journal of language & communication disorders*, *48*(6), 651-665.

Gottman, J., Gonso, J., & Rasmussen, B. (1975). Social interaction, social competence, and friendship in children. *Child Development*, 709-718.

Grueneisen, S. Wyman, E. & Tomasello, M. (2014). "I know that you don't know I know" Children use 2nd order false-belief reasoning for peer coordination. *Child Development,* 86(1): 287-293.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 801-811.

* Hale, C., & Tager-Flusberg, H. (2005). Social communication in children with autism: The relationship between theory of mind and discourse development. *Autism, 9*(2), 157-178.

For Peer Review Only

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129-154.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1-21.

Helland, W. A., Lundervold, A. J., Heimann, M., & Posserud, M. B. (2014). Stable associations between behavioral problems and language impairments across childhood–The importance of pragmatic language problems. *Research in Developmental Disabilities*, *35*(5), 943-951.

*Huang, S. F., Oi, M., & Taguchi, A. (2015). Comprehension of figurative language in Taiwanese children with autism: The role of theory of mind and receptive vocabulary. *Clinical Linguistics & Phonetics*, *29*(8-10), 764-775.

Hughes, C. (2011). *Social understanding and social lives: From toddlerhood through to the transition to school*. Psychology Press.

Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. (2005). Origins of individual differences in theory of mind: from nature to nurture?. *Child Development*, *76*(2), 356-370.

Jansonius-Schultheiss, K., Borgers, M., DeBruin, B., & Stumpel, H. (2006). Renfrew's language scales Dutch adaptation. *Amsterdam: Pro-education, HvA*.

Jones, G., Gobet, F., Freudenthal, D., Watson, S. E., & Pine, J. M. (2014). Why computational models are better than verbal theories: the case of nonword repetition. *Developmental Science*, *17*(2), 298-310.

Jones, C. D., & Schwartz, I. S. (2009). When asking questions is not enough: An

For Peer Review Only

observational study of social communication differences in high functioning children with autism. *Journal of Autism and Developmental Disorders*, *39*(3), 432-443.

Joseph, R. M., McGrath, L. M., & Tager-Flusberg, H. (2005). Executive dysfunction and its relation to language ability in verbal school-age children with autism. *Developmental Neuropsychology*, *27*(3), 361-378.

Ketelaars, M., Cuperus, J., Jansonius, K., & Verhoeven, L. (2010). Pragmatic language impairment and associated behavioural problems. *International Journal of Language & Communication Disorders, 45*(2), 204-214.

* Ketelaars, M., Jansonius, K., Cuperus, J., & Verhoeven, L. (2012). Narrative competence and underlying mechanisms in children with pragmatic language impairment. *Applied Psycholinguistics, 33*(02), 281-303.

Kemple, K., Speranza, H., & Hazen, N. (1992). Cohesive discourse and peer acceptance: Longitudinal relations in the preschool years. *Merrill-Palmer Quarterly (1982-)*, 364-381.

Kingdon, D., Cardoso, C., & McGrath, J. J. (2016). Research Review: Executive function deficits in fetal alcohol spectrum disorders and attention□deficit/hyperactivity disorder–a meta□analysis. *Journal of Child Psychology and Psychiatry*, *57*(2), 116-131.

Korkman, M., Kirk, U., & Kemp, S. L. (2003). Bilan neuropsychologique de l'enfant. ECPA.

Küntay, A., Nakamura, K., & Ateş-Şen, A. B. (2014). Crosslinguistic and crosscultural approaches to pragmatic development. In D. Matthews (Ed.) *Pragmatic Development in First Language Acquisition* (pp. 317-341). Amsterdam: John Benjamins.

Lai, C. L. E., Lau, Z., Lui, S. S., Lok, E., Tam, V., Chan, Q., ... & Cheung, E. F. (2017). Meta□analysis of neuropsychological measures of executive functioning in children

For Peer Review Only

and adolescents with high□functioning autism spectrum disorder. *Autism Research*, *10*(5), 911-939.

Landry, O., & Al-Taie, S. (2016). A meta-analysis of the Wisconsin Card Sort Task in autism. *Journal of Autism and Developmental Disorders*, *46*(4), 1220-1235.

* Leonard, M., Milich, R., & Lorch, E. (2011). The role of pragmatic language use in mediating the relation between hyperactivity and inattention and social skills problems. *Journal of Speech, Language and Hearing Research, 54*(2), 567-579.

Lewis, F. M., Woodyatt, G. C., & Murdoch, B. E. (2008). Linguistic and pragmatic language skills in adults with autism spectrum disorder: A pilot study. *Research in Autism Spectrum Disorders*, *2*(1), 176-187.

Lloyd, P., Camaioni, L., & Ercolani, P. (1995). *Test of Referential Communication: (TREC) skills.* London : The Psychological Corporation.

Lloyd, P., Peers, I., & Foster, C. (2001). *The Listening Skills Test.* London: The Psychological Corportation.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study . *Child Development, 74*(4), 1130-1144.

*Losh, M., & Capps, L. (2003). Narrative ability in high-functioning children with autism or Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *33*(3), 239-251.

Lord, C., Risi, S., Lambrecht, L., Cook Jr, E., Leventhal, B., DiLavore, P., . . . Rutter, M. (2000). The Autism Diagnostic Observation Schedule - Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

For Peer Review Only

* Losh, M., Martin, G., Klusek, J., Hogan-Brown, A., & Sideris, J. (2012). Social communication and theory of mind in boys with autism and fragile X syndrome. *Frontiers in Psychology, 3*, 1-12.

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*: 890-904.

* Maridaki-Kassotaki, K., & Antonopoulou, K. (2011). Examination of the relationship between false-belief understanding and referential communication skills. *European Journal of Psychology of Education , 26*(1), 75-84.

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language*, *85*(3), 451-466.

*Massaro, D., Valle, A., & Marchetti, A. (2014). Do social norms, false belief understanding and metacognitive vocabulary influence irony comprehension? A study of five-and seven-year-old children. *European Journal of Developmental Psychology, 11*(3), 292-304.

Matthews, D., Butcher, J., Lieven,& Tomasello, M. (2012). Two□and four□year□olds learn to adapt referring expressions to context: effects of distracters and feedback on referential communication. *Topics in Cognitive Science*, *4*(2), 184-210.

Matthews, D. (2014). *Pragmatic Development in First Language Acquisition* (Vol. 10). Amsterdam: John Benjamins.

Mayer, M. (1969). *Frog Where Are You?* New York: Puffin Pied Piper.

McDonald, S., & Van Sommers, P. (1993). Pragmatic language skills after closed head injury: Ability to negotiate requests. *Cognitive Neuropsychology, 10*(4), 297-315.

For Peer Review Only

* McEvoy, R., Rogers, S., & Pennington, B. (1993). Executive function and social communication deficits in young autistic children. *Journal of Child Psychology and Psychiatry, 34*(4), 563-578.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56-63.

*Mewhort□Buist, T. A., & Nilsen, E. S. (2013). What are you really saying? Associations between shyness and verbal irony comprehension. *Infant and Child Development*, *22*(2), 180-197.

Miller, M., Young, G. S., Hutman, T., Johnson, S., Schwichtenberg, A. J., & Ozonoff, S. (2015). Early pragmatic language difficulties in siblings of children with autism: implications for DSM□5 social communication disorder?. *Journal of Child Psychology and Psychiatry*, *56*(7), 774-781.

Milligan, K., Astington, J., & Dack, L. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding . *Child Development , 78*(2), 622-646.

* Miniscalco, C., Rudling, M., Rastam, M., Gillberg, C., & Johnels, J. (2014). Imitation (rather than core language) predicts pragmatic development in young children with ASD: a preliminary longitudinal study using CDI parental reports. *International Journal of Language & Communication Disorders, 49*(3), 369-375.

For Peer Review Only

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100. Chicago

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8-14.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *Early social communication scales (ESCS*). Coral Gables, FL: University of Miami

Murphy, S. M., Faulkner, D. M., & Farley, L. R. (2014). The behaviour of young children with social communication disorders during dyadic interaction with peers. *Journal of Abnormal Child Psychology*, *42*(2), 277-289.

Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329-336.

*Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Research*, *2*(6), 334-347.

*Nicholson, A. T., Whalen, J. M., & Pexman, P. M. (2013). Children's Processing of Verbal Irony: The Earliest Moments. *Canadian Journal of Experimental Psychology*, *67*(4), 292.

*Nilsen, E. S., Glenwright, M., & Huyder, V. (2011). Children and adults understand that verbal irony interpretation depends on listener knowledge. *Journal of Cognition and Development*, *12*(3), 374-409.

* Nilsen, E., & Graham, S. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58*(2), 220-249.

For Peer Review Only

* Nilsen, E., & Graham, S. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development, 83*(4), 1400-1415.

Norbury, C. F. (2014). Practitioner review: Social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry*, 55(3), 204-216.

Norbury, C. F., & Bishop, D. V. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287-313.

*Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: a cross-disorder comparison. *Journal of Child Language*, *41*(3), 485-510.

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: a validation of the Children's Communication Checklist—2. *International Journal of Language & Communication Disorders*, *39*(3), 345-364.

O'Neill, D. K. (2007). The language use inventory for young children: A parent-report measure of pragmatic language development for 18-to 47-month-old children. *Journal of Speech, Language, and Hearing Research*, *50*(1), 214-228.

O'Neill, D. (2009). Language Use Inventory: An assessment of young children's pragmatic language development for 18-to 47-month-old children [Manual]. *Waterloo, Ontario, Canada: Knowledge in Development*.

O'Neill, D. (2012). Components of pragmatic ability and children's pragmatic language development. In H. Schmid (Ed.), *Cognitive Pragmatics* (Vol. 4, pp. 261-287). Walter de Gruyter.

O'Neill, D. K. (2014). Assessing pragmatic language functioning in young children. In D. Matthews (Ed.) *Pragmatic Development in First Language Acquisition* (pp. 363-386).

For Peer Review Only

Amsterdam: John Benjamins.

*O'Reilly, K., Peterson, C. & Wellman, H. (2014). Sarcasm and advanced Theory of Mind understanding in children and adults with prelingual deafness. *Developmental Psychology, 50*(7): 1862-1877.

*Pellicano, E. (2013). Testing the predictive power of cognitive atypicalities in autistic children: evidence from a 3-year follow-up study. *Autism Research, 6*(4), 258-267.

Pelletier, J., & Astington, J. W. (1998). Metacognitive vocabulary test. *Unpublished ms., Institute for Child Studies, University of Toronto*.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, 689-700.

Pesco, D., & O'Neill, D. K. (2012). Predicting later language outcomes from the Language Use Inventory. *Journal of Speech, Language, and Hearing Research*, *55*(2), 421-434.

Peterson, C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language, 15*(1), 123-145.

Peterson, C., Garnett, M., Kelly, A., & Attwood, T. (2009). Everyday social and conversation applications of theory-of-mind understanding by children with autism-spectrum disorders or typical development. *European Child & Adolescent Psychiatry, 18*(2), 105-115.

*Pexman, P. M., Zdrazilova, L., McConnachie, D., Deater-Deckard, K., & Petrill, S. A. (2009). "That Was Smooth, Mom": Children's Production of Verbal and Gestural Irony. *Metaphor and Symbol*, *24*(4), 237-248.

Phelps-Terasaki, D., & Phelps-Gunn, T. (2007). *TOPL-2 Test of Pragmatic Language: Examiner's Manual.* San Antonio, TX: The Psychological Corporation.

For Peer Review Only

Philofsky, A., Fidler, D., & Hepbum, S. (2007). Pragmatic language profile of school-age children with autism spectrum disorders and Williams syndrome. *American Journal of Speech-Language Pathology, 16*(4), 368-380.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*(2), 127-152.

Racine, T. P., & Carpendale, J. I. (2007). The role of shared practice in joint attention. *British Journal of Developmental Psychology*, *25*(1), 3-25.

Ren, X., Wang, T., & Jarrold, C. (2016). Individual differences in frequency of inner speech: Differential relations with cognitive and non-cognitive factors. *Frontiers in Psychology*, *7*.

Renfrew, C. E. (2010). *Bus Story Test: A test of narrative speech*. Speechmark.

*Resches, M., & Pereira, M. P. (2007). Referential communication abilities and theory of mind development in preschool children. *Journal of Child Language*, *34*(01), 21-52.

* Rinaldi, P., Baruffaldi, F., Burdo, S., & Caselli, M. (2013). Linguistic and pragmatic skills in toddlers with cochlear implant. *International Journal of Language & Communication Disorders , 48*(6), 715-725.

* Rints, A., McAuley, T., & Nilsen, E. (2015). Social communication is predicted by inhibitory ability and ADHD traits in preschool-aged children a mediation model . *Journal of Attention Disorders, 19*(10), 901-911.

Rosenblueth, A., & Wiener, N. (1945). The role of models in science. *Philosophy of Science*, 12(4), 316-321.

Rosenthal, R. (1991). *Meta-analytic Procedures for Social Research* (Vol. 6). Sage.

For Peer Review Only

Russell, R., & Grizzle, K. (2008). Assessing child and adolescent pragmatic language competencies: Toward evidence-based assessments. *Clinical Child and Family Psychology Review, 11*(1-2), 59-73.

Russell, J., Jarrold, C., & Hood, B. (1999). Two intact executive capacities in children with autism: Implications for the core executive dysfunctions in the disorder. *Journal of Autism and Developmental Disorders*, *29*(2), 103-112.

Saborit, C., & Julián, J. (2005). *L'avaluació del llenguatge infantil. ELI.* . Universitat Jaume I de Castelló (UJI). Col· lecció: Educació.

Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3□Year□Old children make relevance inferences in indirect verbal communication. *Child Development, 84*(6), 2079-2093.

Seibert, J., Hogan, A., & Mundy, P. (1987). Assessing social communication skills in infancy. *Topics in Early Childhood Special Education, 7*, 32-48.

Semel, E. M., Wiig, E. H., & Secord, W. (1995). *CELF3: Clinical Evaluation of Language Fundamentals*. Psychological Corporation, Harcourt Brace.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals (CELF-4)* The Psychological Corporation.

Sherrod, L., & Lamb, M. (1981). Infant social cognition: An introduction. In L. Michael, & L. Sherrod, *Infant Social Cognition: Empirical and Theoretical Considerations* (pp. 1-10). Hillsdale: NJ: Erlbaum.

Sjöwall, D., Roth, L., Lindqvist, S., & Thorell, L. B. (2013). Multiple deficits in ADHD: executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry*, 54(6), 619-627.

*Slomkowski, C., & Dunn, J. (1996). Young children's understanding of other people's beliefs and feelings and their connected communication with friends. *Developmental Psychology*, *32*(3), 442.

For Peer Review Only

Sparrow, S., Balla, D., & Cichetti, D. (1984). *Vineland Adaptive Behaviour Scales.* American Guidance Services. MN: Circle Pines.

Spitzberg, B., & Adams, T. (2007). *CSRS, the Conversational Skills Rating Scale: an Instructional Assessment of Interpersonal Competence.* NCA, National Communication Association.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 360-407.

St Clair, M. C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders*, *44*(2), 186-199.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute 2nd order beliefs. *Developmental Psychology*, *30*(3), 395.

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of 2nd□order mental state attributions. *British Journal of Developmental Psychology*, *13*(2), 191-204.

Tager-Flusberg, H. (1996). Brief report: Current theory and research on language and communication in autism. *Journal of Autism and Developmental disorders, 26*(2), 169-172.

Tager-Flusberg, H. (1999). A psychological approach to understanding the social and language impairments in autism. . *International review of psychiatry, 11*(4), 325-334.

Tager-Flusberg, H., Paul, R., Lord, C., Volkmar, F., Paul, R., & Klin, A. (2005). Language and communication in autism. *Handbook of Autism and Pervasive Developmental Disorders*, *1*, 335-364.

*Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A

For Peer Review Only

comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics*, *16*(3), 241-256.

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, *15*(3), 249-275.

Volden, J. (2002). Features leading to judgements of inappropriacy in the language of speakers with autism: A preliminary study. *Journal of Speech Language Pathology and Audiology*, *26*(3), 138-146.

*Volden, J., Coolican, J., Garon, N., White, J., & Bryson, S. (2009). Brief report: pragmatic language in autism spectrum disorder: relationships to measures of ability and disability. *Journal of Autism and Developmental Disorders, 39*(2), 388-393.

Volden, J., & Phillips, L. (2010). Measuring Pragmatic Language in Speakers With Autism Spectrum Disorders: Comparing the Children's Communication Checklist-2 and the Test of Pragmatic Language. *American Journal Of Speech-Language Pathology*, *19*(3), 204-212.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review*, *20*(4), 766-772.

Welch-Ross, M. K. (1997). Mother–child participation in conversation about the past: Relationships to preschoolers' theory of mind. *Developmental Psychology*, *33*(4), 618.

Wellman, H. M., & Liu, D. (2004). Scaling of theory□of□mind tasks. *Child Development*, *75*(2), 523-541.

Whitehouse, A. J., Watt, H. J., Line, E. A., & Bishop, D. V. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language & Communication Disorders*, *44*(4), 511-528.

For Peer Review Only

*Whyte, E., & Nelson, K. (2015). Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders, 54*, 2-14.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436-452.

Williams, K. T. (1997). *Expressive Vocabulary Test Second Edition (EVT™ 2)*. Pearson.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 12*(1), 103-128.

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language, 62*(1), 89-106.

For Peer Review Only

Appendix A

Search strategy

Web of Science, SCOPUS and PsycINFO were searched by the second author between February and March 2016. Search terms were combined to cover the following domains: 1) pragmatics (pragmatic, social communication, audience design, common ground, conversation, discourse), 2) child development (child*, infan*, devel*, toddler, pre-school, preschool) and 3) the three underpinning cognitive domains (vocabulary, syntax, grammar, formal language, structural language; social cognition, social-cognition, theory of mind, false belief, joint attention, perspective taking, mental state, mentali*; executive function*, memory, cognitive flexibility, mental flexibility, set switching, inhibition, inhibitory control, executive control). Any further papers identified from other sources (e.g., references section of initially identified papers) were included. The third author expanded this search in July 2016 to include the term irony as a pragmatic key word. Case studies or studies with small groups of brain damaged individuals or individuals with rare clinical disorders were excluded along with intervention studies and studies of bilingual children. Only papers that reported a measure of the strength of association between two continuous variables were included. Studies that reported group comparisons (e.g., incidence of pragmatic impairment associated with incidence of language impairment), while reflective of the same relationship, were not considered. A meta-analysis calculator was used to report effect sizes as Cohen's d (Cohen, 1992). We consider d=0.2 a 'small' effect size, 0.5 a 'medium' effect size and 0.8 a 'large' effect size. For studies which presented data based on the same sample, a single effect size was calculated (Rosenthal, 1991).

For Peer Review Only

APPENDIX B

Table 1: Study characteristics and findings. (Effect sizes reported in parentheses as Cohen's d. Acronyms are spelt out at the end of the table.)

Author (year)	N, Population, Mean age (years; months)	Pragmatic measure	Executive function measure	Mentalising measure	Formal language measure	Findings (effect size, Cohen's d)
Adachi et al., (2004)	29, ADHD, 9;6 54, HFPDD, 9;8	Metaphor and Sarcasm Scenario Test (A written test that child need to be able to read)			Verbal IQ	For children with ADHD, VIQ was correlated with metaphor (0.89) but not with sarcasm (0.15) or landmir (-0.10) scores. For HFPDD children, VIQ was correlated with metaph (1.24) but not with sarcasm (0.17) or landmine (0.04) scores
Akbar, Loomis, and Paul (2013)	62, ASD, 8;7	CASL Pragmatic Judgement Subtest	WM – Letter number sequencing task of the WISC O – NEPSY-II animal sort subtest CF – D-KEFS trail making subtest I – D-KEFS		CELF 4 (Core Language standard score)	CASL was correlated with CELF-4 (2.35) CASL was correlated with WM (1.90) and O (0.92) bu not with CF or I. CASL correlated positively with parent-report Vinelan Communication Scale (1.17), teacher report Vineland Communication Scale (1.25), and negatively with teacher report WM (1.51)

For Peer Review Only

			colour-word interference subtest Parent-report and teacher-report - BRIEF			
Andrés-Roqueta, Adrian, Clemente, and Katsos (2013)	93, DLD, 5;4	ELI Pragmatics Subtest (Receptive and Expressive items related to figurative language, politeness (metacognitive)	I – Matching Familiar Figures test	Change of location task Unexpected contents task	CEG (receptive grammar) Sentence recall (expressive grammar) ELI (receptive and expressive vocabulary)	Pragmatic score correlated with ToM (1.28), Inhibition (1.31) grammar-receptive (1.58), grammar-expressive (0.89), vocabulary-receptive (1.28), vocabulary-expressive (1.62) NB Items on ELI test vary in terms of whether they would traditionally be classified as pragmatic.
Angeleri and Airenti, 2014;	100, TD, 4;9	Irony comprehension and joke comprehension (combined for overall humour score)		1st and 2nd order ToM tasks	Italian version of PPVT-R	The overall humor score was significantly correlated with the ToM score (*0.80*), with the PPVT-R score (*1.35*), and with children's age (*1.28*). Analyses looking at each type of pragmatic and ToM task, tended to reve correlations between ToM tasks and irony stories but n correlations between ToM tasks and control stories. Path analyses suggested that language ability affected both humour comprehension and ToM, and ToM had n independent causal effect on humor comprehension.
Bacso and Nilsen, 2017	109, TD, 5;0	Referential communication task (production of unambiguous	WM Digit Span subtest from the WISC 4th		Picture Naming task from the WPPSI-III	Looking at children's initial descriptions of referents, th number of descriptors they produced was correlated wi expressive vocabulary (0.77) working memory (0.85), and cognitive flexibility (0.70) but not inhibition.

For Peer Review Only

		descriptions)	Edition I Red dog-Blue dog task CF Object classification task for children			Once age and vocabulary were partialled out, the correlation held with working memory (0.56) and working cognitive flexibility (0.43) Looking at children's repairs of initially inadequate descriptions, the number of new descriptors they produced was correlated with expressive vocabulary (0.63) working memory (0.43), and cognitive flexibility (0.70) but not inhibition. Once age and vocabulary were partialled out, a correlation held for working cognitive flexibility (0.43)
Banasik, 2013	46, TD, 5;1	Irony comprehension task (force choice touch screen task)		Reflection on thinking test: 9 tasks including tests of: Appearance-reality, 1st- and 2nd order belief, deception, emotion understanding		Recognition of irony did not correlate with children's score on the ToM tests but did correlate with a measure of the quality of justification of their ToM responses (0.68) NB No age effect was found in group of 3-5 year olds.
Bernard and Deleau (2007)	81, TD, seen longitudinally at 3;8, 4;2 and 4;8	Communicative perspective taking measure (collapsing across three dimensions: 1: social status, 2: common ground		1st order ToM (Unexpected transfer task Unexpected contents task Unexpected pictures task)	Language score composite (Receptive vocab) and comprehension of relative clauses) derived from subtest from the	At time 3;8: Communicative perspective taking was correlated with language ability (1.28), and false belief understanding (0.85). At 4;2: Communicative perspective taking was correlated with language ability (1.12), and false belief understanding (0.70). At 4;8

For Peer Review Only

		3. Gricean maxims.)			ISADYLE.	Communicative perspective taking was correlated with language ability (0.85), and false belief understanding (0.47). Three regression analyses were run to explain variance in: CPT at time2 from measures taken at time1; CPT time 3 from measures at time 1 and CPT time 3 from measures at time 2. After controlling for age and CPT at the relevant prior time point, language scores predicted significant additional variance (between 8 and 10%) but FB score did not. Need to bear in mind that FB tasks at time 1 were nearing floor.
Bishop and Norbury (2005)	17, DLD, 6-10 25, SCD, 6-10 14, HFA, 6-10	Principal component from: CCC (teacher/therapist report) pragmatic composite and communication scales of SCQ, ADOS-G	Generativity - Composite of two ideational fluency tasks (Use of Objects and Pattern Meanings)			Generativity was correlated with the pragmatic composite (0.93) Generativity was not correlated with SCQ communication scale but was correlated with the CCC pragmatic composite (0.45) and with the ADOS–G communication scale (1.11). Both the latter correlations remained the same size when age and language were partialled out.
Blain-Brière, Bouchard and Bigras (2014)	70, TD, 4;8	Semi-structured conversation	SC – Prohibited toy protocol WM – Backwards digit span			No relationship between EF and PSCS-P-complexity w strong enough to reach significance (0.49) Talkativeness was negatively correlated with inhibition (-0.58) Responsiveness was positively associated with working memory (0.61)

For Peer Review Only

			CF – DCCS P/I – Tower of Hanoi			Fluidity (speech free of hesitation and repetition), was positively correlated with all EF measures except flexibility (Self-control: 0.63, Inhibition: 0.63, Working memory: 0.52, Planning: 0.65) Measures of self-control, flexibility and planning show very little association with pragmatic function other tha fluidity.
Blom and Boerma (2016)	84 DLD 5;9 45 TD 5;9 (with a follow up at 6;9)	Narrative production and comprehension (Macrostructure) - Multilingual Assessment Instrument for Narratives (LITMUS-MAIN)	WM – Backwards digit span		PPVT III (Dutch Version) TAK sentence repetition task	For the children with DLD, narrative comprehension w associated with vocabulary at both time points (0.63), with sentence repetition at time 1 (0.75) and with working memory at time 1 (0.75). Narrative production was associated with sentence repetition at both time points (0.54) and with working memory at both time points (0.47, 0.82) For the TD children, narrative comprehension was associated with sentence repetition at time 1 (1.03) and with working memory at time 1 (0.82).
Caillies et al. (2014)	15 TD, 9;0 15 ADHD 9;0	Irony comprehension	WM Digit Span subtest from the WISC IV and Sentence Repetition subtest of the French version of the NEPSY I. Auditory Attention and Response Set, and Statue	Two 2nd order false belief tests	Verbal reasoning - Similarities subtest of the WISC-IV	For children with ADHD, 2nd order false belief understanding related to both an 'explanation' question (What did [IRONIST] mean when s/he said X?) (1.42) and to a question about the ironist's belief (1.85). Verbal reasoning was also correlated with both measures; irony explanation (1.91), ironist's belief (2.27), and to an even greater degree than ToM (1.85). EF measures did not correlate with comprehension scores. For the typical group, ToM only related to the question about the ironists belief. Verbal reasoning was not correlated with either measure. Inhibitory control was

For Peer Review Only

			from French version of the NEPSY			correlated with both measures irony explanation (1.58), ironist's belief (2.87), but working memory was not. Neither age nor verbal IQ was controlled for in these analyses. Authors note small sample size.
Capps et al., (1998)	15, ASD, 11;11 15, DD, 9;5	Semi-structured conversation		1st order ToM (Smarties task Sally-Anne task)	CELF (language age)	For the ASD group, the amount of contingent, relevant and new information they provided was correlated with ToM scores (1.5). However, when language age was accounted for, the association with ToM was no longer significant. For the developmentally delayed control group, langua age was correlated with contingent, relevant, new info, (2.14) but ToM was not.
Capps, Losh and Thurber, (2000)	13, ASD, 12.6 13, DD, 9.8 13, TD, 6.0	Narratives elicited using wordless picture book		1st order ToM (Smarties task, puppet and interactive versions of the Sally-Anne task)	CELF language age	For the ASD group, scores of ToM tasks were correlate with the following narrative qualities: syntactic diversit (2.41), evaluative statements (1.35), evaluative diversit (1.85), mental state terms (2.49) and negatively correlated with affective state terms (-2.49). Theory of mind continued to be marginally correlated with the mental state terms (1.28), and affective state terms, (-1.25) when language ability was controlled for For the developmentally delayed group, there were no significant correlations between scores of ToM tasks ar narrative qualities.
Dahlgren and Dahlgren Sandberg (2008)	30, ASD, 10;1 30, TD, 9;6	Referential communication task (describing cards that listener could not see)		1t order ToM (ToM1) 2nd order ToM (ToM2)	Verbal IQ	For children with ASD, verbal IQ was related to the number of relevant features mentioned and referential efficiency in a sample of 30 7- to 14-year-olds with AS For children with ASD, ToM1 was correlated with both the number of relevant features mentioned when describing a target (1.10) and the efficiency with which these (and not irrelevant features) were mentioned (1.0

For Peer Review Only

						However, these measures were also associated with verbal IQ so the specificity of association is unclear. N.B., ToM1 was a binary measure (pass / fail), only fiv of the thirty children with ASD failed 1st-order ToM. No associations with 2nd order ToM were observed for TD children or those with ASD (possibly due to lack of variance).
Davies, Andrés-Roqueta, and Norbury (2016)	18, DLD, 5-10;11 18, TD, 5-10;11	Reference production task Comprehension and judgement task (describing referents a listener cannot see and selecting referents based on visual and discourse context)			Receptive and expressive grammar (sentence recall) subtests from the ELI. Vocabulary subtest from the WISC-IV	Across both groups there were significant correlations between production of optimal utterances in the contras condition and all formal language measures (combined effect size 0.93) Across both groups, performance on the judgement tasl significantly correlated with receptive grammar and vocabulary (combined effect size 2.27) Correlations are also reported by group For children with DLD, there were correlations betwee production of optimal utterances on a referential communication task and receptive grammar: (1.49) and vocabulary (1.01) but not sentence recall (-0.12) For TD children there was an association with receptive grammar only (2.49)
De Marchena and Eigsti, 2016	18, ASD, 12;7-16;11 18, TD, 12;2-17;11	Narrative task – private condition and shared condition yielding an adaptation to common ground measure.			PPVT	Receptive vocabulary was not found to correlate with th common ground measure, within or across groups. NB Authors note that participants were selected to be in normal range for vocabulary. When the groups were collapsed, scores on the Social Responsiveness Scale (SRS) were negatively correlated with the common ground measure.
De Rosnay, Fink, Begeer,	129, TD, 6;6	Mindful Conversation		Ten-item ToM battery	PPVT-III	PPVT-III and MCC were significantly correlated (0.63) Total ToM was significantly correlated with MCC (0.8

For Peer Review Only

Slaughter, and Peterson (2014)		Competence		consisting of standard false-belief tasks		and this relationship remained after age and PPVT-III and shyness were controlled for (0.54) NB Some questions on the MCC are directly tap ToM (e.g., 'Does the child have difficulty understanding oth people's thoughts?').
Farrant, Maybery, and Fletcher (2010)	Study 1: 99, TD, 5;4 Study 2: 93 TD 5;1 and 30 DLD 5;3	Conversation Skill Rating Scale		Socio-emotional engagement, joint attention and imitation were measured retrospectively using scales developed for this project		Study 1 – CSRS was significantly related to socio-emotional engagement (0.93), joint attention (1.42) and imitation (0.95) Study 2 – CSRS was significantly correlated with socio emotional engagement (0.70), joint attention (0.80) and imitation (0.95). NB Reliability of retrospective reports is unclear.
Fernández (2013)	115, TD, 4;8- 8;8	Narrative production coded for evaluation, accurate cohesion, psychological cohesion and coherence (plus combined pragmatic score)		ToM1 ToM2 Messy room story Faux pas story	Spanish adaptation of the PPVT-R	Pragmatic language scores were correlated with numbe of utterances (0.89) and number of clauses (1.62) TVIP scores were only significantly correlated with narrative coherence (0.54) There was no correlation between any of four measures of narrative quality and 1st order ToM scores (perhaps due to a ceiling effect on the latter) There was a correlation with 2nd order ToM and coherence scores (0.47) A regression model predicting overall pragmatic score found, after entering gender and length of narrative, 2n order ToM explained an additional 5% of variance
Filippova and Astington (2008)	24, TD, 5;8 24, TD, 7;7 24, TD, 9;7	Irony comprehension	WM – Numbers subtest of the children's memory scale	Combined: 2nd order false belief stories, Strange stories,	Receptive vocabulary PPVT-III	Irony was significantly positively correlated with 2nd order ToM (1.19), PPVT-III (1.42), WM (0.63) and prosodic understanding (0.75). Once age, vocabulary, and digit-span were controlled for, a marginally significant association between ToM

For Peer Review Only

				Faux pas stories		and Irony scores remained (0.45) Regression models showed that once age, memory, attunement to prosody and receptive vocabulary were controlled for, ToM was a significant predictor of irony comprehension When all other variables are controlled for, vocabulary also explained unique variance
Fukumura (2016)	20, ASD, 8;10 20, TD, 8;4 (plus additional study with adolescents)	Referential communication task			BPVS; Expressive vocabulary sub-test of the WASI	For both the ASD group, neither receptive nor expressi vocabulary was correlated with the tendency to use mo adjectives in the shared rather than privileged context, with the number of egocentric adjectives in the privileged context. There were large correlations between the number of adjectives produced in the shared context and both receptive *(1.46)* and expressive *(d=0.98)* vocabulary. In the TD group, adjective use was not correlated with either vocabulary measure.
Gillespie-Lynch et al., (2015)	23, TD, 31 TD sibs at high risk ASD, 10 high risk sibs who went on to have ASD, 7;5	CCC-2 Pragmatic composite (parent report)		ESCS		Groups analysed separately. No association between joint attention variables at 12 months or 18 months and school age pragmatic language for children with ASD for TD children (collected approximately 6 years later) NB Behaviours observed during infancy occurred with low frequency.
Gillis and Nilsen (2014)	40, TD 5;0 36, TD, 7;1	Referential communication – Ambiguity detection task	CF – Object classification task for children		Picture vocabulary subtest of the TOLD	Significant negative correlation between scores on the CF task and ratings of how helpful the ambiguous clues were for 5-year-olds (-1.15, i.e, the better the CF, the more likely they were to say ambiguity was unhelpful). This

For Peer Review Only

						relationship remained after controlling for age and receptive language. No significant correlations were found for 7-year-olds. Significant negative correlation between TOLD (forma language) and ambiguity detection for both groups (5 years -0.77; 7 years –0.72).
Godbee and Porter (2013)	26, WS, 18;3 26, TDCA (Chronological Age matched) 18;0, 26, TDMA (Mental Age matched), 5;9	Non-literal speech stories (sarcasm, metaphor, simile comprehension)	Verbal working memory (WJ Revised)		Expressive vocabulary (WJ Revised)	For TDCA and TDMA controls combined, each of the cognitive measures assessed using the WJ-R COG (including expressive vocabulary, verbal working memory, perceptual integration, inferential reasoning a overall cognitive ability) was significantly and positive correlated with each of the measures of non-literal language comprehension (no individual results reporte by all effect sizes d >1.39). For the WS group, measures of sarcasm comprehensio were not correlated with any cognitive measures. Metaphor comprehension was correlated with verbal working memory (0.84) but not expressive vocabulary. Simile comprehension was correlated with verbal working memory (0.88) but not expressive vocabulary. NB small sample size and large age range (early childhood to late adulthood)
Hale and Tager-Flusberg (2005)	57, ASD, 4-13;11	Natural language with parent during free play coded for topic contingent discourse		Ten-item ToM battery in 3 sequenced batteries covering: Desire and a pretense,	Combined: Receptive vocabulary PPVT-III and Expressive Vocabulary test	Time 1 – significant correlation between contingent discourse and vocab (0.87) and ToM (1.12). Time 2 – significant correlation between contingent discourse and vocab (0.95) and ToM (1.31) At both time points, regression analysis confirmed that

For Peer Review Only

				perception/knowledge, 1st and 2nd order false belief, lies jokes, moral judgement.		ToM explained additional variance (8%) once age, IQ and vocab score were controlled for. Time 1 vocabulary was the only significant longitudina predictor of time 2 contingent discourse.
Huang, Oi and Taguchi (2015)	50, HFASD, 10;2 50, TD, 10;7	40 figurative language tasks (comprehension of metaphor, irony, sarcasm, indirect reproach and indirect request)		ToM Battery – 1st and 2nd order ToM	PPVT-R, Verbal IQ (WISC III)	For TD children, receptive vocabulary was correlated with metaphor comprehension (0.62) but not with any other measures of figurative language. For HFASD children, receptive vocabulary was significantly correlated with metaphor comprehension (1.13) as was verbal IQ. No correlations with other measures of figurative language were observed. No correlations with ToM measures were reported (instead children were classified into one of 3 groups fo this measure)
Ketelaars, Jansonius, Cuperus, and Verhoeven (2012)	77, SCD, 5;6 77, TD, 5;6	Dutch adaptation of the Renfrew bus story test. Narratives coded for: productivity (length), organization (relevant content), and cohesion (use of cohesive devices)	Principal component of: Planning, inhibition, cognitive flexibility (Tower task Auditory attention and response set from NEPSY) and working memory (number recall from Kaufman Assessment	1st order ToM: Three change of location tasks	Composite of: Receptive and sentence comprehension (from Dutch Language test for children) Expressive vocabulary (from Dutch Renfrew Word Finding Vocabulary Test)	For TD children, narrative productivity (but not conten or cohesion) was positively correlated with EF (0.63) a ToM (0.85). Once language was controlled for, only ToM was predictive of narrative productivity. For SCD children (who tended to convey less plot content, often omitting initiating events and the story outcome) EF was correlated with narrative productivity (0.79) and organisation (0.45). Once language was controlled for, EF was predictive o SCD children's narrative productivity (explaining an additional 9% of variance)

For Peer Review Only

			Battery for Children)			
Kuijper, Hartman and Hendriks (2015)	46, ASD 37, ADHD 38, TD	Narrative production: Referent maintenance and to reintroduction	I – stop signal task WM – n-back task	1st and 2nd order ToM	Vocabulary from Dutch WISC-III and Dutch PPVT-III (not analysed)	None of the cognitive predictors explained variance in the rate at which children appropriately maintained referenced. In a series of mixed effects logistic regression models, cognitive variables analysed (1st and 2nd order ToM, W and I) were found to be associated with appropriate referent reintroduction. In a multivariate model with all four cognitive predicto considered simultaneously (fitted to data from all three groups), reintroduction was predicted by 2nd order ToM and working memory. Note, formal language was not controlled in these analyses.
Leonard et al (2011)	54, TD, 10;6	CCC-2 composite of scales E-J (parent ratings)			OWLS. Combined Oral Expression and Listening Comprehension. Verbal IQ (KBIT-2)	No correlations between CCC2 and either OWLS or verbal IQ. NB This study includes CCC2 subscales I and J, which cover social relations and interests (i.e., non pragmatic features of autism). Not clear whether a correlation would be expected.
Losh and Capps (2003)	28, high-functioning ASD, 11.3 22, TD, 10.6	Semi-structured conversational storytelling, storybook narratives (coded for length,		Happé's strange stories	Verbal IQ (WISC-III)	Narrative measures were not correlated with Verbal IQ or ToM, although associations with emotion understanding were observed.

For Peer Review Only

		grammatical complexity, evaluation and structure)				
Losh, Martin, Klusek, Hogan-Brown, and Sideris (2012)	28, ASD, 9;2, 40 FXS-ASD, 10;7 21, FXS only, 9;7 21, DS, 10;10 20, TD, 4;10 (All groups male)	Pragmatic Judgement subtest of CASL CCC-2 (teacher ratings)		Basic battery: intentionality, understanding of desires, false belief (reduced verbal load), appearance-reality and perspective taking. Advanced battery: perspective taking, diverse desires, diverse belief, false belief and knowledge access		Correlations between ToM and performance on the CASL for ASD (1.35), FXS (0.77), DS (1.19) and TD (1.28) Correlations between ToM and scores on the CCC-2 initiation subscale for ASD (1.35) and coherence subscale for FXS (0.77).
Maridaki-Kassotaki, and Antonopoulou (2011)	76, TD, 5;6	Listening Skills Test. Test of Referential Communication		Composite of: Unexpected transfer test Deceptive box test Deceptive object test		False belief was correlated with two out of four components of the LIST: referent identification (0.47) and comprehension of directions (0.61) It was also correlated with one out of the three components of the TREC: the ability to detect ambiguit in oral messages and respond adequately as listener (0.66). It did not correlate with the component 'adequa message to speaker'.

For Peer Review Only

						NB Many items on this LIST could be classified as calling on semantic knowledge whereas the TREC requires children to unambiguously describe a target in an array of distractors (and to do the receptive equivalent)
Massaro et al., (2014)	34, TD, 5;8 36, TD, 7;3	Irony comprehension: socially shared (SS) irony and situationally defined (SD) irony		1st order (unexpected transfer) and 2nd order FB tasks ToM	Metacognitive vocabulary test; Italian standardisation of PPVT-R	Collapsing across age groups and partialling out age, there was a significant correlation between both irony tasks and both vocabulary tasks and some correlation with FB1: SS irony: correlated with PPVT (0.56); and MVT (0.47 and FB1 (0.54) SD irony: correlated with PPVT (0.74); and MVT (0.5 Regression analyses by age revealed nothing was predictive of 5 year olds' comprehension. For 7 year olds, only vocabulary was a predictor of SS irony. Vocabulary and MVT were predictors of SD irony. NB: Only one item tapping 2nd order ToM – lack of variance could explain some null findings.
McEvoy, Rogers, and Pennington (1993)	17, ASD, 5;1 13, DD, 4;2 16, TD, 3;2	ESCS: Joint Attention, Social Interaction and Behaviour regulation scales.	I – AB error task I – Delayed response task CF – Spatial reversal task – CF – Alternation task		Verbal tasks from the BSID Picture naming and identification and verbal tasks from the Stanford-Binet Intelligence Scale	Correlation analyses run for three groups combined for children contributing data to the spatial reversal task (focused on because most sensitive to group difference Perseverative errors on the spatial reversal task were correlated with the Joint Attention scale of ESCS and (-1.25) and the Social interaction scale (-0.98) Verbal ability was correlated with the Social interactio scale (-0.82) Regression analysis controlling for group membership

For Peer Review Only

						and verbal ability revealed that joint attention was a significant predictor of spatial reversal. NB The ESCS largely assesses non-verbal communication and some scales could equally be categorised as a measure of mentalising.
Mewhort-Buist and Nilsen, 2013	88, TD, 9;10	Irony comprehension (understanding speaker belief and intent for criticism and compliments)		2nd order false belief task	Picture Vocabulary subtest of the TOLD-I:4	When a speaker made an ironic criticism, receptive vocabulary skills were correlated with understanding th speaker's belief (0.37) and their intent (0.68). ToM skills were not correlated with understanding (although the measure of understanding speaker belief was almos at ceiling). When a speaker made an ironic compliment, vocabular skills were significantly associated with understanding the speaker's belief (0.37) but not their intent. ToM wa correlated with understanding the speaker's belief (.36) but not their intent. NB Large number of measures correlated, with some towards ceiling.
Miniscalco, Rudling, Rastam, Gillberg, and Johnels (2014)	34, ASD, 3;5 and 4;6	Swedish version of CDI:WS Pragmatics scale from Swedish CDI:WS			CDI words and gestures (CDI: WG) CDI:WS	Partial correlations (accounting for age): Pragmatics scales at time 1 and 2 were significantly positively correlated with vocabulary (time 1 (1.76); tir 2 (1.25)) and grammar (time 1 (2.20); time 2 (1.07)). In a regression analysis predicting time 2 pragmatic score, with age, time elapsed since time 1 and time 1 pragmatic score as control, vocab and grammar did not predict pragmatic language outcomes (but the 'imitatin

For Peer Review Only

						adults score did). NB The pragmatic scale assessed whether children use language to talk about the past or future and whether they engage in pretend play.
Nadig, Vivanti and Ozonoff (2009)	17, HFA, 11;3 17, TD, 10;8	Referential communication tasks			CELF 4	Level 1 adaptation (when both speaker and addresses c see all the objects, providing sufficient disambiguating information to select one) was correlated with formal language level for the TD group (1.07) but not the HFA group. Level 2 adaptation (when an addressee cannot see all th objects, only using descriptions that make sense from their point of view) was not correlated with language level for the TD group but it was for the HFA group (0.85). Level 3 adaptation (efficiently providing indirect clues hidden object identity) was correlated with language level for both the TD (marginally significant correlatio and the HFA group.
Nicholson et al., 2013)	31, TD, 9;2	Comprehension of ironic criticisms (overt forced choice response and eye movement measures) CCC-2		Empathy Quotient (EQ-C)	CCC-2 (subscales A-D)	No measure of irony comprehension was correlated wit the CCC2 structural language or pragmatic subscales. The structural language subscales of the CCC2 were correlated with the pragmatic subscales (1.67) Comprehension choice scores were correlated with empathy quotient (0.95) and some eye-movement

For Peer Review Only

		(subscales E-H)				measures. NB large number of measures correlated in analyses.
Nilsen and Graham, (2012)	34, TD, 4;1 (then longitudinally followed to 4;6 and 5;0)	Referential communication – Task assessing object choice, looking time, and message evaluation.	I – Day-night I – Grass-snow		PPVT III	Inhibition was not related to looking time or object choice. Inhibition was not related to message evaluations withi the same assessment period For the knowledgeable-ambiguous condition, inhibition at 4 years was correlated message evaluation both at 4;6 (1.35) and 5;0 (0.95) when verbal skills wer controlled for. Inhibition was not correlated with performance in any o the other conditions Vocabulary skills were not significantly related to message evaluations at any age
Nilsen and Graham, (2009) – experiment one	60, TD, 5;0	Referential communication task with objects in/outside visual common ground. Production and comprehension (object choice and looking time) measures	WM – Backwards digit span WM – Memory for objects taken from WISC-III I – Red dog/blue dog I – Tapping task CF – Flexible item selection		PPVT III	The production measure was only correlated with memory for objects (0.61) and this did not hold when a and verbal skills were controlled for. All three measures of comprehension (of egocentric interpretation) were significantly negatively correlated with performance on inhibition tasks but not other measures of EF (Red/blue dog; -0.65, Tapping task; -0.52). After controlling for age and verbal skills, correlations remained between red/blue dog task and both looking time and choice of referential alternative (0.52) and between the tapping task and looking time.

For Peer Review Only

						PPVT was correlated with how often two objects were chosen in privileged ground condition (-0.54)
Nilsen and Graham (2009) – experiment two	47, TD, 3;10	Referential communication task with objects in/outside visual common ground. Comprehension measures of egocentrism (object choice and looking time measures)	I – Bear/dragon task (conflict inhibition) I – Gift delay task (delay inhibition)		PPVT	PPVT was correlated with how often a referential alternative was picked in privileged ground condition -.77 but not other measures of egocentric comprehension (although association was in same direction). The bear/dragon task was correlated with duration of egocentric eye gaze (-0.98) and remained so when verb skills and age were controlled for. It was not correlated with the object choice measures of comprehension. The gift delay task was significantly correlated with the choice comprehension measures (-0.71), however only the relationship between gift delay and both objects chosen remained significant when age and verbal skills were controlled for (-0.68). There was no correlation between the gift delay task and duration of eye gaze In a regression model, children's performance on the conflict inhibition task was the only variable that accounted for unique variance in the looking time measure (18%).
Nilsen, Glenwright and Huyder (2011)	53, TD, 8;8	Irony comprehension – forced choice measure		2nd order false belief tasks	Picture Vocabulary subtest of the TOLD-Primary 3	In this task children heard a story about two protagonis a speaker who made a sarcastic remark, and a listener who was either in a position to understand it was sarcastic or not. A measure of children's sensitivity to the listener's knowledge state affecting their understanding of the speaker's intentions was significantly correlated with 2nd order ToM (0.70) but not with their understanding of the listener's interpretation of humorous intent. Receptive vocabular was not correlated with either measure although it was

For Peer Review Only

						correlated with a measure of understanding what the listener would believe to be the true state of affairs (0.56).
Norbury, Gemmel, and Paul (2014)	22, DLD, 6;7 – 15;4 26, ASD, 6;6 – 15;9 27, TD, 6;9 – 15;2	Narrative elicitation task, wordless picture book coded for internal state language, relevant content, pragmatic errors, and macrostructure			CELF-4, Verbal IQ (WISC or BPVS)	In the DLD group, language ability was not correlated with pragmatic errors. In the ASD group language ability was negatively correlated with pragmatic errors (1.22) However, it was also negatively correlated with relevant propositions (-1.35), suggesting more verbally able children may be more verbose but not in an adaptive way.
O'Reilly, Peterson and Wellman (2014)	10, native signing deaf, 9;0, 32, deaf of hearing parents (late signers), 9;3 39 hearing children, 8;8	Comprehension of sarcasm		1st and 2nd order false belief	Syntax subtest of CELF-Preschool translated into Auslan or NSS (no test administered for hearing children)	For all children combined, the ToM measure was correlated with sarcasm measure (1.12) For the deaf children only, this correlation was re-run partialling out age and language ability and an correlation between ToM and sarcasm comprehension remained (.90)

For Peer Review Only

Pellicano (2013)	37, ASD, 6;7 at time 1, follow up 3 years later (time 2)	ADOS-G (higher scores reflect greater difficulty with social communication)	Aggregate of: P – Tower of London task CF – Teddy-bear set-shifting task IC – Luria's hand-game P – Mazes task from the Wechsler Preschool and Primary Scales of Intelligence –Revised	Sum of: 1st-order unexpected contents task, 1st-order unexpected location task, 2nd order unexpected location task	PPVT-III	Individual differences in children's verbal ability (-0.77 ToM (-0.79) and EF (-1.15) scores at time 1 were significantly and negatively related to ADOS-G scores time 2 (whereas time 1 age, non-verbal ability and central coherence bore no such relation). When age, verbal ability, and non-verbal ability at time were partialled out, only early differences in EF scores remained related to later ADOS-G scores (0.21) In a regression analysis, only EF (not ToM or Central Coherence) predicted unique variance (16%) above and beyond age verbal and non-verbal ability.
Pesco and O'Neill (2012)	348 children between 18 and 47 months when LUI collected and mean age 5;8 when language outcomes collected. Oversampling to ensure representation of	LUI parent completed (language total score)			DELV-NR (total language composite score), The CELF-P2 (core language score); CCC2 (language composite)	For children with LUI total score collected between 12 and 24 months (N= 112), there was a correlation with t DELV-NR language composite at 5;8 (0.4) but not wit the other two language measures. For children with LUI total score collected between 24 and 29 months (N= 94), there was a correlation with all three language outcome measures at 5;8 (d between .6 and .73) For children with LUI total score collected between 30 and 35 months (N= 67), there was a correlation with all three language outcome measures at 5;8 (average d = 1.21) For children with LUI total score collected between 36 and 41 months (N= 32), there was a correlation with all three language outcome measures at 5;8 (d between .7 and 1.4). For children with LUI total score collected between 42

For Peer Review Only

	children with weak pragmatic skills on LUI					and 47 months (N= 43), there was a correlation with all three language outcome measures at 5;8 (d between .3 and .67) Thus the LUIs predictive validity changed with age and peaked approximately when children turn 3 years of ag
Pexman et al. (2009)	118 older TD, 9;9, and 118 younger TD 6;9	Production of irony during naturalistic play with two other family members			PPVT	No correlations observed between receptive vocabulary and irony production. NB broad definition of irony used.
Resches and Pereira (2007)	74, TD, 4;6	Referential communication task – treasure hunt where director child helps another who was not previously party to information about location of treasure		ToM – 2 tasks assessing knowledge-ignorance and 1st order false belief		ToM was correlated with the number of accurate descriptions children produced (1.96), an association that remained once age was controlled for (0.87). ToM was also negatively correlated with the number of ambiguous descriptions produced (-1.58) but once age was controlled for this association did not hold. NB pragmatic ability was a property of dyads, large number of correlations run.
Rinaldi, Baruffaldi, Burdo, and Caselli (2013)	23, deaf, 1;7-3;0 (correlations N = 11)	Italian version of the Social Conversational Skills Rating Scale. Parent report with assertiveness			Short form of the Italian version of the MacArthur-Bates CDI Words and Sentences	Significant correlation between word production and th pragmatic subscales assertiveness (1.67) and responsiveness (2.02)

For Peer Review Only

		and responsiveness scales			(expressive vocabulary analysed)	
Rints, McAuley, and Nilsen (2015)	36, TD, 3;7	CASL – Pragmatic Judgement subtest LUI (part 3)	I – Statue subtest of the NEPSY-II SWAN parent rating scale: total scores reflecting inattention and hyperactivity-impulsivity			Controlling for age, children who made more movemer errors on the Statue task also obtained lower scores on the CASL (-1.12) Children who were rated as more inattentive or hyperactive-impulsive by their parents were also rated having poorer pragmatic skills on the LUI (.80)
Slomkowski and Dunn (1996)	36 TD (3;4 for ToM assessment 3;11 for communication)	Naturalistic conversation with a peer coded for connectedness		1st order ToM (change of location tasks)		False belief scores were correlated with mean length of connected episode (.84), mean length of play episode (.93) and mean length of pretend episode (.70)
Tager-Flusberg and Sullivan (1995)	27, ASD, 16;8 27 DD, 12;6	Narrative production based on wordless picture book		1st order ToM (4 false belief tasks)	PPVT CELF Sentence Structure and Formulated Sentences	No correlations were observed for the Developmentally Delayed group. For the ASD group, ToM scores were correlated with narrative measures including number of propositions (1.12), and number of connectives (1.03). (NB no clea division between formal and pragmatic narrative measures as this was not the focus of the study)
Volden et al., (2009)	37, ASD, 8;6	TOPL VABS			CELF-3	70% of variance in TOPL was explained by a model including CELF expressive, CELF receptive and a non-

For Peer Review Only

		(Communication and Socialization scales) ADOS communication				verbal cognitive score. Only the CELF measures were significant predictors. So combined formal language scores explain majority of variance in TOPL. CELF expressive, CELF receptive and TOPL account f 30% of the overall variance in VABS Communication Scale. But TOPL did not explain unique variance. No predictive value of any measures for the Socialization scale. Only TOPL scores predicted ADOS communication.
Whyte and Nelson (2015)	26, ASD, 9;8 69, TD, 8;10	CASL (Pragmatic Judgment subtest and Nonliteral Language subtest)		Children's version of the reading the mind in the eyes task (Baron-Cohen et al., 2001)	Syntax construction subtest of the CASL Receptive and expressive vocabulary from Verbal IQ subtests from the Kaufman Brief Intelligence Test, 2nd edition	Pragmatic judgment scores were correlated with syntax scores both for TD children (2.87) and children with ASD (2.97) Pragmatic judgment scores were correlated with vocabulary scores both for TD children (3.22) and children with ASD (2.49) Pragmatic judgment scores were correlated with ToM scores both for TD children (1.39) and children with ASD (1.54) When controlling for vocabulary and syntax, ToM abilities were correlated with pragmatic judgment score for TD children (0.52) but not for children with ASD. Nonliteral language abilities were correlated with synta age for TD children (2.98), and for children with ASD (2.27). Nonliteral language abilities were correlated with vocabulary age for TD children (2.87), and for children with ASD (2.76). Nonliteral language abilities were correlated with TOM for TD children (1.76), and for children with ASD (1.3 These correlations held when vocabulary and syntax were controlled for (TD 0.93, ASD 0.80)

For Peer Review Only

Note: Acronyms used in the table are as follows:

Population:
TD = Typically Developing, ASD = Autism Spectrum Disorder, DLD = Developmental Language Disorder, SCD= Social Communication Disorder, FXS Fragile X Syndrome, WM = working memory, O = organization, CF = cognitive flexibility, I = inhibition, SC = self-control, P = planning

Pragmatic tests:
ADOS-G = Autism Diagnostic Observation Schedule – General (Lord et al., 2000)
CASL = Comprehensive Assessment of Spoken Language, (Carrow-Woolfolk, 1999)
CCC, CCC2 = Children's Communication Checklist 2 (Bishop, 2003)
ELI = Evaluacion del Lenguaje Infantil (Saborit & Julian, 2005)
LITMUS-MAIN = Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012)
LUI = Language Use Inventory (O'Neill, 2009)
MCDS = Mindful Conversational Difficulties Scale (Peterson, Garnett Kelly & Attwood, 2009)
SCQ = Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003)
TOPL - Test of Pragmatic Language (Phelps-Terasaki & Phelps-Gunn, 1992)
VABS - Vineland Adaptive Behavior Scales (Sparrow, Balla, Cicchetti, 1997)

Executive function tests:
BRIEF = Behavior Rating Inventory of Executive Functioning (Gioia, Isquith, Guy, & Kenworthy, 2000)
DCCS = The Dimensional Change Card Sort (Zelazo, 2006)
D-KEFS = Delis-Kaplan Eexecutive Function System (Delis et al., 2001)
NEPSY-II = Neuro-Psychological Assessment of Children (Korkman, Kirk, & Kemp, 2007)
SWAN = Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior rating scale. (Swanson, n.d.)
WISC-IV = Wechsler Intelligence Scale for Children (Wechsler, 2003)
WJ Revised = Woodcock–Johnson (Revised) Tests of Cognitive Ability (Woodcock, 1989)

Theory of Mind tests:
ESCS Early Social Communication Scale (Mundy et al., 2003)

Formal language tests:

For Peer Review Only

BPVS = The British picture vocabulary scale (Dunn & Dunn, 2009)
BSID Bayley Scales of Infant Development (Bayley, 1969)
CDI (WG, WS) = The MacArthur Bates Communicative Development Inventories (Words & Gestures, Words & Sentences. Fenson et al., 1996)
CEG = Comprension de Estructuras Gramaticales (Mendoza et al. 2005)
CELF 4, CEFL Preschool 2 = Clinical Evaluation of Language Fundamentals (Semel, Wiig, Secord, & Langdon, 2006).
DELV-NR = Diagnostic Evaluation of Language Variation - Norm-Referenced (Seymour, Roeper, & de Villiers, 2005)
ELI = Evaluacion del Lenguaje Infantil (Saborit and Julian 2005)
ISADYLE = Instruments pour le Screening et l'Evaluation Approfondie
des Dysfonctionnements du Langage chez l'Enfant (Piérart, Comblain, Gregoire, & Mousty, 2009)
KBIT-2 = Kaufman Brief Intelligence Test -2 (Kaufman & Kaufman, 2004)
OWLS = Oral and Written Language Scales (Carrow-Woolfolk, 1995)
PPVT (R, III) = Peabody Picture Vocabulary Test (multiple editions: Dunn & Dunn, 1981, 2007)
TAK = Taaltoets Alle Kinderen Verhoeven & Vermeer, 2001)
TOLD = Test of Language Development Primary 3rd Edition (Newcomer & Hammill, 1997)
WASI = Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999)
WPPSI = Wechsler Preschool and Primary Scale of Intelligence (multiple editions: Wechsler, Scales & Index, 2012)

For Peer Review Only