



City Research Online

City St George's, University of London

Citation: Harvey, A., Spicer-Cain, H., Botting, N. & Henry, L. (2025). Narrative abilities of autistic and non-autistic adolescents: The role of Mentalising and Executive Function. *Autism Research*, 18(1), pp. 152-165. doi: 10.1002/aur.3272

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/34023/>

Link to published version: <https://doi.org/10.1002/aur.3272>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Narrative abilities of autistic and non-autistic adolescents:

The role of Mentalising and Executive Function

[Running title: **NARRATIVE SKILLS IN AUTISTIC ADOLESCENTS**]

Anna Harvey, Helen Spicer-Cain, Nicola Botting, Lucy Henry

Department of Language and Communication Science,

City St George's, University of London

Northampton Square, EC1V 0HB

Corresponding author: Dr Anna Harvey (anna.harvey.3@city.ac.uk)

ORCID ID: 0000-0001-9984-1595

Co-authors:

Dr Helen Spicer-Cain (helen.spicer-cain.2@city.ac.uk)

Prof Nicola Botting (nicola.botting.1@city.ac.uk)

Prof Lucy Henry (lucy.henry.1@city.ac.uk)

Acknowledgements: Not applicable.

Lay summary:

Previous research suggests that autistic young people often experience challenges with spoken narrative skills, and that this may be linked to cognitive differences in ‘mentalising’ (inferring what others are thinking or feeling) and high-level thinking skills (executive function, EF). This study investigated how mentalising and EF contributed to the structure and coherence of narratives produced by a large sample of autistic and non-autistic adolescents with similar cognitive and language levels. Mentalising ability consistently predicted both narrative structure and coherence, whether participants were autistic or not, but EF did not predict either aspect of narrative performance in this sample.

Abstract

Spoken narrative skills are important for adolescents in their everyday lives. Previous research suggests that producing well-structured and coherent narratives may be challenging for autistic young people. Mentalising, also known as ‘advanced Theory of Mind’ (ToM) and ‘Executive Function’ (EF) are two cognitive abilities frequently explored in relation to autism, both of which may be implicated in narrative ability. The present study investigated these relationships in a group of autistic adolescents (N=44) aged 11-15 years and a comparable non-autistic group (N=54) that did not significantly differ on age, sex, non-verbal cognitive ability, or receptive/expressive language skills. Participants were assessed on a video-based spoken narrative task, scored for both overall structure (‘story grammar’) and narrative coherence. A battery of tasks measuring Mentalising and EF (working memory, inhibition, shifting, generativity) was also administered.

Relationships between scores on cognitive measures and narrative performance were investigated using hierarchical linear regression analyses. Mentalising scores were found to significantly predict narrative performance across all outcome measures and were a stronger predictor than diagnostic group. Diagnostic group predicted narrative structure ('story grammar') scores but not coherence scores. EF scores were not predictive of narrative ability in this sample. Mentalising skills appear to play an important role for both autistic and non-autistic adolescents in the generation of narrative structure and coherence within spoken accounts.

Keywords: Adolescents, Executive Functioning, Social Cognition & Theory of Mind, Language

Background

Narrative, or storytelling, is a sophisticated form of discourse drawing on a range of linguistic, cognitive and pragmatic abilities (Norbury et al., 2014). Narrative skills are fundamental to human communication, social development and learning, and are essential for both social and academic success (Petersen et al., 2008). Autism is a neurodevelopmental difference with an estimated prevalence of 1.57% amongst school-aged children in the UK (Baron-Cohen et al., 2017). Autism is known to impact on young people's communication abilities (American Psychiatric Association, 2013), particularly in terms of pragmatics, that is, the effective use of language in social contexts (Naigles & Chin, 2015). Previous research into the narrative skills of autistic children and adolescents has resulted in conflicting findings across various experimental tasks (Baixauli et al.,

2016), with differences in the group matching strategies used by research groups further complicating the picture. However, there is evidence to suggest that autistic young people tend to find generating well-structured and coherent verbal accounts more challenging than non-autistic young people (Baixauli et al., 2016; Conlon et al., 2019; Hewitt et al., 2019).

Structure and coherence are two closely related, yet conceptually distinct, aspects of narrative production (Harvey et al., 2023). The overall structure (macrostructure) of a narrative account reflects how the content of the story is organised. 'Story grammar' (Stein & Glenn, 1979) is a common approach to macrostructural analysis that considers how narratives are structured around key story elements, such as setting, initiating event, plan, action, consequence, internal response and resolution. Whilst coherent narratives must have a logical story structure, the concept of narrative coherence was also considered in the present study because it encompasses the broader ways in which narrators create holistic and meaningful accounts. For example, using accurate reference chains when describing the actions of individual characters, so that the listener can easily follow the story, or leaving out irrelevant or incongruous information, which could confuse the audience (Harvey et al., 2023).

Spoken narrative abilities are important for adolescents across many different contexts; for example, sharing experiences with their peers, demonstrating their learning in the classroom, or telling someone about their day (Petersen et al., 2014). To better support the functional communication skills of autistic young people, it is of interest to consider some of the underlying cognitive factors that may be implicated in the production of well-structured and coherent narratives. Mentalising, also known as 'advanced Theory of Mind'

(ToM) and Executive Function are two cognitive abilities that have been extensively researched in relation to autism (Rajendran & Mitchell, 2007), although the relationship between these constructs remains unclear (Hill, 2004). Mentalising refers to the cognitive ability "to attribute mental states to another person and to infer their underlying intentions, thoughts, emotions and motivation" (Colle et al., 2008, p.28). The term 'advanced ToM' is also sometimes used to distinguish these more nuanced 'mind-reading' abilities from the classic first and second order 'false belief' tasks used in earlier ToM research (White et al., 2009). A key feature of narrative is the ability to provide information to the audience about the motivation, emotional responses and psychological states of the characters (Mar, 2004). Children and adolescents who find it challenging to infer the mental states of others may struggle to reflect these aspects in their spoken narratives (Capps et al., 2000). In pragmatic terms, such narrators might also have difficulty taking the needs of their listener into account and anticipating which elements of background information are necessary for them to make sense of the story (Loveland et al., 1990).

There is considerable evidence that autistic individuals respond differently to neurotypical people in mentalising tasks, across the developmental trajectory (e.g., Baron-Cohen et al., 1997; Kaland et al., 2008; White et al., 2009). However, the assertion that mentalising may provide a general cognitive explanation for autism attracts increasing criticism on both empirical and epistemological grounds (Gernsbacher & Yergeau, 2019), with many researchers rejecting the 'deficit' model of autism (Astle & Fletcher-Watson, 2020). Recent research into mentalising has also called into question the notion that difficulties or differences in discerning the mental states of others are uniquely associated with autism. Cotter et al. (2018) carried out a systematic review of meta-analyses that

examined social cognition, including mentalising, in different clinical populations. These authors concluded that difficulties with mentalising tasks were found across a wide range of psychiatric, neurological and developmental conditions, reflecting a broader shift within psychological research away from categorical labels and towards a more transdiagnostic approach. It has become apparent that different diagnostic labels may have similar underlying factors (Caspi & Moffitt, 2018) and that specific diagnoses do not necessarily correspond to consistent behavioural presentations (Jones et al., 2021). Moreover, there is an increasing body of evidence to suggest that there is considerable individual variation in mentalising ability even within the neurotypical population (Devine, 2021).

Previous research has demonstrated associations between performance on mentalising tasks and specific aspects of narrative ability. For example, studies focusing on autistic participants have found scores on various mentalising measures to be significantly correlated with the use of mental state terms, syntactic diversity and evaluation (Capps et al., 2000); and to predict narrative coherence and cohesion in story retellings (Hilvert et al., 2016). Relatively weaker mentalising ability has also been linked to increased difficulties in autistic young people's ability to accurately reproduce temporal and causal order in generalised event narratives (Loth et al., 2007). Research including non-autistic comparison groups has also demonstrated significant relationships between mentalising and narrative skills regardless of participants' diagnostic status, in relation to mental state terms (Kuijper et al., 2017); emotional descriptors (Siller et al., 2014); and other measures of discourse pragmatics, such as referencing and causal conjunctions (Kuijper et al., 2017). At the macrostructural level, mentalising ability has been shown to be positively associated with the production of more coherent, structured and elaborative written

narratives, by both autistic and non-autistic young people alike (Hilverd et al., 2020). However, despite providing evidence that mentalising ability is important for adolescents' narrative generation, few previous studies have also considered the role of EF within the same sample. There is also little prior work that has investigated associations between these underlying cognitive skills and the coherence of spoken accounts by autistic and non-autistic young people.

Executive function (EF)

EF is an umbrella term encompassing a range of higher-order cognitive processes that are considered essential for “complex, novel and goal-oriented behaviours” (Jones et al., 2018), although there is some disagreement over how best to characterise this domain of cognitive functioning. Miyake et al. (2000) proposed a triadic model that identified the core components of EF as ‘updating of working memory’ (WM), ‘response inhibition’ and ‘set shifting’. This is arguably the most replicated and empirically supported EF framework (Jewsbury et al., 2016), with its core structure successfully replicated in children (Lehto et al., 2003) and across the adult lifespan (Fisk & Sharp, 2004; Vaughan & Giovanello, 2010). Fisk and Sharp (2004) have also argued for including generativity as a fourth key function, since this was found to load onto a distinct factor in their model. Although research evidence demonstrates executive functioning differences in autistic individuals (see Demetriou et al., 2018, for a meta-analysis), individual research findings relating to EF and autism are inconsistent. Previous studies broadly confirm difficulties with shifting tasks, while evidence for difficulties with WM, inhibition and generativity is more equivocal (Pennington & Ozonoff, 1996). However, there are notable issues with measuring EF

related to unusually spiky profiles (Christ et al., 2011); task administration formats (Kenworthy, 2008) and the highly heterogeneous nature of the autistic population (Mottron, 2004; Sergeant et al., 2002)

Research indicates that EF may be implicated in narrative structure; for instance, in the ability to structure the narrative around key plot elements, organise these in a coherent manner, and stay on topic by ignoring irrelevant information (Ketelaars et al., 2012; Mar, 2004). EF have also been linked specifically to narrative coherence (Dealy et al. 2019; Scholtens et al., 2014; Kuijper et al., 2017; Bourke et al., 2020). Autistic individuals tend to have difficulties with complex executive tasks (Hill, 2004; Sergeant et al., 2002), which could affect the ability to organise and relate the content of a narrative. Bourke et al. (2020) suggest that challenges with mental flexibility might disadvantage autistic narrators in generating invented stories or embellishing events. However, empirical evidence for the contribution of EF to narrative ability in this population is limited. Some previous studies have indicated a relationship between WM and pragmatic narrative ability in autism, possibly due to the challenge of simultaneously processing and organising linguistic information while monitoring what has already been said (Baixauli-Forte et al., 2019; Schuh et al., 2016). Kuijper et al. (2017) also noted that autistic children who had poorer WM capacity produced stories that were shorter and simpler but found no association between measures of inhibition and narrative ability. Conversely, Greco et al. (2023) found no relationship between WM and narrative performance in their sample but observed that children with better inhibition skills tended to produce more fluent narratives, with fewer instances of self-repair. Overall, the contribution of specific domains of EF to

narrative skills in autistic individuals is still poorly understood (Greco et al., 2023), suggesting that further exploration of this area is warranted.

The present study

This study investigated the relationship between the cognitive factors described above and spoken narrative skills in autistic and non-autistic adolescents with typical-range cognitive and linguistic abilities. The research questions were:

1. Is there an association between mentalising ability or performance on EF tasks and narrative structure ('story grammar')?
2. Is there an association between mentalising ability or performance on EF tasks and narrative coherence?
3. After accounting for mentalising and EF ability, does diagnostic status predict narrative ability?

Based on previous literature, we predicted that mentalising ability would contribute to narrative structure and narrative coherence scores across the whole sample. We also tentatively predicted associations between some aspects of EF and narrative performance. However, due to limited previous research, we were not confident in making predictions about the specific EF components involved. We anticipated that diagnostic status would predict narrative ability, i.e., that group differences in the cognitive skills described above would act as a limiting factor in the performance of the autistic group.

Methods

Recruitment

Ethical approval for this study was obtained from the Department of Language and Communication Science Proportionate Review Committee at City, University of London (ETH1920-1434) on 14.07.2020. Informed consent was obtained verbally from all participants and in written form from their parents before participating in the study. Recruitment took place through social media channels, autism research networks (Autistica and the Cambridge Autism Research Database) and secondary schools. A small incentive of an online shopping voucher was offered to participants.

Inclusion criteria

Participants included in the autistic group were required to have a formal diagnosis of autism. Diagnostic status was supported by administering the Social Responsiveness Scale (SRS-2; Constantino & Gruber, 2012), a parental report screening instrument which identifies the presence and level of autistic traits (T-scores of 59 or below are considered within typical limits. However, participants with a formal autism diagnosis were not excluded if they scored slightly below this threshold, as lower scores could be due to measurement error, or might reflect the impact of interventions aimed at developing participants' communication and social skills). Participants in both groups were not excluded from the study due to additional diagnoses, such as dyslexia or ADHD. Participants were excluded if they did not reside in the United Kingdom; did not fall within the stipulated age range at the time of assessment; did not speak English fluently; or were not able to communicate verbally in full sentences. To ensure that our autistic and non-

autistic groups were comparable, participants were also excluded if they scored lower than 2 SD below the mean on the 'Matrix Reasoning' subtest of the WASI-II (Wechsler, 2011). This measure of non-verbal cognitive ability was used to estimate whether participants had intellectual abilities that were likely to fall within the borderline to typical range (i.e., corresponding to a full-scale IQ score of more than 70), without subjecting them to excessive testing.

Participants

Study design

The study involved two groups: autistic adolescents aged 11-15 years and a non-autistic comparison group. The wider study sample comprised 110 participants, and is described in Harvey et al. (2024). Due to missing data for some of the cognitive measures, the study reported in this article included 98 of these participants (44 autistic, 54 non-autistic). Participants were matched at a group level on chronological age, sex, non-verbal cognitive ability and scores on receptive vocabulary and expressive language measures. Participants were assessed at one timepoint on two narrative tasks and a battery of cognitive measures.

To investigate group differences, participants were compared on key background variables using independent samples *t*-tests (**Table 1**). No significant differences were found for age, non-verbal cognitive ability, receptive vocabulary or expressive language skills; however, *p*-values for non-verbal cognitive ability and expressive language did not meet the threshold recommended for group matching by Mervis and Klein-Tasman (2004). All background variables were therefore controlled in subsequent analyses. As

anticipated, scores on the Social Responsiveness Scale (SRS-2) differed significantly between the groups, reflecting their diagnostic status ($p < .001$). The groups did not differ significantly in terms of sex, as demonstrated by a chi-square test ($\chi^2(1) = .934$, $p = .334$).

Table 1. Mean (SD) scores and ranges for background variables (age, non-verbal cognitive ability, receptive vocabulary and expressive language) and SRS-2 scores for autistic and non-autistic groups, with group differences.

Variables	Autistic group (N=44; 31M, 13F)		Non-autistic group (N=54; 32M, 22F)		Group differences
	<u>M (SD)</u>	<u>Range</u>	<u>M (SD)</u>	<u>Range</u>	
Age (months)	159.98 (16.75)	132 - 191	158.56 (15.96)	133 - 190	$t(96) = -.43$, $p = .669$
Wechsler Abbreviated Scale of Intelligence – II: <i>Matrix Reasoning</i> (T-scores: $M = 50$, $SD = 10$)	50.82 (9.05)	32 - 71	53.96 (10.05)	38 - 77	$t(96) = 1.61$, $p = .111$
British Picture Vocabulary Scale – 3 (standardised scores: $M = 100$, $SD = 15$)	102.73 (15.47)	70 - 135	104.76 (12.36)	84 - 131	$t(96) = .72$, $p = .471$
Clinical Evaluation of Language Fundamentals –5 UK: Recalling	10.14 (2.70)	5 - 16	10.69 (3.43)	5 - 19	$t(96) = .87$, $p = .389$

Sentences (scaled
scores: $M= 10$, $SD=$
3)

Social	77.95 (10.48)	50 - 90	51.89 (11.74)	39 - 84	$t(96)= -$
Responsiveness					11.47,
Scale -2 (T-scores:					$p< .001^{***}$
$M= 50$, $SD= 10$)					

Data was collected from parents on participants' ethnicity and any additional neurodevelopmental or psychiatric diagnoses. Both groups comprised mostly white participants, although the autistic group was more ethnically diverse overall. There were some participants in both groups with additional diagnoses, (e.g., ADHD, Dyslexia); however, these were more common among the autistic participants (see supplementary materials for detailed demographics).

Procedure

Due to the impact of the Covid pandemic and the ensuing national lockdowns, all data collection for the study was carried out online over Zoom. Although most of the assessment measures used were not designed to be administered remotely, research evidence indicates that online video-based administration of formal language assessments shows good reliability and validity when compared to face-to-face administration (Ciccia et al., 2011; Waite et al., 2010). Remote administration of cognitive assessments, including executive function tasks, has also been shown to produce

comparable results to in-lab administration (Collins et al., 2022). Despite this, the remote administration of assessments could potentially be challenging for some participants, particularly those with additional support needs. However, none of the autistic participants in the present study (who had low support needs) experienced difficulties with this procedure. In fact, several individuals provided positive feedback that the online assessment method corresponded to their preferences, for reasons such as being in a familiar environment and being able to switch off their camera if they wished.

Ninety-five participants were assessed at their homes. Parents remained present during the sessions but were instructed not to prompt their child. Three of the autistic participants were assessed at school, supported by a member of staff; however, these sessions were otherwise identical in format, and all participants were assessed in a quiet room, with one adult present. Assessment sessions were led by the first author, a qualified Speech and Language Therapist with extensive experience of working with autistic clients. Sessions lasted between 60-90 minutes, with breaks offered to participants as often as desired. Two participants opted to complete their assessments across two shorter sessions (30-45 minutes). A visual timetable was used to introduce each activity. To ensure an identical procedure, the assessor followed a script, and tasks were presented in the same order for each participant.

Background measures

Participants were assessed on their non-verbal cognitive skills using the WASI-II: 'Matrix Reasoning' subtest (Wechsler, 2011). This is a measure of 'perceptual reasoning', in

which participants complete a matrix by choosing the correct picture and is scored using T-scores (internal consistency of .86-.87; test-retest stability of .76-.81). Receptive vocabulary was assessed using the British Picture Vocabulary Scale (BPVS-3; Dunn & Dunn, 2009), a standardised measure that requires participants to choose the picture that best represents a spoken word from a choice of four (reliability of .91). The published procedure was followed for both measures, except that stimuli were presented via screen-share and participants were instructed to say the number of their chosen response rather than point to the item. Expressive language was assessed using the Clinical Evaluation of Language Fundamentals (CELF-5) 'Recalling Sentences' subtest (Wiig et al., 2013), in which participants were asked to repeat spoken sentences and responses were scored according to the number of errors, before being converted to scaled scores (reliability of .82 - .90).

Narrative task

Participants viewed two short video clips (3-4 mins each). After each one, they were immediately asked to provide a free recall account of 'what happened', with no further prompts given. To reflect the content of communication in real-life contexts, the videos were chosen to reflect familiar social situations and approximated everyday uses of narrative skills, e.g., an adolescent telling an adult about their day at school. Video A was an animated sequence with no dialogue, showing a misunderstanding between two strangers. Video B was a live-action sequence in which a secondary-aged student arrives late to his English lesson and gets in trouble. Narrative retellings were recorded and transcribed following the session, then assessed for their overall structure

(macrostructure). A 'story grammar' framework was used, with the events of the stimulus videos coded according to seven key story elements (setting, initiating event, plan, action/attempt, consequence, internal response and resolution). Participants' narratives were scored against this framework, and the total number of story elements was calculated. Narrative coherence was scored using a novel framework created for this study (the '6Cs'). A rating scale (0-3) was used to assess six dimensions of coherent storytelling (context, characterisation, chronology, causality, cohesion, and congruence). These scores were then summed to create a total coherence score, out of a maximum of 18. For a detailed description of the narrative task and both scoring methods, see supplementary materials. Inter-rater reliability measures were carried out for 36% of narrative transcripts, ranging from 'good' to excellent' for narrative structure (ICC: Video A: $M = 0.90$; Video B: $M = 0.89$) and 'moderate' to 'good' for narrative coherence (ICC: Video A: 0.74; Video B: 0.83).

Mentalising tasks

Two assessments of mentalising ability were used to obtain a composite score. These were five short 'mental state' stories ('Strange Stories'; Happé, 1994, reproduced in White et al., 2009) and a complementary, film-based version of this task ('Silent Film Task', Devine & Hughes, 2013). Participants were presented with the written story text accompanied by an audio recording of the researcher reading this aloud and finishing with a 'mentalising' question (e.g., "Why did the burglar do that?"). Participants were also presented with five short clips from a classic silent film and asked to respond to questions which required them to infer the thoughts, feelings or motivations of the characters (e.g.,

“Why did the men hide?”). Both measures used the same 0-2 coding system, in which a correct mental-state response received 2 points, whereas a factual answer lacking any explicit reference to mental state received only 1 point. Participants were asked 11 mentalising questions in total, with a maximum possible score of 22.

Executive Function (EF) task battery

A battery of assessments was devised to measure the four principal components of EF identified above:

WM task: Working Memory Test Battery for Children: ‘Listening Recall’ subtest (WMTB-C; Pickering & Gathercole, 2001)

Participants listened to sentences read out by the examiner and stated whether each sentence was true or false. Following this, participants recalled the final word from all sentences in serial order. The task gradually increased in difficulty, starting with one sentence and moving up to two, three, or more sentences in blocks of 6 trials. To progress to higher blocks, participants had to correctly recall all sentence final words in serial order on at least 4/6 trials within a block; scores reflected total number of trials correct.

Inhibition task: ‘Red or Green?’ (Tatool Online - adapted from ‘Simon Task’ - von Bastian, Souza & Gade, 2016)

In this computer-based task, red or green coloured circles appeared on the screen, either to the right- or left-hand side. Participants responded by pressing a predetermined arrow key (left or right) for each colour, regardless of where the circle appeared on the screen. Following six practice trials, participants were presented with 80 randomised trials. Before each trial, a blank screen was displayed. The stimulus then appeared on either the left or the right of the screen for 2000ms. To score the task, the total response time for congruent and incongruent trials was recorded and converted to seconds. The number of correct responses for each condition was then divided by the total response time, to give a rate of correct responses per second. The difference between these scores was calculated, allowing both reaction speed and accuracy to be expressed in one score (the 'interference cost').

Shifting task: 'Size? Alive?' (Tatool Online - adapted from 'Animacy/Size Shifting Task' - von Bastian, Souza, & Gade, 2016)

In this computer-based task, participants pressed the left and right arrow keys to categorise hand-drawn images by either animacy (animal or object) or size (larger or smaller than a football). Each trial was accompanied by a visual cue to remind participants of the relevant sorting dimension. Following six practice trials, each condition included 25 assessed trials. After animacy and size conditions were completed, participants completed 50 'mixed' trials (i.e., shifting between these conditions). Trials followed a set order, with stimuli remaining on-screen until a button was pressed to move on to the next trial. The total response times for the two non-switch blocks were summed. As for the

inhibition task, an overall rate of correct responses was generated for condition, and then the difference was calculated to determine the 'switch cost'.

Generativity task: Category Fluency test (adapted from Clinical Evaluation of Language Fundamentals – Fourth Edition, CELF-4 UK: 'Word Associations' subtest; Semel et al., 2006)

In this task, participants rapidly named as many items as they could think of within a given category. The researcher modelled the task, using the category 'furniture'. Participants were timed for one minute generating responses for the category 'animals' and then again for 'food'. Responses were transcribed and counted to produce a raw score. Repeated items, or items that did not correspond to the category were not scored.

Data analysis

Mean scores on all study variables for each group are displayed in **Table 2**.

Scores were analysed using IBM SPSS Statistics, version 29. A preliminary review of the data included a Principal Components Analysis (Jolliffe, 2002) of the four EF measures. Varimax rotation was used, and two factors were identified as those with Eigenvalues of >1.0 (Kaiser rule; Kaiser, 1960). Individual tasks were included in the factors if the correlation with overall factor score was $r > 0.5$. WM - Listening Recall task and Generativity – Category Fluency task loaded onto one factor (EF1) and Inhibition – 'Red or Green?' task and Shifting - 'Size/Alive?' task loaded onto another (EF2). These two components

were used as the EF measures in all subsequent analyses. Pearson correlations were used to investigate the relationship between Mentalising and EF scores, revealing low levels of correlation across these tasks (Mentalising/EF1: $r = .388$, $p < .001$; Mentalising/EF2: $r = -.001$, $p = .991$).

In all analyses, narrative scores for Videos A and B were considered separately, due to differences in the content and presentation of both videos (Video A was a non-verbal animated sequence, whereas Video B featured real actors and dialogue). Since these stimuli placed different demands on participants in terms of processing and comprehension, combining the scores from the two videos might have obscured potentially differing patterns of results. There was a moderately large positive correlation for Story grammar scores across Videos A and B ($r = .648$, $p < .001$), and a low positive correlation for Coherence scores across both videos ($r = .459$, $p < .001$). This indicated that although the two video tasks captured somewhat similar narrative abilities, they were not directly comparable.

Hierarchical linear regressions were carried out for each narrative measure (i.e., narrative structure and narrative coherence were analysed separately for each stimulus video). The background (henceforth 'control') variables of age (months), non-verbal cognitive scores (WASI-II: 'Matrix Reasoning'), receptive vocabulary scores (BPVS-3) and expressive language scores (CELF-5: 'Recalling Sentences') were entered in the first step. In the second step, Mentalising scores and the two EF measures (EF1, EF2) were entered, to determine whether these added any unique variance beyond the control variables alone. In the final step, Group was entered as a dummy variable, to investigate the impact of diagnostic status on narrative ability.

Additional exploratory analyses

Despite the overall group difference observed on SRS-2 scores, 13 non-autistic participants scored above the cut-off for ‘typical’ limits on this measure. This raised the possibility that these individuals might either fulfil diagnostic criteria for autism or exhibit sub-threshold autism-like characteristics. However, research suggests that the presentation of some non-autism-related learning or behavioural difficulties may also result in elevated SRS-2 scores (Cholemkery et al., 2014; Hus et al., 2013; Wigham et al., 2012). In either case, we were concerned that the presence of non-autistic participants with elevated scores on this measure might alter our results. For this reason, we ran a parallel set of exploratory analyses with these participants excluded (N=85), which are presented in supplementary materials.

Results

Table 2. Mean (SD) scores and ranges for study variables (narrative structure, narrative coherence, Mentalising and EF measures for autistic and non-autistic groups, with group differences.

Variables	Autistic group (N=44; 31M, 13F)		Non-autistic group (N=54; 32M, 22F)		Group differences
	<u>M (SD)</u>	<u>Range</u>	<u>M (SD)</u>	<u>Range</u>	
Narrative structure (/26)					
Video A	13.18 (4.40)	4-22	16.35 (4.30)	5-24	$t(96)= 3.59, p<.001^{***}$
Video B	11.80 (3.79)	3-19	15.69 (3.71)	8-23	$t(96)= 5.12, p<.001^{***}$

Narrative coherence (/18)					
Video A	11.77 (3.03)	3-16	13.04 (1.57)	9-16	$t(61)= 2.51, p=.015^*$
Video B	11.55 (2.99)	4-16	12.91 (2.30)	6-17	$t(79)= 2.48, p=.015^*$
Mentalising (/22)					
	13.91 (3.71)	6-20	15.46 (3.17)	8-21	$t(96)= 2.24, p=.028^*$
EF measures:					
WM (SS: $M=100, SD=15$)	106.05 (22.07)	65-144	111.50 (26.21)	59-144	$t(96)= 1.10, p=.275$
Inhibition (interference cost, ms)	.20 (.17)	-.18 - .57	.14 (.19)	-.27 - .81	$t(96)= -1.56, p=.123$
Shifting (switch cost, ms)	.36 (.19)	-.06 - .88	.42 (.17)	.04 - .84	$t(96)= 1.64, p=.104$
Generativity (total no. of items)	40.30 (11.17)	17-75	47.06 (11.43)	23-73	$t(96)= 2.94, p=.004^{**}$

Narrative structure (story grammar):

Video A (animated, no dialogue)

A hierarchical linear regression model was used to investigate the contribution of the control variables (age, non-verbal cognitive scores, receptive vocabulary scores and expressive language scores); mentalising and EF scores; and Group to narrative structure scores for Video A. Step 1 of the regression model (control variables) was non-significant, $R^2= .085, F(4, 93)= 2.171, p= .078, \text{adj. } R^2= .046$. Step 2 of the model was statistically significant, $R^2= .257, F(7, 90)= 4.458, p< .001, \text{adj. } R^2= .200$, explaining an

additional 17% of unique variance once the control variables had been accounted for ($\Delta R^2=.172$). Mentalising was the only significant predictor of story grammar scores for Video A ($\beta= .386$, $p< .001$), with EF scores failing to reach significance (EF1: $\beta= .226$, $p= .056$; EF2: $\beta= -.083$, $p= .377$). In the final model (Step 3), the inclusion of Group significantly predicted 4% of additional variance overall, with the autistic group showing lower scores than the non-autistic group, $\Delta R^2=.043$, $F(8, 89)= 4.779$, $p< .001$ (Group: $\beta= -.222$, $p= .022$). Mentalising remained significant as a predictor variable ($\beta= .348$, $p= .001$). See **Table 3**.

Table 3. Hierarchical linear regression predicting total narrative structure ('story grammar') scores for Video A from control variables, Mentalising and EF scores, and diagnostic group (final model).¹

Variables	B (95% CI)	SE B	β	p
Constant	9.175 (-2.552 - 20.903)	5.902	-	.124
Age	.018 (-.034 - .070)	.026	.064	.491
Receptive vocabulary	-.013 (-.096 - .071)	.042	-.038	.764
Expressive language	.032 (-.311 - .375)	.173	.022	.853
Non-verbal ability	.001 (-.105 - .107)	.053	.002	.986
Mentalising**	.459 (.186 - .732)	.138	.348	.001**
EF1	.727 (-.350 - 1.804)	.542	.158	.183
EF2	-.338 (-1.171 - .496)	.419	-.073	.423
Group*	-2.043 (-3.778 - -.308)	.873	-.222	.022*

Note: N=98. $R^2= .085$ for Step 1 ($p=.078$). $\Delta R^2= .172$ for Step 2 ($p< .001^{***}$). $\Delta R^2 =.043$ for Step 3 ($p< .001^{***}$).

Narrative structure ('story grammar'):

Video B (real actors, with dialogue)

A hierarchical linear regression model was used to investigate the contribution of the control variables (age, non-verbal cognitive scores, receptive vocabulary scores and expressive language scores); mentalising and EF scores; and Group to narrative structure scores for Video B. Step 1 of the regression model was statistically significant, $R^2 = .113$, $F(4, 93) = 2.970$, $p = .023$, adj. $R^2 = .075$, although none of the control variables were individually significant. Step 2 of the model was also significant, $R^2 = .261$, $F(7, 90) = 4.547$, $p < .001$, adj. $R^2 = .204$, explaining an additional 15% of unique variance once the control variables had been accounted for ($\Delta R^2 = .148$). Mentalising was the only significant predictor of 'story grammar' scores for Video B ($\beta = .421$, $p < .001$), with EF scores failing to reach significance (EF1: $\beta = .028$, $p = .811$; EF2: $\beta = -.111$, $p = .237$). In the final model (Step 3), the inclusion of Group significantly predicted 13% of additional variance overall, with the autistic group showing lower scores than the non-autistic group, $\Delta R^2 = .133$, $F(8, 89) = 7.240$, $p < .001$ (Group: $\beta = -.390$, $p < .001$). As for Video A, Mentalising remained significant as a predictor variable ($\beta = .354$, $p < .001$). See **Table 4**.

Table 4. Hierarchical linear regression predicting total narrative structure ('story grammar') scores for Video B from control variables, Mentalising and EF scores, and diagnostic group (final model).

Variables	B (95% CI)	SE B	β	p
Constant	3.651 (-6.310 - 13.612)	5.013	-	.468
Age	.031 (-.013 - .075)	.022	.120	.166
Receptive vocabulary	.016 (-.055 - .087)	.036	.051	.663
Expressive language	.106 (-.185 - .397)	.147	.079	.471
Non-verbal ability	.021 (-.069 - .111)	.045	.048	.646
Mentalising***	.426 (.194 - .658)	.117	.354	<.001***

EF1	-.387 (-1.301 - .528)	.460	-.092	.403
EF2	-.396 (-1.104 - .312)	.356	-.094	.270
Group***	-3.278 (-4.752 - -1.804)	.742	-.390	<.001***

Note: N=98. $R^2 = .113$ for Step 1 ($p=.023^*$). $\Delta R^2 = .148$ for Step 2 ($p< .001^{***}$). $\Delta R^2 = .133$ for Step 3 ($p< .001^{***}$).

Narrative coherence:

Video A (animated, no dialogue)

A hierarchical linear regression model was used to investigate the contribution of the control variables (age, non-verbal cognitive scores, receptive vocabulary scores and expressive language scores); mentalising and EF scores; and Group to narrative coherence scores for Video A. Step 1 of the regression model (control variables) was non-significant, $R^2 = .071$, $F(4, 93) = 1.772$, $p = .141$, adj. $R^2 = .031$. Step 2 of the model was statistically significant, $R^2 = .230$, $F(7, 90) = 3.841$, $p = .001$, adj. $R^2 = .170$, explaining an additional 16% of unique variance once the control variables had been accounted for ($\Delta R^2 = .159$). Mentalising significantly predicted coherence scores for Video A ($\beta = .335$, $p = .002$), as did EF1 scores ($\beta = .273$, $p = .024$). EF2 scores were not significant at Step 2 ($\beta = -.078$, $p = .412$). In the final model (Step 3), although the inclusion of Group significantly predicted 2% of additional variance, $\Delta R^2 = .021$, $F(8, 89) = 3.720$, $p < .001$, Group was not individually significant (Group: $\beta = -.153$, $p = .122$). Mentalising remained significant as a predictor at Step 3 ($\beta = .309$, $p = .005$), whereas EF1 scores fell short of significance ($\beta = .226$, $p = .067$). **See Table 5.**

Table 5. Hierarchical linear regression predicting total narrative coherence scores for Video A from control variables, Mentalising and EF scores, and diagnostic group (final model).

Variables	B (95% CI)	SE B	β	<i>p</i>
Constant***	11.154 (4.803 - 17.505)	3.196	-	<.001***
Age	.003 (-.025 - .031)	.014	.018	.851
Receptive vocabulary	.013 (-.032 - .058)	.023	.075	.570
Expressive language	-.003 (-.189 - .183)	.093	-.004	.976
Non-verbal ability	-.048 (-.105 - .010)	.029	-.192	.103
Mentalising**	.213 (.065 - .361)	.074	.309	.005**
EF1	.545 (-.039 - 1.128)	.294	.226	.067
EF2	-.173 (-.624 - .278)	.227	-.072	.448
Group	-.739 (-1.678 - .201)	.473	-.153	.122

Note: N=98. $R^2 = .071$ for Step 1 ($p=.141$). $\Delta R^2 = .159$ for Step 2 ($p= .001^{**}$). $\Delta R^2 = .021$ for Step 3 ($p< .001^{***}$).

Narrative coherence:

Video B (real actors, with dialogue)

A hierarchical linear regression model was used to investigate the contribution of the control variables (age, non-verbal cognitive scores, receptive vocabulary scores and expressive language scores); mentalising and EF scores; and Group to narrative coherence scores for Video B. Step 1 of the regression model (control variables) was statistically significant, $R^2 = .219$, $F(4, 93) = 6.519$, $p < .001$, adj. $R^2 = .185$; with Age and Receptive vocabulary significantly predicting coherence scores (Age: $\beta = .219$, $p = .020$; Receptive vocabulary: $\beta = .265$, $p = .036$). Step 2 of the model was also significant, $R^2 = .316$, $F(7, 90) = 5.933$, $p < .001$, adj. $R^2 = .263$, explaining an additional 10% of unique variance once the control variables had been accounted for ($\Delta R^2 = .097$). Mentalising was

the only significant predictor of coherence scores for Video B at Step 2 ($\beta = .348, p < .001$), with EF scores failing to reach significance (EF1: $\beta = .005, p = .967$; EF2: $\beta = -.073, p = .418$). At Step 2, Age and Receptive vocabulary were non-significant (Age: $\beta = .167, p = .067$; Receptive vocabulary: $\beta = .182, p = .144$). In the final model (Step 3), although the inclusion of Group significantly predicted 2% of additional variance ($\Delta R^2 = .024, F(8, 89) = 5.737, p < .001$), Group was not individually significant (Group: $\beta = -.167, p = .073$). However, Mentalising remained significant as a predictor variable ($\beta = .319, p = .002$), and Age was also significant in this final model ($\beta = .182, p = .044$). See **Table 6**.

Table 6. Hierarchical linear regression predicting total narrative coherence scores for Video B from control variables, Mentalising and EF scores, and diagnostic group (final model).

Variables	B (95% CI)	SE B	β	<i>p</i>
Constant	-.520 (-7.218 - 6.177)	3.371	-	.878
Age*	.030 (.001 - .060)	.015	.182	.044*
Receptive vocabulary	.041 (-.007 - .089)	.024	.208	.094
Expressive language	.080 (-.115 - .276)	.099	.093	.417
Non-verbal ability	.011 (-.050 - .071)	.030	.039	.725
Mentalising**	.247 (.091 - .404)	.079	.319	.002**
EF1	-.127 (-.742 - .489)	.310	-.047	.684
EF2	-.178 (-.654 - .298)	.240	-.066	.459
Group	-.906 (-1.897 - .085)	.499	-.167	.073

Note: N=98. $R^2 = .219$ for Step 1 ($p < .001^{***}$). $\Delta R^2 = .097$ for Step 2 ($p < .001^{**}$). $\Delta R^2 = .024$ for Step 3 ($p < .001^{***}$).

Discussion

The present study investigated associations between narrative ability and performance on measures of Mentalising and EF in a sample of autistic and non-autistic adolescents with comparable linguistic and non-verbal skills. In line with predictions, Mentalising was the most important variable in predicting narrative ability, whether assessed using measures of narrative structure (story grammar) or narrative coherence, across both videos. In fact, Mentalising emerged as a more consistent predictor of narrative scores than diagnostic group, remaining significant in all models even after Group was added in the final step. By contrast, contrary to predictions, EF scores did not significantly predict narrative scores in any model, once all variables were included, and this finding was consistent for measures of both narrative structure and coherence across both videos.

It was noteworthy that Group was a significant predictor of participants' narrative structure scores for both videos, with the autistic group scoring lower than the non-autistic group, but did not predict narrative coherence scores for either video. Similarly, although age and receptive vocabulary scores predicted narrative coherence scores for Video B in some steps of the regression model, the overall pattern of results indicated that the background variables (age, non-verbal cognitive ability, receptive vocabulary, expressive language) were not key factors contributing to narrative performance in this sample of adolescents with typical-range cognitive and verbal abilities. Additional exploratory analyses excluding non-autistic participants with elevated SRS-2 scores (Appendix B) confirmed our overall findings that Mentalising predicts both narrative structure and coherence, irrespective of Group.

These findings highlight the importance of mentalising skills for producing well-structured and coherent accounts of events that take place within social contexts. Mentalising may play a role in two different ways (see Ketalaars et al., 2012): a) helping individuals to comprehend the original events that have transpired (or, in the present study, interpreting the content of the videos); and b) generating a verbal account of these events in order to relay the most salient details to their listener, whilst providing sufficient contextual information

Our study strongly suggests that young people with poorer mentalising skills may struggle with providing spoken narrative accounts, regardless of their diagnostic status. The findings align with previous research in this area demonstrating associations between mentalising ability and various aspects of narrative performance in autistic samples (e.g., Capps et al., 2000; Hilvert et al., 2016; Loth et al., 2007) and in samples including both autistic and non-autistic children (Hilvert et al., 2020; Kuijper et al., 2017; Siller et al., 2014). Our findings also uphold recent work indicating that individual differences in mentalising abilities transcend diagnostic categories (e.g., Cotter et al., 2018; Devine 2021).

Neither of the dimensions of EF measured in this study were found to significantly predict either narrative structure or coherence in our sample. It is possible, however, that the large number of predictor variables may have meant that we were unable to detect small, but real, effects for some of these aspects of EF in relation to narrative scores. An alternative explanation could be that the experimental tasks used in the present study might not accurately reflect how participants use their executive skills in 'real-life' contexts (see Kenworthy et al., 2008).

Another speculative explanation for these null findings is that our sample may have been comprised of young people with relatively strong EF skills, even within the autistic group. The practical requirements of participation in the study, such as the ability to focus during a one-hour assessment session, could have introduced a sampling bias whereby adolescents with poorer EF skills were less likely to volunteer. Our recruitment strategy, which principally targeted adolescents attending mainstream educational settings, could have also inadvertently 'screened out' participants with more pronounced executive difficulties who might struggle to cope with the demands of a mainstream secondary school (see Jacobson et al., 2011) and therefore might be more likely to attend specialist or alternative provisions. Since we cannot discount the possibility that our sample was not representative of EF skills within the wider autistic population, we consider that further research in this area is warranted to explore the potential impact of executive difficulties on spoken discourse.

While Group was a significant predictor for narrative structure scores across both videos, with autistic adolescents scoring lower than their non-autistic counterparts, Mentalising remained significant even after Group was added to the regression model in the final step. In relation to narrative coherence, however, Group did not significantly predict additional variance in scores for either video when entered in the final step of the regressions (after controlling for all other variables). This suggests that telling coherent stories might be an area of relative strength for autistic narrators. These findings reflect a broader shift in developmental research towards a neurodiversity perspective, with increasing recognition that there can be common areas of strength and challenge between individuals who may or may not have a particular diagnostic label. Astle and Fletcher-Watson (2020) argue

that rather than focusing on diagnoses, the aim of research should be “to establish which dimensions are important for understanding individual outcomes, irrespective of the diagnostic category applied” (p.432). Our study findings indicate that adolescents’ narrative abilities should not be assumed based on diagnosis (or lack thereof) and suggest that supporting and facilitating mentalising skills might be a useful approach for any young person who struggles with spoken discourse. The findings also lend support to previous research linking mentalising ability in older children and adolescents to broader social outcomes (Bosacki & Wilde Astington, 1999; Devine & Hughes, 2013), since challenges with narrative skills are known to have a negative impact upon young people’s social competence (see Petersen et al., 2014).

Limitations

The autistic adolescents included in our sample demonstrated broadly typical-range cognitive and linguistic skills and presented with low support needs. This means that the findings may not be generalisable to the wider autistic population, who may have a co-occurring intellectual disability, language disorder, or other complex support needs. Although groups were matched on participant sex, the overall sample was not balanced in terms of male and female participants. This resulted from difficulties recruiting enough autistic girls to the study, a common issue in autism research (Shefcyk, 2015). Since research indicates that autistic girls may present with a different linguistic profile to boys (e.g., Conlon et al., 2019), the comparative lack of female participants may limit the generalisability of our findings.

To provide a representative sample of the broader population, no participants in either group were excluded because of additional diagnoses, such as Dyslexia. However, many more participants in the autistic group had co-occurring diagnoses than participants in the non-autistic comparison group, particularly ADHD (see Appendix A). This reflects the high prevalence of multiple diagnoses amongst autistic people, with autism/ADHD being a common dual diagnosis (Stevens et al., 2016). This may introduce a potentially confounding variable when interpreting the results of this study, and the impact of co-occurring conditions on narrative ability warrants further study.

A further limitation of the present study was that only a 'moderate' to 'good' level of inter-rater reliability was achieved for the '6Cs' coherence framework. We note that this novel assessment tool requires further piloting and refinement to reduce subjectivity in scoring and improve its reliability.

In conclusion, mentalising skills appear to be fundamental in supporting adolescents' production of well-structured and coherent verbal narrative accounts, regardless of whether they have an autism diagnosis or not. Further research in this area could consider the factors contributing to poorer mentalising ability, and whether developing adolescents' mentalising skills might improve their ability to successfully generate spoken narratives in their daily lives.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable?. *Child Development*, 77(6), <https://doi.org/10.1111/j.1467-8624.2006.00968.x>

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington, D.C.: American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>

Astle, D. E., & Fletcher-Watson, S. (2020). Beyond the core-deficit hypothesis in developmental disorders. *Current Directions in Psychological Science*, 29(5), 431–437. <https://doi.org/10.1177/0963721420925518>

Baixauli, I., Colomer, C., Roselló, B., & Miranda, A. (2016). Narratives of children with high-functioning autism spectrum disorder: A meta-analysis. *Research in Developmental Disabilities*, 59, 234-254. <https://doi.org/10.1016/j.ridd.2016.09.007>

Baixauli-Fortea, I., Miranda Casas, A., Berenguer-Forner, C., Colomer-Diago, C., & Roselló-Miranda, B. (2019). Pragmatic competence of children with autism spectrum disorder: Impact of theory of mind, verbal working memory, ADHD symptoms, and structural language. *Applied Neuropsychology. Child*, 8(2), 101-112. <https://doi.org/10.1080/21622965.2017.1392861>

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or

Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 38(7), 813-822.
<https://doi.org/10.1111/j.1469-7610.1997.tb01599.x>

Baron-Cohen, S., Scott, F., Allison, C., Williams, J., Bolton, P., Matthews, F., & Brayne, C. (2008). Prevalence of Autism-spectrum conditions: UK school-based population study. *British Journal of Psychiatry*.
<https://doi.org/10.1192/bjp.bp.108.059345> h

Bosacki, S., & Wilde Astington, J. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development*, 8(2), 237-255. <https://doi.org/10.1111/1467-9507.00093>

Bourke, L., Marriott-Fellows, M., Jones, A., Humphreys, L., Davies, S. J., Zuffiano, A., & López-Pérez, B. (2020). Writing with imagination: The influence of hot and cold executive functions in children with autism characteristics and typically developing peers. *Reading & Writing*, 33(4), 935-961.
<https://doi.org/10.1007/s11145-019-09989-w>

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. <http://doi.org/10.1023/a:1005126915631>

Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175(9), 831-844. <https://doi.org/10.1176/appi.ajp.2018.17121383>

- Cholemkery, H., Kitzerow, J., Rohrmann, S., & Freitag, C. M. (2014). Validity of the social responsiveness scale to differentiate between autism spectrum disorders and disruptive behaviour disorders. *European Child & Adolescent Psychiatry*, 23(2), 81-93. <https://doi.org/10.1007/s00787-013-0427-5>
- Christ, S. E., Kester, L. E., Bodner, K. E., & Miles, J. H. (2011). Evidence for selective inhibitory impairment in individuals with autism spectrum disorder. *Neuropsychology*, 25(6), 690. <https://doi.org/10.1037/a0024256>
- Ciccia, A. H., Whitford, B., Krumm, M., & McNeal, K. (2011). Improving the access of young urban children to speech, language and hearing screening via telehealth. *Journal of Telemedicine and Telecare*, 17(5), 240-244.
<https://doi.org/10.1258/jtt.2011.100810>
- Colle, L., Baron-Cohen, S., Wheelwright, S., & van der Lely, Heather K. J. (2008). Narrative discourse in adults with high-functioning autism or asperger syndrome. *Journal of Autism and Developmental Disorders*, 38(1), 28-40.
<https://doi.org/10.1007/s10803-007-0357-5>
- Collins, C. L., Pina, A., Carrillo, A., Ghil, E., Smith-Peirce, R. N., Gomez, M., Okolo, P., Chen, Y., Pahor, A., Jaeggi, S. M., & Seitz, A. R. (2022). Video-based remote administration of cognitive assessments and interventions: A comparison with in-lab administration. *Journal of Cognitive Enhancement*, 6(3), 316-326.
<https://doi.org/10.1007/s41465-022-00240-z>

Conlon, O., Volden, J., Smith, I. M., Duku, E., Zwaigenbaum, L., Waddell, C., Szatmari, P., Mirenda, P., Vaillancourt, T., Bennett, T., Georgiades, S., Elsabbagh, M., & Ungar, W. J. (2019). Gender differences in pragmatic communication in school-aged children with autism spectrum disorder (ASD). *Journal of Autism & Developmental Disorders*, 49(5), 1937-1948. <https://doi.org/10.1007/s10803-018-03873-2>

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale: SRS-2*. Torrance, CA: Western Psychological Services.

Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C. Y., & Barnett, J. H. (2018). Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience and Biobehavioral Reviews*, 84, 92-99. <https://doi.org/10.1016/j.neubiorev.2017.11.014>

Dealy, J., Mudrick, H., & Robinson, J. (2019). Children's narrative story stem responses: Contributions of executive functioning and language proficiency to relationship representations. *Social Development*, 28(1), 168-185. <https://doi.org/10.1111/sode.12335>

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., ... & Guastella, A. J. (2018). Autism spectrum disorders: a meta-analysis of executive function. *Molecular Psychiatry*, 23(5), 1198-1204. <https://doi.org/10.1038/mp.2017.75>.

Devine, R. T. (2021). Individual differences in theory of mind in middle childhood and adolescence. In R. T. Devine & S. Lecce (Eds.), *Theory of mind in middle childhood and adolescence: Integrating multiple perspectives* (pp. 55–76).

Routledge. <https://doi.org/10.4324/9780429326899-5>

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development*, 84(3), 989-1003. <https://doi.org/10.1111/cdev.12017>

Dunn, L. M., & Dunn, D. M. (2009). *The British picture vocabulary scale* (3rd ed.). GL Assessment Limited.

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology*, 26(7), 874-890. <https://doi.org/10.1080/13803390490510680>

Gernsbacher, M. A., & Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a theory of mind. *Archives of Scientific Psychology*, 7(1), 102–118. <https://doi.org/10.1037/arc0000067>

Greco, G., Choi, B., Michel, K., & Faja, S. (2023). Here's the story: Narrative ability and executive function in autism spectrum disorder. *Research in Autism Spectrum Disorders*, 101, 102092. <https://doi.org/10.1016/j.rasd.2022.102092>

Happé, F. (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. <https://doi.org/10.1007/BF02172093>

Harvey, A., Spicer-Cain, H., Botting, N., & Henry, L. (2024). *Everyday narrative skills in autistic adolescents* [Manuscript submitted for publication]. Department of Language and Communication Science, City St George's, University of London.

Harvey, A., Spicer-Cain, H., Botting, N., Ryan, G., & Henry, L. (2023). Assessing 'coherence' in the spoken narrative accounts of autistic people: A systematic scoping review. *Research in Autism Spectrum Disorders*, 102, 102108. <https://doi.org/10.1016/j.rasd.2023.102108>

Henry, L. A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and Adolescent Mental Health*, 15(2), 110-119. <https://doi.org/10.1111/j.1475-3588.2010.00557.x>

Hewitt, L. E. (2019). Narrative as a critical context for advanced language development in autism spectrum disorder. *Perspectives of the ASHA Special Interest Groups*, 4(3), 430-437. https://doi.org/10.1044/2019_PERS-SIG1-2018-0021

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, 24(2), 189-233. <https://doi.org/10.1016/j.dr.2004.01.001>

Hilvert, E., Davidson, D., & Gámez, P. B. (2016). Examination of script and non-script based narrative retellings in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 29, 79-92. <http://doi.org/10.1016/j.rasd.2016.06.002>

- Hilvert, E., Davidson, D., & Gámez, P. B. (2020). Assessment of personal narrative writing in children with and without autism spectrum disorder. *Research in Autism Spectrum Disorders*, 69, 101453. <https://doi.org/10.1016/j.rasd.2019.101453>
- Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*, 54(2), 216-224. <https://doi.org/10.1111/j.1469-7610.2012.02589.x>
- Jacobson, L. A., Williford, A. P., & Pianta, R. C. (2011). The role of executive function in children's competent adjustment to middle school. *Child Neuropsychology*, 17(3), 255–280. <https://doi.org/10.1080/09297049.2010.535654>
- Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell-Horn-Carroll model. *Journal of Experimental Psychology: General*, 145(2), 220 – 245. <https://doi.org/10.1037/xge0000119>
- Jolliffe, I. T. (2002). Principal component analysis and factor analysis. In *Principal component analysis* (2nd ed., pp.150-166). Springer.
- Jones, C. R. G., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J. S., Tregay, J., Happé, F., & Charman, T. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, 11(1), 95-109. <https://doi.org/10.1002/aur.1873>

Jones, J. S., The Calm Team, & Astle, D. E. (2021). A transdiagnostic data-driven study of children's behaviour and the functional connectome. *Developmental cognitive neuroscience*, 52, 101027. <https://doi.org/10.1016/j.dcn.2021.101027>

Kaiser, H. (1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151.

<https://doi.org/10.1177/001316446002000116>

Kaland, N., Callesen, K., Møller-Nielsen, A., Mortensen, E., & Smith, L. (2008).

Performance of children and adolescents with Asperger syndrome or high-functioning autism on advanced theory of mind tasks. *Journal of Autism and Developmental Disorders*, 38(6), 1112-1123. <https://doi.org/10.1007/s10803-007-0496-8>

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review*, 18(4), 320–338. <https://doi.org/10.1007/s11065-008-9077-7>

Ketelaars, M. P., Jansonius, K., Cuperus, J., & Verhoeven, L. (2012). Narrative competence and underlying mechanisms in children with pragmatic language impairment. *Applied Psycholinguistics*, 33(2), 281-303.

<https://doi.org/10.1017/S014271641100035X>

Kuijper, S. J. M., Hartman, C. A., Bogaerds-Hazenberg, S. T. M., & Hendriks, P. (2017). Narrative production in children with autism spectrum disorder (ASD) and

children with attention-deficit/hyperactivity disorder (ADHD): Similarities and differences. *Journal of Abnormal Psychology*, 126(1), 63-75.

<http://doi.org/10.1037/abn0000231>

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, 21(1), 59-80. <https://doi.org/10.1348/026151003321164627>

Loth, E., Gómez, J. C., & Happé, F. (2007). Event schemas in autism spectrum disorders: The role of theory of mind and weak central coherence. *Journal of Autism and Developmental Disorders*, 38(3), 449–463.

<https://doi.org/10.1007/s10803-007-0412-2>

Loveland, K. A., McEvoy, R. E., & Tunali, B. (1990). Narrative story telling in autism and Down's syndrome. *British Journal of Developmental Psychology*, 8(1), 9-23.

<http://doi.org/10.1111/j.2044-835X.1990.tb00818.x>

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, 42(10), 1414–1434.

<https://doi.org/10.1016/j.neuropsychologia.2003.12.016>

Mervis, C. B., & Klein-Tasman, B. P. (2004). Methodological issues in group-matching designs: Alpha levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders*, 34(1), 7-17.

<https://doi.org/10.1023/B:JADD.0000018069.69562.b8>

- 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. <https://doi.org/10.1006/cogp.1999.0734>
- Mottron, L. (2004). Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders*, 34(1), 19-27. <https://doi.org/10.1023/B:JADD.0000018070.88380.83>
- Naigles, L. R., & Chin, I. (2015). Language in children with autism spectrum disorders. *The Cambridge Handbook of Child Language* (pp. 637-658). Cambridge University Press. <https://doi.org/10.1017/CBO9781316095829.029>
- Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders*, 38(3), 287-313. <http://doi.org/10.1080/136820310000108133>
- 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. <http://doi.org/10.1017/S030500091300007X>
- Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 37(1), 51-87. <http://doi.org/10.1111/j.1469-7610.1996.tb01380.x>

- Petersen, D. B., Brown, C. L., Ukrainetz, T. A., Wise, C., Spencer, T. D., & Zebre, J. (2014). Systematic individualized narrative language intervention on the personal narratives of children with autism. *Language, Speech, and Hearing Services in Schools, 45*(1), 67-86. http://doi.org/10.1044/2013_LSHSS-12-0099
- Petersen, D. B., Gillam, S. L., & Gillam, R. B. (2008). Emerging procedures in narrative assessment: The Index of Narrative Complexity. *Topics in Language Disorders, 28*(2), 115-130. doi: 10.1097/01.TLD.0000318933.46925.86
- Pickering, S., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*. Hove: Psychological Corporation.
- Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review, 27*(2), 224-260. <http://doi.org/10.1016/j.dr.2007.02.001>
- Scholtens, S., Rydell, A., Bohlin, G., & Thorell, L. (2014). ADHD symptoms and attachment representations: Considering the role of conduct problems, cognitive deficits and narrative responses in non-attachment-related story stems. *Journal of Abnormal Child Psychology, 42*(6), 1033-1042. <https://doi.org/10.1007/s10802-014-9854-0>
- Schuh, J. M., Eigsti, I. M., & Mirman, D. (2016). Discourse comprehension in autism spectrum disorder: Effects of working memory load and common ground. *Autism Research, 9*(12), 1340-1352. <https://doi.org/10.1002/aur.1632>
- Semel, E., Wiig, E. H., & Secord, W. (2006). *Clinical evaluation of language fundamentals—Fourth Edition UK (CELF-4 UK)*. New York: Pearson Education

Sergeant, J. A., Geurts, H., & Oosterlaan, J. (2002). How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder?. *Behavioural Brain Research*, 130(1-2), 3-28. [https://doi.org/10.1016/s0166-4328\(01\)00430-2](https://doi.org/10.1016/s0166-4328(01)00430-2)

Shefcyk, A. (2015). Count us in: Addressing gender disparities in autism research. *Autism: The International Journal of Research and Practice*, 19(2), 131-132. <https://doi.org/10.1177/1362361314566585>

Siller, M., Swanson, M. R., Serlin, G., & Teachworth, A. G. (2014). Internal state language in the storybook narratives of children with and without autism spectrum disorder: Investigating relations to theory of mind abilities. *Research in Autism Spectrum Disorders*, 8(5), 589-596.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In Freedle, R. O. (Ed.), *New directions in discourse processing* (pp. 53–120). New Jersey: Ablex Publishing Corporation.

Stevens, T., Peng, L., & Barnard-Brak, L. (2016). The comorbidity of ADHD in children diagnosed with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 31, 11-18. <https://doi.org/10.1016/j.rasd.2016.07.003>

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, 25(2), 343-355. <https://doi.org/10.1037/a0017729>

- von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, 45(1), 108-115. <https://doi.org/10.3758/s13428-012-0224-y>
- Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010). Internet-based telehealth assessment of language using the CELF-4. *Language, Speech & Hearing Services in Schools*, 41(4), 445-458. [https://doi.org/10.1044/0161-1461\(2009/08-0131\)](https://doi.org/10.1044/0161-1461(2009/08-0131))
- Wechsler, D. (2011). *WASI-II: Wechsler abbreviated scale of intelligence*. PsychCorp.
- White, S., Hill, E., Happe, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing mentalizing impairments in autism. *Child Development*, 80(4), 1097-1117. <https://doi.org/10.1111/j.1467-8624.2009.01319.x>
- Wigham, S., McConachie, H., Tandos, J., Le Couteur, A. S., & Gateshead Millennium Study core team. (2012). The reliability and validity of the social responsiveness scale in a UK general child population. *Research in Developmental Disabilities*, 33(3), 944-950. <https://doi.org/10.1016/j.ridd.2011.12.017>
- Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical evaluation of language fundamentals: CELF-5*. Pearson.

Appendices

- A:** Additional demographic information
 - B:** Narrative structure (story grammar) scoring frameworks
 - C:** Narrative coherence ('6Cs') scoring framework
 - D:** Correlations between study variables
 - E:** Exploratory analyses excluding non-autistic participants with elevated SRS-2 scores
-

Appendix A: Additional demographic information

Ethnicity of study participants (by parent report):

Autistic group (N= 44)	Non-autistic group (N= 54)
White British: 32 (72.7%)	White British: 46 (85.2%)
White Other: 7 (15.9%)	White Other: 4 (7.4%)
White and Black Caribbean: 1 (2.3%)	White and Black Caribbean 1 (1.9%)
Black British: 1 (2.3%)	Black Caribbean: 1 (1.9%)
Black Caribbean: 1 (2.3%)	Asian (Pakistani): 1 (1.9%)
Asian - Bangladeshi: 1 (2.3%)	Other – Arab: 1 (1.9%)
Other – Kurdish: 1 (2.3%)	

Number of study participants with neurodevelopmental or psychiatric diagnoses other than autism (by parent report):

Autistic group (N=44)	Non-autistic group (N=54)
ADHD:12 (27.3%)	Dyslexia: 5 (9.3%)
Dyslexia: 3 (6.8%)	ADHD: 2 (3.7%)
Dyspraxia/DCD: 3 (6.8%)	Dyspraxia/DCD: 2 (3.7%)
Sensory Processing Disorder: 2 (4.5%)	Anxiety: 1 (1.9%)
Anxiety: 1 (5.7%)	
Mild Learning Difficulties: 1 (2.3%)	
Dysgraphia: 1 (2.3%)	
Epilepsy: 1 (2.3%)	

Appendix B: Narrative structure (story grammar) scoring framework

Story grammar element:	Score:
Setting (S)	
Initiating event (IE)	
Plan (P)	
Action/Attempt (A)	
Consequence (C)	
Internal response (IR)	
Resolution (R)	
Story grammar total score:	

Video A:

Main story events	Story grammar code	Included in narrative? (0/1)
Episode 1		
An old lady/woman/grandma	S	
At the train station	S	
She wants to get some biscuits from the vending machine/wants to buy a snack/wants something to eat	P	
She puts some money in the machine	A	
The packet of biscuits gets stuck/won't come out	IE	
She is annoyed/frustrated/angry	IR	
She hits/kicks/bangs the vending machine	A	
The packet of biscuits comes out/falls down	C	
She is pleased/happy	IR	
She takes the biscuits/puts the packet in her bag	R	
Episode 2		
The old lady goes into the train station/onto the platform/sits down on a bench/ sits next to a teenager/man/boy/	S	
She opens the packet to eat a biscuit/she starts eating the biscuits	P	
The teenager takes a biscuit/also starts eating the biscuits	IE	
She is annoyed/frustrated/angry/ She thinks he is eating her biscuits	IR	
He eats another biscuit/keeps taking the biscuits	A	
The old lady tells the teenager off/shouts at him	C	
They fight over the biscuits/she tries to take the last biscuit off him	A	

The teenager offers half of the last biscuit to the old lady/tries to share the last biscuit with her	C	
The old lady crumbles the biscuit in her hand/throws it down/rejects his offer	R	
Episode 3		
The old lady gets on the train	S	
The ticket inspector comes/she needs to show her ticket/she gets out her ticket	IE	
She opens her bag/she puts the ticket away/she looks in her bag	A	
She sees her (unopened) packet of biscuits in the bag	C	
The old lady realises she was eating the teenager's biscuits/realises that he was actually being generous	IR	
The old lady is regretful/sorry/feels bad for getting angry with him	IR	
The train leaves/The teenager walks away/puts the empty packet in the bin	R	

Video B:

Main story events	Story grammar code	Included in narrative? (0/1)
Episode 1		
A boy/Joe/ a group of boys/Year 7s	S	
Playing football/outside/on the playing field/at breaktime.	S	
The bell rings/it's time to go to class/Joe's friends say they need to go	IE	
Joe keeps playing/falls over/gets tackled	A	
He gets mud on his shirt/gets his uniform dirty	C	
Joe's friend offers his water bottle to clean the shirt/his friend tries to help	A	
Joe refuses/says no/doesn't want it to look like he's wet himself	A	
Joe is worried about his teacher's reaction/doesn't want to get in trouble	IR	
He decides to clean his shirt before going to his lesson	P	
Episode 2		
Joe goes into the school/is in the school corridor/hallway	S	
He is lost/ he can't find his classroom/ he doesn't know where to go	IE	
He asks a group of girls for directions to his classroom	P	

The girls give him the wrong directions/make fun of him	A	
Joe is fed up/upset/disappointed	IR	
An older student approaches Joe and tells him where to go / Someone else helps him by telling him the right directions.	R	
Episode 3		
Joe goes to his classroom/enters the classroom	S	
He is late to the lesson /The teacher/Mr Drew asks Joe why he is late	IE	
Joe tries to apologise/says sorry	P	
Mr Drew makes fun of Joe in front of the class/ Mr Drew is mean to Joe/ The teacher tells him off	A	
The whole class laughs at Joe	C	
Joe is annoyed/angry/upset	IR	
Joe mutters “get lost” to the teacher/tells Mr Drew to “get lost”	A	
Mr Drew is cross/annoyed	IR	
Mr Drew asks Joe to repeat what he said	C	
Joe pretends he said that he “got lost”	A	
Mr Drew tells Joe to sit down/ Joe goes to his seat.	R	

Appendix C: Narrative coherence ('6Cs') scoring framework

	0	1	2	3
Context	No attempt to provide any information about the time setting or location of the events described.	Some aspect of setting is mentioned during the narrative, but lacks specificity (e.g., "they're outside"; "it's late") OR: Narrator provides enough contextual information that the setting can be correctly inferred (minimum of <u>TWO</u> contextual clues).	Narrator provides <u>specific</u> information about either location or time setting (e.g. "at the train station"; "in the classroom", "it's break-time") OR: Narrator provides <u>non-specific</u> information about both location and time.	Narrator orients the listener to both the location and the time setting (and at least one of these is specific).
Characterisation	Characters are underdeveloped. No use of internal state terms in the narrative (i.e., cognitive or emotional)	Some limited information is provided about the motivations of the characters (i.e., at least one internal state term is used in the narrative to describe thoughts or feelings).	A range of internal state terms are used in the narrative to describe thoughts or feelings. The motivations of the main or supporting characters are evident.	More than one character in the narrative is well-developed, with a range of internal state terms providing information about their thoughts or feelings. The motivations of the main and/or supporting characters are evident.
Chronology	Story events do not follow a logical chronological sequence.	The narrative follows a logical chronological sequence, but a significant number of story events are omitted (<u>less than 10 main events reported in total</u>).	Story events follow a logical chronological sequence, with any timeline violations explained by the narrator. However, the narrative ends abruptly and lacks a clear resolution.	Story events follow a logical chronological sequence, with any timeline violations explained by the narrator. The narrative includes a clear ending or resolution.
Causality	No attempt to link cause and effect in the narrative. Even if story events are listed in the correct order, it is not clear how they are related (e.g., consequences are presented without the causes being mentioned).	There are some identifiable instances of cause and effect in the narrative (e.g., characters' actions are followed by direct consequences). However, the narrator does not use any causal conjunctions to highlight these relationships.	The narrator attempts to highlight cause and effect relationships between story events. At least one causal conjunction is used in the narrative (e.g., 'so', 'because', 'since', 'as', 'even though', 'consequently', 'therefore', 'as a result', 'as a	The narrator uses a range of causal conjunctions to explain cause and effect relationships between key story events (e.g., 'so', 'because', 'since', 'as', 'even though', 'consequently', 'therefore', 'as a result'...)

			result'...)	
Cohesion	Referencing is unclear, making the story difficult to follow.	Referencing is somewhat unclear, but despite the presence of ambiguous references, the listener is still able to follow the story easily.	Referencing is generally clear, with no more than two ambiguous references.	Referencing is accurate throughout the entire story; (i.e., all pronouns can be traced back to an appropriate antecedent).
Congruence	Off-topic, bizarre or extraneous utterances impede comprehension of the story.	Presence of more than one off-topic, bizarre or extraneous utterance; however, these comments do not impede overall comprehension of the story.	Utterances are generally on-topic, with not more than one off-topic or bizarre remark.	All utterances are pertinent to the events being described in the narrative.

Appendix D: Correlations between study variables

	1.	2.	3.	4.	5.	6.	7.	8.
1. Structure (video A)	-							
2. Structure (video B)	$r = .648$, $p < .001^{**}$ *	-						
3. Coherence (video A)	$r = .613$, $p < .001^{**}$ *	$r = .475$, $p < .001^{**}$ *	-					
4. Coherence (video B)	$r = .414$, $p < .001^{**}$ *	$r = .689$, $p < .001^{**}$ *	$r = .459$, $p < .001^{***}$	-				
5. Mentalising	$r = .456$, $p < .001^{**}$ *	$r = .480$, $p < .001^{**}$ *	$r = .395$, $p < .001^{***}$	$r = .488$, $p < .001^{**}$ *	-			
6. WM	$r = .273$, $p = .007^{**}$	$r = .196$, $p = .053$	$r = .346$, $p < .001^{**}$	$r = .282$, $p = .005^{**}$	$r = .333$, $p < .001^{***}$	-		
7. Inhibition	$r = -.132$, $p = .195$	$r = -.174$, $p = .087$	$r = -.097$, $p = .344$	$r = -.195$, $p = .054$	$r = -.119$, $p = .242$	$r = -.056$, $p = .586$	-	
8. Shifting	$r = .068$, $p = .505$	$r = .050$, $p = .622$	$r = .077$, $p = .453$	$r = .138$, $p = .176$	$r = .191$, $p = .060$	$r = .194$, $p = .055$	$r = .093$, $p = .360$	-
9. Generativity	$r = .298$, $p = .003^{**}$	$r = .213$, $p = .035^*$	$r = .261$, $p = .009^{**}$	$r = .158$, $p = .119$	$r = .258$, $p = .010^*$	$r = .375$, $p < .001^{**}$ *	$r = -.130$, $p = .201$	$r = .136$, $p = .183$

Appendix E: Exploratory analyses excluding non-autistic participants with elevated SRS-2 scores (N=13)

Table 1. Mean (SD) scores for control variables (age, non-verbal cognitive ability, receptive vocabulary and expressive language) and SRS-2 scores for autistic and non-autistic groups, with group differences.

Variables	Autistic group (N=44)	Non-autistic group (N=41)	Group differences
Age (months)	159.98 (16.75)	158.32 (16.21)	$t(83) = -.46, p = .644$
Non-verbal ability: WASI-II - Matrix Reasoning (T-scores: $M = 50, SD = 10$)	50.82 (9.05)	54.78 (9.68)	$t(83) = 1.95, p = .054$
Receptive vocabulary: BPVS-3 (standardised scores: $M = 100, SD = 15$)	102.73 (15.47)	106.78 (11.86)	$t(83) = 1.35, p = .181$
Expressive language: CELF-5 UK - Recalling Sentences (scaled scores: $M = 10, SD = 3$)	10.14 (2.70)	11.61 (3.27)	$t(83) = 2.27, p = .026^*$
Autistic traits: Social Responsiveness Scale-2 (T-scores: $M = 50, SD = 10$)	77.95 (10.48)	46.15 (5.13)	$t(83) = -17.95, p < .001^{***}$
<i>Note: T-scores >59 are indicative of social communication difficulties</i>			

Narrative structure (story grammar)

Video A

Step 1 of the regression model (control variables) was statistically significant, $R^2 = .115$, $F(4, 80) = 2.607$, $p = .042$, adj. $R^2 = .071$, although expressive language was the only individual variable that was significant ($\beta = .253$, $p = .046$).

Step 2 of the model was also statistically significant, $R^2 = .319$, $F(7, 77) = , p < .001$, adj. $R^2 = .257$, explaining an additional 20% of unique variance once the control variables had been accounted for ($\Delta R^2 = .203$). Mentalising significantly predicted story grammar scores for Video A ($\beta = .434$, $p < .001$), as did EF1 scores ($\beta = .283$, $p = .024$). However, EF2 scores were not significant ($\beta = -.025$, $p = .793$).

In the final model (Step 3), the inclusion of Group significantly predicted 2% of additional variance overall, $\Delta R^2 = .019$, $F(8, 76) = , p < .001$, although Group was not an individually significant predictor (Group: $\beta = -.151$, $p = .146$). Mentalising remained significant as a predictor variable ($\beta = .410$, $p < .001$), but EF1 did not retain its significance in this final step ($\beta = .230$, $p = .074$). See **Table 2**.

Table 2. Hierarchical multiple regression predicting total narrative structure ('story grammar') scores for Video A from control variables, Mentalising and EF scores, and diagnostic group (final model).

Variables	B (95% CI)	SE B	β	p
Constant	9.772 (-3.104 - 22.648)	6.465	-	.135
Age	.022 (-.033 - .078)	.028	.079	.424
Receptive vocabulary	-.039 (-.126 - .048)	.044	-.117	.376
Expressive language	.064 (-.309 - .438)	.187	.043	.733
Non-verbal ability	-.021 (-.133 - .090)	.056	-.044	.703
Mentalising**	.530 (.242 - .818)	.145	.410	<.001***
EF1	1.046 (-.103 - 2.195)	.577	.230	.074
EF2	-.112 (-.966 - .742)	.429	-.025	.794

Group*	-1.392 (-3.279 - .496)	.948	-.151	.146
--------	------------------------	------	-------	------

Note: N=85. $R^2 = .115$ for Step 1 ($p = .042^*$). $\Delta R^2 = .203$ for Step 2 ($p < .001^{***}$). $\Delta R^2 = .019$ for Step 3 ($p < .001^{***}$).

Video B

Step 1 of the regression model (control variables) was statistically significant overall, $R^2 = .119$, $F(4, 80) = 2.702$, $p = .036$, adj. $R^2 = .075$, although none of the control variables were individually significant.

Step 2 of the model was also significant, $R^2 = .286$, $F(7, 77) = 4.397$, $p < .001$, adj. $R^2 = .221$, explaining an additional 17% of unique variance once the control variables had been accounted for ($\Delta R^2 = .167$). Mentalising was the only significant predictor of story grammar scores for Video B ($\beta = .449$, $p < .001$), while EF scores were not significant (EF1: $\beta = .125$, $p = .322$; EF2: $\beta = -.062$, $p = .533$).

In the final model (Step 3), the inclusion of Group significantly predicted 12% of additional variance overall, $\Delta R^2 = .119$, $F(8, 76) = 6.459$, $p < .001$ (Group: $\beta = -.381$, $p < .001$), with Mentalising remaining significant as a predictor variable ($\beta = .389$, $p < .001$). See **Table 3**.

Table 3. Hierarchical multiple regression predicting total narrative structure ('story grammar') scores for Video B from control variables, Mentalising and EF scores, and diagnostic group (final model).

Variables	B (95% CI)	SE B	β	p
Constant	7.914 (-3.147 - 18.975)	5.554	-	.158
Age	.019 (-.028 - .067)	.024	.076	.421
Receptive vocabulary	.002 (-.073 - .077)	.037	.007	.958

Expressive language	.045 (-.276 - .366)	.161	.033	.780
Non-verbal ability	.003 (-.093 - .099)	.048	.007	.949
Mentalising***	.456 (.208 - .704)	.124	.389	<.001***
EF1	-.027 (-1.014 - .960)	.496	-.007	.957
EF2	-.247 (-.981 - .486)	.368	-.061	.504
Group***	-3.176 (-4.797 - -1.554)	.814	-.381	<.001***

Note: N=85. $R^2 = .119$ for Step 1 ($p = .036^*$). $\Delta R^2 = .167$ for Step 2 ($p < .001^{***}$). $\Delta R^2 = .119$ for Step 3 ($p < .001^{***}$).

Narrative coherence

Video A

Step 1 of the regression model (control variables) was non-significant, $R^2 = .078$, $F(4, 80) = 1.686$, $p = .161$, adj. $R^2 = .032$.

Step 2 of the model was statistically significant, $R^2 = .275$, $F(7, 77) = 4.179$, $p < .001$, adj. $R^2 = .209$, explaining an additional 20% of unique variance once the control variables had been accounted for ($\Delta R^2 = .198$). Mentalising significantly predicted narrative coherence scores for Video A ($\beta = .385$, $p = .001$), as did EF1 scores ($\beta = .332$, $p = .010$). However, EF2 scores were not significant ($\beta = -.073$, $p = .465$).

In the final model (Step 3), the inclusion of Group significantly predicted 2% of additional variance overall, $\Delta R^2 = .019$, $F(8, 76) = 3.954$, $p < .001$, although Group was not an individually significant predictor (Group: $\beta = -.150$, $p = .161$). Mentalising and EF1 scores both remained significant as predictor variables (Mentalising: $\beta = .361$, $p = .003$; EF1: $\beta = .280$, $p = .036$). Non-verbal cognitive ability also significantly predicted narrative coherence scores in this final model ($\beta = -.249$, $p = .040$).

See **Table 4**.

Table 4. Hierarchical multiple regression predicting total narrative coherence scores for Video A from control variables, Mentalising and EF scores, and diagnostic group (final model).

Variables	B (95% CI)	SE B	β	<i>p</i>
Constant***	13.485 (6.267 - 20.704)	3.624	-	<.001***
Age	-.006 (-.037 - .025)	0.16	-.036	.724
Receptive vocabulary	.014 (-.035 - .062)	.024	.076	.579
Expressive language	-.068 (-.278 - .141)	.105	-.083	.517
Non-verbal ability*	-.066 (-.128 - -.003)	.031	-.249	.040*
Mentalising**	.254 (.092 - .415)	.081	.361	.003**
EF1*	.691 (.047 - 1.336)	.323	.280	.036*
EF2	-.177 (-.656 - .302)	.240	-.073	.464
Group	-.751 (-1.810 - .307)	.531	-.150	.161

Note: N=85. $R^2 = .078$ for Step 1 ($p = .161$). $\Delta R^2 = .198$ for Step 2 ($p < .001^{***}$). $\Delta R^2 = .019$ for Step 3 ($p < .001^{***}$).

Video B

Step 1 of the regression model was statistically significant, $R^2 = .212$, $F(4, 80) = 5.372$, $p < .001$, adj. $R^2 = .172$; however, receptive vocabulary was the only control variable that individually predicted narrative coherence scores for Video B ($\beta = .301$, $p = .024$).

Step 2 of the model was also statistically significant, $R^2 = .316$, $F(7, 77) = 5.076$, $p < .001$, adj. $R^2 = .254$, explaining an additional 10% of unique variance once the control variables had been accounted for ($\Delta R^2 = .104$). In this step, Mentalising was the only significant predictor of narrative coherence scores for Video B ($\beta = .345$, $p = .003$), with receptive vocabulary becoming non-significant ($\beta = .192$, $p = .147$). EF scores were not significant (EF1: $\beta = .116$, $p = .349$; EF2: $\beta = -.066$, $p = .498$).

In the final model (Step 3), the inclusion of Group significantly predicted 4% of additional variance overall, $\Delta R^2 = .039$, $F(8, 76) = 5.234$, $p < .001$ (Group: $\beta = -.219$, $p = .034$), with Mentalising remaining significant as a predictor variable ($\beta = .311$, $p = .006$). See **Table 5**.

Table 5. Hierarchical multiple regression predicting total narrative coherence scores for Video A from control variables, Mentalising and EF scores, and diagnostic group (final model).

Variables	B (95% CI)	SE B	β	p
Constant	3.353 (-3.909 - 10.615)	3.646	-	.361
Age	.018 (-.013 - .049)	.016	.113	.248
Receptive vocabulary	.044 (-.005 - .093)	.025	.230	.080
Expressive language	-.011 (-.222 - .199)	.106	-.013	.915
Non-verbal ability	.001 (-.062 - .064)	.032	.004	.971
Mentalising**	.230 (.067 - .392)	.082	.311	.006**
EF1	.104 (-.544 - .752)	.325	.040	.750
EF2	-.167 (-.649 - .315)	.242	-.065	.492
Group*	-1.153 (-2.218 - -.088)	.535	-.219	.034*

Note: N=85. $R^2 = .212$ for Step 1 ($p < .001^{***}$). $\Delta R^2 = .104$ for Step 2 ($p < .001^{***}$). $\Delta R^2 = .039$ for Step 3 ($p < .001^{***}$).