SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

DEPARTMENT OF PSYCHOLOGY

DOCTORAL PROGRAM IN

PSYCHOLOGY LINGUISTICS AND COGNITIVE NEUROSCIENCE

(XXXV Cycle)

CURRICULUM: THE MIND BRAIN AND BEHAVIOUR

The relation between Executive Function and Narrative Competence in Early Childhood

Doctoral Thesis by

Nicoletta Scionti

Supervisors

Prof. Gian Marco Marzocchi

Prof. Laura Zampini

Coordinator

Prof. Simona Sacchi

ACADEMIC YEARS 2019-2022

Declaration

I declare that the work presented in this thesis is my own.

Where information has been derived from other sources, I confirm that this has been reported in the thesis.

____

Nicoletta Scionti

In loving memory of

My Dad

Index

Abstract

Executive function (EF) and narrative competence (NC) are two fundamental skills in human life. Both dimensions have been extensively investigated, but research on their relationship is still scarce.

The dissertation presents three studies conducted within this PhD project, aiming to deepen the knowledge of the relationship between NC and EF in Italian preschoolers.

The first chapter presents a meta-analysis on the association between these dimensions across childhood and adolescence. Results showed that the strength of this association changes over time and across different types of narrative and executive competence. Early childhood resulted in the period of life where EF and NC are more associated.

Following recent evidence that narrative assessment may serve to address EF in the adult population (Cannizzaro & Coelho, 2013), the second chapter aims to analyze which indices of narrative performance could be useful to address EF in children. Narrative performances were extensively investigated in a sample of 39 preschoolers, and the relation with EF tasks was analyzed taking into consideration possible confounding variables. Results showed that anticipations and anaphoric use of articles within oral narratives might be useful as an ecological measure of a child's working memory and interference control, respectively.

Finally, the third chapter presents a study on the differences between children and adults in referential cohesion, the ability to adjust referential forms by discourse function (introduction, maintenance, re-introduction of characters) when referring to more or less accessible character entities in a narrative context. Then, cognitive mechanisms underlying children's referential choices that diverged from adults were investigated. Results showed the fundamental role of interference control processes in referential choices made by children.

Introduction

Narratives represent an essential device for human communication, and the onset of the use of narratives represents a critical step in language development studies. Evaluation of children's oral narratives is of significant interest to researchers and practitioners, as being a proficient narrator is an important skill in children's and adults' lives. Oral narrative skills are a key component of most school curricula, and several studies support the importance of narrative abilities to academic and social success for both typically developing children and children with language and learning disabilities (Fazio et al., 1996; O'Neal et al., 2004). Extant research reports that good narrative skills are positively associated with structural language, literacy and social skills (Dickinson & McCabe, 2001; Griffin et al., 2004; Johnston, 2008).

Telling stories is a multi-componential complex competence. It requires the child to be able to consider and integrate linguistic, cognitive, and social abilities (Boudreau & Chapman, 2000; Karmiloff & Karmiloff-Smith, 2002). Among cognitive abilities implied, the contribution of executive functions (EF) within narration is not clear. Narrative discourse is organised around a typical story structure (i.e., the story's grammar), which includes an introduction, a provision of the setting and description of the characters in the story, a problematic situation that shapes the protagonist's goal, attempts to solve the problem, and a conclusion (e.g., Pinto, 2003; Stein, 1988; Stein, Glenn & Freedle, 1979). Telling well-structured stories was found to be related to EF in adults, and this leads some authors to claim that narrative assessment might be valid as an ecological measure of EF. EF is an umbrella term encompassing a broad set of higher-order cognitive abilities implied in goal-directed behaviour. It enables us to alter overlearned behavioural patterns when they become unsatisfactory, understand complex or abstract concepts, solve new problems, and manage relationships (Van der Linden et al., 2000). Like narrative skills, EF is a predictor of great relevance to many developmental outcomes. A large body of research has demonstrated substantial links between EF and academic achievement, literacy, health, wealth, and criminality (Moffitt et al., 2011) in children of various ages with and without neurodevelopmental disorders (see Best et al., 2009; Müller et al., 2008, for reviews). Despite their undoubted importance in human development, EF has been traditionally quite difficult to define and operationalize. There is no unanimous agreement on which domains to include in the construct of EF. It seems to be initially unitary or undistinguishable (e.g., Wiebe et al., 2011) and then differentiated across development, but when and how EF differentiates is still unclear. In adults and 8- to 13-years children, three specific core domains were mostly identified: inhibition, updating of working memory and shifting (Miyake et al., 2000; Lehto et al., 2003). However, research with younger children usually yields a smaller number of factors. Especially for preschool age, the debate on the structure of EF is still open. Furthermore, another issue concerning this construct is that no single behaviour can be directly tied to EF, and there are no "pure measures" tapping EF skills.

The poor ecological validity is a relevant issue in EF assessment. Performance-based tests seem unable to capture children's EF processes relevant to real-world outcomes (Gioia et al., 2008). Informant-based rating scales supposed to address the ecological validity of EF are weakly related to performance at the EF task. Moreover, the results of rating scales may be biased by several factors as the informant characteristics and the context in which different informants judge child behaviour (Barkley, 2006).

Twenty years ago, Grafman and Litvan (1999) proposed that experimental tasks may be promising tools for exploring these complex behaviours due to their close association with real-life activities. In this light, this dissertation aims to examine if there are premises to suppose that the assessment of narrative competence (NC) may represent an ecologically valid way to explore the workings of EF in children. To do that, we deepen the knowledge of the relationships between EF and NC across development.

There are several reasons to expect that NC and EF are associated. For instance, the development in NC emerges in concert with developments in EF. Evidence from imaging studies indicates that these skills depend upon overlapping neural substrates, mainly frontal lobe function, and deficits across these skill sets are observed in adults with traumatic brain injuries (Mar, 2004; Coelho et al., 1995).

Telling a good story requires the individual to set the goal of linking all of the story elements in a coherent manner, retrieving the appropriate semantic information, syntactic structures, and morphological features that would express the causal links between various story elements. It also requires to indicate the characters' motivations and reactions and monitoring the narrative while it is being produced. To tell a coherent story, children need to set up a hierarchical goal and plan and monitor the organization of the narrative events around the story schema. Especially, the story's episodic structure (goal – attempt – outcome) represents a type of goal-directed thinking and action similar to knowledge stored and processed by adults in the prefrontal cortex, which psychologists use to refer with the general meaning of EF. Furthermore, the capacity of individuals to refer to characters within the story in a cohesive way is an aspect that may be influenced by EF. For instance, the capacity to re-introduce and mark given referents appropriately within narration requires the child to consider the referent's accessibility from the listener's perspective and seems to be associated with working memory and inhibitory processes (Hendriks et al., 2014; Kuijper et al., 2015). Indeed, in order to re-introduce a character that is no more topical within the narrative discourse, the narrator has to keep track of the character's topicality within the discourse and inhibit pronominal forms to use more explicit referencing expressions (full noun phrases) that make the listener easier to assign the referent unambiguously.

Despite all these reasons to expect EF and NC to be related across development, evidence on their relationship is unclear, and research addressing the question in childhood is scarce.

Therefore, the first chapter of this dissertation presented a meta-analytic study we conducted to determine the overall strength of the relationship between NC and EF across childhood, if and when the strength of this relationship changes over development. Furthermore, we examine potential moderators to understand if the strength of the relation varies between typically- vs atypically developing children by different EF domains and narrative levels. The work aimed to clarify the state of the art and orient the research on this topic.

The second chapter gets to the heart of the matter and presents a study on the indices of narrative performance that might be useful to address children's EF working in an ecological way. Several indices of narrative performance were collected in a sample of 39 preschoolers, and through a regression method, we explored which indices were more related to performance at traditional EF tasks.

Finally, the third chapter deepens into the relationship between EF and referential cohesion, a crucial aspect of NC. The study compares referential choices made by children vs adults across narrative discourse and analyses some EF mechanisms underlying referential choices made by children within narratives.

CHAPTER 1

The relation between narrative skills and executive functions across childhood: a systematic review and meta-analysis

1.1 INTRODUCTION

Narratives represent an essential device for human communication and are a vehicle for cultural transmission.

The onset of the use of narratives represents a critical step in language development studies, where storytelling represents a real and contextualized request for children. Therefore, it is seen by many authors as a "naturalistic" approach to studying language development (Schraeder et al., 1999). Evaluation of children's oral narratives is of significant interest to researchers and practitioners, as being a proficient narrator is an important skill in children's and adults' lives. Oral narrative skills are a key component of most school curricula, and several studies support the importance of narrative abilities to academic and social success for both typically developing children and children with language and learning disabilities (Fazio et al., 1996; O'Neal et al., 2004). Extant research reports good narrative skills are positively associated with structural language, literacy and social skills (Dickinson & McCabe, 2001; Griffin et al., 2004; Johnston, 2008).

Telling stories is a multi-componential complex competence. It requires the child to be able to plan and execute their production of the story's plotline by using appropriate vocabulary, grammar, and syntax. Studies on the development of narrative skills have identified that stories have a typical structure, or story grammar (Stein & Glenn, 1979), following a "schema" that children and adults use to understand, interpret, and produce stories. According to Stein and Glenn's (1979) story-grammar model, stories must include a setting and an episode system at a minimum. An episode consists of an introduction, a provision of the setting and description of the characters in the story, a problematic situation that shapes the protagonist's goal, attempts to solve the problem, and a conclusion (e.g., Pinto, 2003; Stein, 1988; Stein et al., 1979). Stories may also include multiple episodes organized in a linear or a hierarchical manner, resulting in more complexity (e.g., stories with multiple embedded episodes within a particular story arc).

Developmental studies reveal that the acquisition of narrative proficiency is a slow process, which emerges in the preschool years and is not fully developed until adulthood (Berman & Slobin, 2013). In early childhood, there is a disproportionate emphasis on characters' actions in narratives without a link to the plot line (van den Broek et al., 1996). At 2 years, narratives are descriptions of character actions, and labels posited without a link to a central theme. Between 3 and 4 years, narratives generally include some local connections between adjacent story events and simple inferences across the story episode. At 4 years, children begin to use structural components of narratives, which generally include causal connections between events. However, until 5, children still show difficulty conceiving an overall plot or overarching goal.

It is not until 6-7 years old that children are able to produce "true narratives". At this age, their narratives follow a logical progression of events, including sub-plots and understanding of time frames. After 7 years old, narratives are generally well-structured. Progress in literacy acquisition seems to play a significant role in this passage. The narrative generation process is thought to draw critically on reading skills. For example, Abbott and Berninger (1993) found that reading contributes significantly to the quality of narrative composition for children in the first three grades.

Empirical findings suggest that reading and writing draw on shared knowledge yet are separate skills with distinct developmental trajectories (Berninger et al., 2002; Fitzgerald & Shanahan, 2000). In a study with 120 third-grade children, Olinghouse (2008) found that reading skills directly influenced compositional quality. There are aspects of continuity and discontinuity in the transition from oral to written narrative composition during this period. Studies on typically developing children provide evidence that children who master writing preserve their narrative skills in the transition between the codes (Bigozzi & Vettori, 2016). However, for those children who did not master it efficiently (e.g., children with learning disabilities and other neurodevelopmental disorders), written narrative composition becomes an obstacle.

At 8–10 years, children generally manage structural components correctly and demonstrate to know how to tell a story to another person. After 10 years, narratives are more complex, detailed and structurally coherent. Children use various linking devices (e.g., prepositions, conjunctions and adverbs) and demonstrate more effort to engage the listener's attention and adapt to different audiences.

Across development, oral and written narratives can be studied at the macro- and microstructure levels. Microstructure refers to specific features of the language used to convey ideas, including the use of decontextualized language and grammatical complexity (e.g., Justice et al., 2010; Mäkinen et al., 2014). In contrast, macrostructure refers to global narrative features, particularly the ability to produce a story that is overall well structured, coherent, and cohesive. During development, a remarkable increase involves the macrostructural level (e.g., Castilla-Earls et al., 2015), particularly in the transition from preschool- to school-age (e.g., Roch et al., 2016; Zanchi & Zampini, 2021). This period is characterized by the rapid qualitative increase of executive functions (EF).

EF refers to a broad set of neurocognitive processes underlying goal-directed control of thought, behaviour, and emotion that allow for adaptation to environmental demands (Shonkoff & Phillips, 2000). Like narrative skills, EF are predictor of great relevance to many developmental outcomes. A large body of research has demonstrated substantial links between EF and academic achievement, literacy, health, wealth, and criminality (Moffitt et al., 2011) in children of various ages with and without neurodevelopmental disorders (see Best et al., 2009; Müller et al., 2008, for reviews).

There is no unanimous agreement around which domains include the construct of EF. Scholars studying EF deal with the problem that EF are initially unitary or undistinguishable (e.g., Wiebe et al., 2011), but they differentiate across development. To date, when and how they differentiate is still unclear. In the adult population, three specific core domains were mostly identified: inhibition, updating of working memory and shifting (Miyake et al. 2000). This finding was replicated in research with 8- to 13-years children (see Lehto et al., 2003). However, research with younger children usually yields a smaller number of factors. Especially for preschool age, the debate on the structure of EF is still open. This period is the most critical for the rapid changes occurring in child neurodevelopment. So far, some studies have found a single factor for all EF (Wiebe et al., 2011), and other studies have proposed a two-factor model instead (Monette et al., 2015; Scionti & Marzocchi, 2021; Usai et al., 2014).

Furthermore, studies on children differ from studies on adults in broader processes of defining EF. For instance, Diamond (2013) includes working memory and cognitive flexibility instead of updating and set-shifting, which are more specific processes. Indeed, working memory here refers to a domain-general system that can store and process information simultaneously. It shows a linear increase from ages 4 to 14 and a levelling off between ages 14 and 15 (Best & Miller, 2010). In contrast, updating is the specific ability to change temporarily stored information in the light of incoming information and is mainly investigated in studies with adults and older school-aged children. Developmental studies showed that updating increases with age along with upgrading of inhibition efficiency and stabilize by the age of 15 years (Carriedo et al., 2016). Cognitive flexibility refers to a tendency to perform in ways that are not fixed or routine, to "think outside the box", or to adapt to changes in the environment; instead, shifting refers to the ability to switch between conflicting operations or different tasks sets. Shifting is a more specific dimension than "cognitive flexibility". However, some authors pointed out that there is no evidence that cognitive flexibility can be considered a general, coherent construct usable in individual-difference research with children (Morra et al., 2017). Very often, the term "cognitive flexibility" in developmental studies is actually used with the meaning of "shifting" (e.g., Monette et al., 2015). The development of successful shifting seems to depend on inhibition and working memory. As Garon et al. (2008) noted, before children can successfully shift between response sets, they must be able to maintain a response set in working memory and then be able to inhibit the activation of a response set to activate an alternative one. Developmental studies reveal that shifting improved from age 4 to adolescence, reaching adult-like levels around 15 (Davidson et al., 2006).

Other authors included different types of inhibition in EF definition, distinguishing inhibition on a behavioural level (response inhibition or behavioural inhibition) and a cognitive/attention level (interference suppression or interference control), both sharing the need to suppress an action or a thought in order to control impulses and stay focused (Diamond, 2013; Nigg, 2017). Studies on their development reveal that at 4 years, these two inhibition processes are already distinguishable (Gandolfi et al., 2014). Improve behavioural inhibition tends to stabilize by the early school years (i.e., from 5-to 8-years; Lehto et al., 2003), whereas a sensitive increase in interference control occurs during elementary school and is followed by slower improvement during early adolescence (Best & Miller, 2010).

Furthermore, with increasing age, complex high-order EF as planning and problem solving become relevant to be included in the construct of EF (Diamond, 2013). They develop particularly late in childhood and undergo a final growth spurt during the beginning of adolescence (Anderson et al., 2001; Welsh & Pennington, 1988). Research on these processes has examined chiefly the development of performance at Tower-like task across different age groups and found age effects only for the more complex problems (McCormack & Atance, 2011).

1.1.1 NC and EF: Are they linked?

There are different reasons to expect that EF and NC are related across development. In general, literature frequently reported significant relationships between EF and different aspects of language skills. Especially during the preschool years, language skills undergo rapid development: vocabulary overgrows, the use of syntactic rules becomes more adult-like, and the ability to use language in narratives improves (e.g., Odato, 2013; Song et al., 2015; Tomasello, 2000). At the same time, the preschool years are characterized by a substantial improvement in EF that are commonly impaired in children with language disorders (e.g., Gooch et al., 2016).

The fact that developments in NC emerge in concert with developments in EF suggests a potential developmental relationship between these abilities. Evidence from imaging studies indicates that these skills depend upon overlapping neural substrates, mainly frontal lobe function, and deficits across these skill sets are observed in adults with traumatic brain injuries (Mar, 2004; Coelho et al., 1995).

Telling a good story requires the individual to set the goal of linking all of the story elements in a coherent manner, retrieving the appropriate semantic information, syntactic structures, and morphological features that would express the causal links between various story elements, and also indicate the characters' motivations and reactions, and monitor the narrative while it is being produced. In order to tell a coherent story, children need to set up a hierarchical goal and plan and monitor the organization of the narrative events, and this seems to engage EF (Mozeiko et al., 2011):

- *shifting* may be involved in the generation of complete episodes within a narrative discourse, in the selection of informative words, and in the ability to monitor the communicative flow;
- *updating of working memory* may be required to generate and understand sentences as well as recall episodic contents for an accurate organization of a story;
- *inhibition processes* may be critical for monitoring the production of extraneous comments and derailments while telling a story and for the ability to inhibit the semantic competitors while producing words;
- *planning and more complex EF* may be recruited to the extent of coordinating all the processes involved as well as for planning and goal setting of the story (e.g., retelling a narrative containing all of the story elements in the correct sequence; Khan, 2013).

In the same way, NC development may support performance on EF tasks. This seems especially plausible on tasks with long and complex instructions and linguistic stimuli to be processed or producing oral responses (Tonér & Nilsson Gerhold, 2021).

However, both cross-sectional and longitudinal studies are inconsistent regarding the association, and potential causal relation, between EF and NC. For instance (Khan, 2013), in a study on children between 3 and 6 years old, results showed that narrative production is best predicted by high-level EF, measured with planning and shifting tasks. In contrast, other studies investigating the relationship between these domains in 4-5 and 7-8 Turkish children found that narrative production, especially plot complexity, is related to these EF only in the older group, not in the younger age band (Balaban, 2020). Moreover, other studies report no association between planning skills and the quality of written narratives in fourth-grade children (Drijbooms et al., 2017).

A significant relationship between microstructural competence, such as lexical variety and syntax used in narratives, and shifting ability, addressed by the performance at card sorting task, is found in a sample of 47 four-to six-year-old Swedish children. In the same way, EF accounts for 7% of the variation in syntactic complexity in Turkish-speaking preschoolers (Balaban, 2020). Longitudinal research on school-aged Dutch children reveals that the development of syntactic complexity in narratives between fourth and sixth grade is also predicted by planning and behavioural inhibition in fourth grade (Drijbooms et al., 2017). The relationship between syntactic complexity and inhibitory skills is not found at preschool age in typically developing Swedish children (Tonér & Nilsson Gerhold, 2021).

Research on the role of working memory in narratives appears more consistent. A study on children aged 5 to 8 shows that the ability to update working memory is moderately associated with referential adequacy, the macrostructural competence to introduce and maintain a reference to story characters in narratives (McNiven, 2007). Studies on children aged 8 to 11 reveal that working memory and shifting significantly account for plot complexity variance, another macrostructural NC indicator, in written narratives (Balioussis et al., 2013). Even when controlling for vocabulary, working memory correlates with text generation at the word, sentence and text level in a sample of 10-years-old children (Puranik, 2006) and adolescents (Swanson & Berninger, 1996b). According to the authors, it may be involved in translating ideas in memory into linguistic representation, organizing thoughts into temporally sequenced discourse and revising text.

In general, studies on narrative writing show that children with higher updating and inhibitory skills produce longer, coherent narratives. The authors (Swanson & Berninger, 1996b) explain the involvement of these processes by the need to suppress inappropriate lexical representations, select the relevant ones and actively hold and update the representations in WM during writing composition. However, some studies on 5 and 6-year-old children with SLI found a significant correlation between narrative retelling skills and working memory, but not with inhibitory processes (Dodweel & Bavin, 2008; Ketelars et al., 2012).

Furthermore, some studies fail to find a direct relationship between NC and inhibitory and WM updating skills, showing that the influence of these EF domains on NC may totally depend on handwriting skills (Drijbooms et al., 2015). Indeed, studies reported that children with poor handwriting skills tend to use the first linguistic expression that occurs to them to frame their ideas without being concerned about shaping the linguistic expression in response to narrative demands or the reader's needs (Salas & Silvente, 2020; Swanson & Berninger, 1996a; Artico & Penge, 2016). They must devote most or all of their cognitive effort to spelling and handwriting, leaving little resources available for other writing processes. This may limit the amount and quality of text they can generate.

In sum, there is conflicting evidence about the developmental stages at which EF relates to NC. Inconsistent results suggest that the development of these skills can be heterochronous with ones that are deeply conceptually related and developing on different timescales. Even though they develop across the preschool period, it seems they do not do in lockstep. Some aspects of EF may develop before others, and the relation between these aspects and NC may be such that there is specificity in predictive relations over developmental time for microstructural and macrostructural elements (Friend & Phoenix-Bates, 2014). Research with atypically developing populations presenting deficit in both EF and NC show similar inconsistent results. For instance, in children with a diagnosis of ADHD and language impairment, Fernandez et al. (2010) found a significant correlation between macrostructural elements produced in the narration (e.g., episodic structure) and planning skills, but not with phonological working memory. Some studies conducted in children with SLI, instead, found significant association between plot structure and phonological working memory (Dodwell & Bavin, 2008; Duinmeijer et al., 2012).

To date, our understanding of how and when different aspects of NC relate to EF – or which part of EF they relate to – is limited. Integration of divergent findings has become a necessary and important task. The present study takes up this task using a meta-analytic approach in order to examine and explain the variability across findings. Larger sample approaches may indeed improve our knowledge on the relation between EF and NC over developmental time and orient future research on this topic. Currently, to our knowledge, there are no systematic reviews or meta-analyses addressing this issue.

The understanding of how different aspects of NC relate to EF – or which part of EF they relate to – is also clinically relevant since both the skills predict important life outcomes (i.e., academic and social success) and are trainable (Abel et al., 2015; Diamond & Lee, 2011; Scionti et al., 2020; Petersen, 2011). Studies show that children – especially those at risk (e.g., children from backgrounds of poverty, children whose first language is not the one spoken in the country where they live, or children with psychopathological traits) – often exhibit less well developed language and executive skills, facing greater risks to academic success than do their typically-developing or more privileged classmates (Abel et al., 2015). The disadvantages attributed to a lagging NC and EF development increase as children progress through school (Petersen, 2011). Early interventions that support the development of narrative skills in young children have been shown effective to promote NC and academic success at the preschool level (e.g., Spencer & Petersen, 2018). Furthermore, these interventions appear to have positive and substantial long-term effects. Evidence on EF training at preschool age also showed that cognitive training to improve these skills early could be effective (Diamond & Lee, 2011; Scionti et al., 2020)

1.1.2 Aims of the study

The goals of the present meta-analysis are the following:

1. Determine the overall strength of the relation between narrative competence (NC) and executive functions (EF) across childhood and adolescence (3-18 years)
2. Determine *if* the strength of this relationship changes across childhood and *when* it changes across development.
3. Examine potential moderators to understand if the strength of the relation changes:
   - between typically- vs atypically developing children (e.g., attention deficit hyperactivity disorder [ADHD], autism spectrum disorder [ASD], specific language impairment [SLI]).
   - by different EF domains (working memory capacity and updating, behavioural inhibition, interference control, shifting, planning and problem-solving);
   - by different narrative types (oral vs written) and levels (micro vs macrostructural levels).

1.2 METHODS

1.2.1 Operational definitions

We categorized NC based on the characteristics of narratives: written or oral. Both types of narratives included the ability to retell or tell a story in written or oral form. Moreover, we classified measures related to NC by dividing them into micro-structural and macro-structural competence. Micro-structural components have been collapsed into one dimension, including lexical (e.g., number and variety of words produced) and syntactic skills (e.g., indices of number and type of utterance and subordinate sentences produced; the mean length of utterance) in narration. Macro-structural components have been collapsed into one dimension, including the richness of content of the narrative (e.g., the amount of information reported in the narrative), the presence of the key passages in the story (e.g., the ability to structure a coherent story), and the cohesion of the story (e.g., anaphoric use of the article and correct referencing across the narration).

Executive domains have been differentiated according to which primary executive process the tasks assessed, based on the EFs assessment literature (Nigg, 2017; Diamond, 2013; Henry & Bettenay, 2010; McCormack & Atance, 2011). For instance, tasks requiring keeping in mind and actively manipulating auditory or visual information (e.g., backward digit; word or spatial span tasks) were coded as working memory capacity measures. These have been distinguished from tasks that mainly required updating of working memory (e.g., n-back), defined as "the ability to monitor and code incoming information, and to update the content of memory by replacing old items with newer, more relevant, information" (Van der Sluis et al., 2007, p. 428). Forward span-like tests were considered to measure short-term memory since they did not require working memory processes (Alloway et al., 2006); therefore, we did not include them in the meta-analysis.

We considered those tests that required children to suppress a dominant but inappropriate response or to prevent impulsive motor response (e.g., knock and tap task; go/no-go; Head Toes Knees Shoulders) as a measure of "behavioural inhibition" (Nigg, 2017). Instead, tasks requiring the ability to prevent interference due to resource or stimulus competition and filter out irrelevant information within the stimuli that contain both relevant and distracting information (e.g., Stroop-like, local to global and Flanker paradigm) are categorized as "interference control" task (Nigg, 2017).

We categorized tests requiring shifting among different response sets and flexibly adjusting the response according to new rules (e.g., verbal fluency, five-point test, Trail Making Test and Wisconsin Change Card Sort) as measures of "shifting".

We classified tests that required order events mentally in advance and plan the actions (McCormack & Hanley, 2011), such as Tower-like tasks or non-narrative sequences, as measures of planning abilities.

If a study collapsed different tasks in a single general dimension, we included it as a general measure of EF for the purpose of the main analysis (e.g., Ketelaars et al., 2011; Drijbooms et al., 2017). However, in such cases, we could not be able to discern between the various EF domains implied. For this reason, we could not consider that outcomes for the analysis of moderation by EF domains.

1.2.2. Search Strategy

Pre-registration of the present research is available on Prospero (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021282450).

In accordance with the PRISMA statement (Page et al., 2020) we used a systematic search strategy to find the pertinent studies. Using different combinations of the terms "executive functions", "narrative" and their synonymous (see Appendix A), we searched on Pubmed, Psychinfo, Linguistics and Language Behaviors, Proquest Dissertations&theses Global, e-thesis online service (Ethos), DART-Europe E-theses Portal to identify all potentially journal articles and unpublished studies, as doctoral dissertations, that reported data on the relation between EF and NC in children and adolescents. Despite our extensive research of the grey literature, we found only a small amount of unpublished studies (overall, 5 studies and 46 different effect size). Preliminary analyses ruled out the presence of publication bias: the size of the relationship was similar in the published and unpublished studies. Therefore, we also included these studies in the main analysis.

After excluding duplicates, 885 records remained. The first author screened all of them based on title and abstract and according to inclusion and exclusion criteria. As a secondary search, the references of the selected studies (n = 15), in addition to relevant systematic reviews, were checked to find other eligible studies. Full text of the identified papers was reviewed by the first author and EB. Disagreements were solved through discussion. The agreement rate between the two raters was high (81%). Finally, as shown in the flow chart, we identified 25 articles (30 studies) with 287 effects that were eligible for the present meta-analytic review. Details concerning the literature search method and criteria for inclusion and exclusion of studies are shown in Figure B1 (see Appendix B)

1.2.3. Inclusion Criteria

The included studies had to meet the following criteria:

- at least one performance-based test related to EF and one related to micro or macrostructural level of NC
- correlational study with a cross-sectional or longitudinal design
- monolingual participants aged between 3 and 18 years old
- paper is written in English, Italian or Spanish.

1.2.4. Exclusion Criteria

We excluded all the studies where participants are bilingual and older than 18.

All outcomes were based on correlations between one or more EF and NC tasks. Where available, we included correlation with accuracy and reaction times on EF tasks. We did not accept measures of EF aspects collected through teacher and parent reports (e.g., BRIEF) because these measures seem to capture different aspects from tasks (Toplak et al., 2013). At the same way, we did not accept measures of narrative comprehension measured through questions. The only kind of NC tasks included required the child to produce a personal story or to retell a story they heard, in oral or written form. We included the studies only if they reported at least one score at neurocognitive EFs measure and at least one micro or macrostructural competence score at NC task.

1.2.5. Coding

During the coding phase, the first author coded each record according to a predefined coding schema, collecting information about bibliographic information (i.e., title, author(s) and year of publication), sample characteristics (i.e., sample size, mean age and standard deviation of each group, clinical risk status of the sample), characteristics of the narrative tasks (i.e., written versus oral form; microstructural versus macrostructural level) and the kind of EF measure (i.e., working memory capacity, updating of working memory, behavioural inhibition, interference control, shifting and planning) and the correlation indices between the NC and EF tasks.

All the correlation indices between the tasks were included if there were two or more eligible NC and EF measures. We apply the same procedure when multiple groups were suitable for the aims of the meta-analysis, like typically and atypically developing children in the same study (i.e., Peristeri et al., 2020; Ketelaars et al., 2011; Park, 2014) or preschoolers and school-aged children (i.e., Balaban et al., 2015; McNiven et al., 2010).

1.2.6. Meta-Analytic Procedures

We used R version 4.1.2 (R Core Team, 2021), RStudio version 1.4.1103 (RStudio Team, 2020), and the Metafor package (Viechtbauer, 2015; see Assink & Wibbelink, 2016) to conduct the analyses. R code and data are openly available at the following webpage: https://osf.io/b34se/?view_only=e1b7bf6b53a34c8587bd58901db8f84d

Pearson product-moment correlation has been used as the effect size to examine the relationship between NC and EF. The magnitude of the correlation has been interpreted using Cohen's (1988) conventions:

- $r \approx 0.10$ [$z \approx 0.10$]: small effect.
- $r \approx 0.30$ [$z \approx 0.31$]: moderate effect.
- $r \approx 0.50$ [$z \approx 0.54$]: large effect.

Since correlations are restricted in their range (i.e., they can take values between -1 and 1), it can introduce bias when we estimate the standard error for studies with small sample size. Thus, the correlation coefficients collected from the selected studies have been transformed into Fisher's *z*. This transformation entails using the natural logarithm function to remove the range restriction and ensure that the sampling distribution is approximately normal. Fisher's *z* and the standard error of Fisher's *z* were calculated directly in R using the cor and log function.

A positive *z* value reflected a positive association between NC and EF, while a negative effect indicated that when the EF competence increased, NC decreased. We computed Z Fisher transformation using Olkin and Finn's (2008) formula. The summary statistics required for each outcome were the number of participants and the correlation coefficients between NC and EF measures. For one study based on regression analysis (i.e., Khan, 2013), the correlation coefficient has been converted from the β coefficient, according to Peterson and Brown's (2005) procedure.

As discussed, many studies in the dataset reported several correlated relevant outcomes, and some studies comprised multiple groups of individuals (e.g., with typical and atypical development). This caused dependencies in the data. So far, several solutions have been introduced to avoid dependency (Assink & Wibbelink, 2016; Borenstein et al., 2009): analyzing the outcomes as if they were independent (i.e., ignoring the dependency), averaging the dependent outcomes into a single effect size, selecting only one outcome for each study, and multilevel meta-analysis. Ignoring the dependency might bias the results; averaging or eliminating effect sizes, on the other hand, would decrease the power of the analysis and limit the research questions that we could ask, as we would not be able to compare moderation effects by EF and NC domains. We, therefore, conducted a three-level meta-analytic analysis, following Assink & Wibbelink (2016). The meta-analytic model considered three different sources of variance: the participants at level 1, the outcomes at level 2, and the studies at level 3.

We used the rma.mv function of the Metafor package and set the tdist parameter as TRUE. Therefore, we based the test statistics and confidence intervals on the t distribution, applied the Knapp and Hartung (2003) adjustment, and used the Restricted Maximum Likelihood estimation method (REML) for estimating the parameters. Tau^2, the Q-test for heterogeneity (Cochran, 1954) and the I^2 statistic were reported.

Studentized residuals and Cook's distances were used to examine whether studies may be outliers or influential in the model context. Studies with a studentized residual larger than the 100 x [1 - 0.05/ (2 x k)] th percentile of a standard normal distribution were considered potential outliers (i.e., using a Bonferroni correction with two-sided alpha = 0.05 for k studies included in the meta-analysis). Studies with a Cook's distance larger than the median plus six times the interquartile range of the Cook's distances were considered influential.

1.3 RESULTS

1.3.1. Selected studies

Thirty studies were eligible for inclusion, for a total of 287 different outcomes, with 3250 participants with typical development (M_{age} = 8.18) and 346 participants (M_{age} = 8.02) with atypical development (i.e., diagnosis of learning disorder, autism spectrum disorder, language impairment, deafness).

1.3.2. Inspection for publication bias

We explored the funnel plot to investigate potential publication bias and checked for differences in effect sizes between published and unpublished studies. The Egger's regression test, using the standard error of the observed outcomes as moderator, is used to check for funnel plot asymmetry. The funnel plot is presented in Figure S1 (see Supplementary Material 1, left panel, at the following webpage: https://osf.io/b34se/?view_only=e1b7bf6b53a34c8587bd58901db8f84d). No evidence of publication bias emerged, (Egger's t = 1.116, p = .266). A visual inspection shows that only a few studies fall outside the pseudo-confidence interval's triangular region. Next, we compared the effect sizes of published and unpublished studies, as higher effects for published studies might be an important indication of publication bias. We could locate only five unpublished studies, with a total of 46 different outcomes. No evidence of publication bias emerged, $F(1, 285)$ = 0.96, p = .325. On the contrary, the size of the effect was slightly bigger for the five unpublished studies than for the published studies: for the unpublished studies the effect was z = 0.283, SE = 0.041, 95% CI = (0.199, 0.367) and for the published studies the effect was z = 0.233, SE = 0.020, 95% CI = (0.193, 0.273). Since this difference was negligible, we decided to include the five unpublished studies in the main analysis.

Subsequent analysis indicated that the size of the effect was related neither to the year of publication of the study, $F(1, 285)$ = 0.187, p = .665, nor to languages spoken by sample of participants involved in the studies, $F(7, 296)$ = 0.193, p = .986. Moreover, a sample size moderator analysis was performed, which resulted in a non-significant effect (p = .109), suggesting that differences in sample size are not a source of the heterogeneity of the results.

An examination of the studentized residuals revealed that one study (Sacchetti, 2018) had a value larger than ± 3.7537 and may be a potential outlier in the context of this model. According to Cook's distances, four studies (Peristeri et al., 2020; Aran-Filippetti et al, 2015; Veraksa et al., 2020; Vanderberg & Swanson, 2006) could be overly influential.

1.3.3. Research question 1: Exploring the overall association between EF and NC

A total of k = 287 effects were included in the analysis. The observed Fisher r-to-z transformed correlation coefficients ranged from -0.0601 to 1.2111), with the total estimates being positive. The estimated average Fisher r-to-z transformed correlation coefficient based on the random-effects model was $z = 0.241$, $r = 0.236$, (95% CI: 0.2053 to 0.2776;). Therefore, the average outcome differed significantly from zero ($t = 13.134$, $p < 0.0001$), indicating a positive, small association between EF and NC over development. According to the Q-test, the true outcomes appear heterogeneous ($Q_{(286)} = 597.25$, $p < 0.0001$. The estimated variance components were $tau^2_{(level\ 3)} = 0.005$ and $tau^2_{(level\ 2)} = 0.006$. This means that $I^2_{(level\ 3)} = 22.95\%$ of the total variation can be attributed to between-study and $I^2_{(level\ 2)} = 29.95\%$ to within-study heterogeneity. We found that the three-level model provided a significantly better fit compared to a two-level model with level 3 constrained to zero ($\chi^2 = 33.39$, $p < .001$).

The 75% rule (Hunter and Schmidt, 1990) suggests that we should inspect heterogeneity if <75% of the total amount of variance can be attributed to within study sampling variance. Therefore, we proceeded to investigate potential moderators, following the research questions outlined above.

1.3.4. Research Question 2: Exploring if and when the association between EF and NC changes over development.

We investigated the impact of age on the relation between EF and NC through meta-regression to understand if and when the relationship between NC and EF changes over time (see Table 1.1). The mean age of the sample ranged between 4 and 15 years and significantly influenced the effect size so that as age increases, the overall effect size decreases, $F(1, 265) = 6.744$, $p = .009$.

The unstandardized regression coefficient and significance for the slope are reported in Table 1.1, which indicates the impact of each unitary change (i.e., one year) in the moderator on the effect size of the relation between EF and NC.

Table 1.1 Age effect on the relation between EF and NC

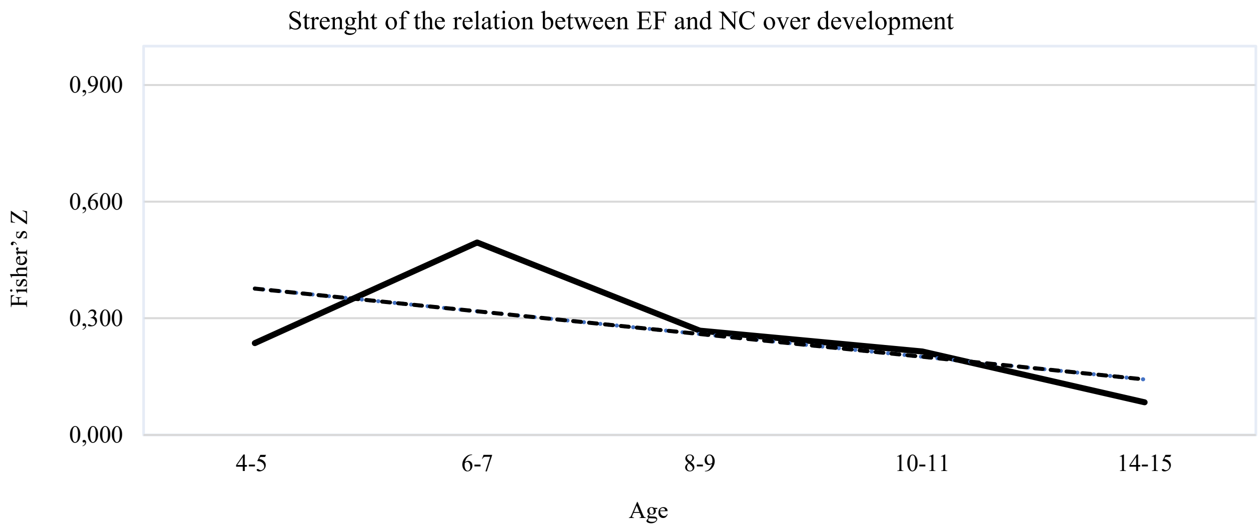
Effect	No. outcomes	No. studies	No. participants	Estimated z	SE	95% CI		p-value
Children's age (years)	*267*	*29*	*3410*	*-0.014*	*0.005*	*-0.025*	*-0.003*	*.009*
Developmental time windows								
Before 7 years	*85*	*13*	*795*	*0.274*	*0.029*	*0.216*	*0.333*	*<.001*
After 7 years	*182*	*16*	*2615*	*0.212*	*0.021*	*0.170*	*0.254*	*< .001*

Note. Italic text indicates the levels of the categorical variables

Looking at the effect size trend over development (see Figure 2), the relationship's turning point appears to be around 7-8 years old. Thus, we performed moderation analysis by dividing the sample into two-time windows (i.e., mean age < 7 years; mean age > 8 years). Results show that this variable significantly impact on the effect size, so that after 7 years old the magnitude of the relation between EF and NC decreases from $z = .274$ to $z = .212$, $F(1, 265) = 3.908$, $p = .049$.

According to these results, we decided to conduct separate meta-analyses to investigate the influence of potential moderators in these two developmental windows (3-7 years; 8-15 years, see Table 1.2).

Figure 2. The relation between EF and NC over development.



Note. The solid line represents the trend of Fisher's Z coefficient over time. Point of the solid line are averaged effect size of the relation between EF and NC in the five time-intervals considered. The dotted line is the trend line of the relation between NC and EF over time. The angular coefficient of the dotted line is negative, indicating that the association between NC and EF decreases over time.

Tables C1.1 and C1.2 (see Appendix C) summarized the characteristics of the studies included respectively in the first and second meta-analysis. In particular, in Table C1.1 we reported the correlations between EF and NC of participants aged 3-7 years old; in Table C1.2, we reported the correlations between EF and NC of participants aged 8-15.

Table 1.2 Moderators of the association between NC and EF before and after 7 years of age.

Effect	**No. outcomes**	**No. studies**	**No. participants**	**Estimated z**	**SE**	**95% CI**		**p-value**
Population (4-7)								
Typically developing	77	9	652	0.248	0.230	0.202	0.294	< .001
Atypically developing	8	4	143	0.436	0.086	0.264	0.607	< .001
Population (8-15)								
Typically developing	106	11	2412	0.221	0.026	0.169	0.273	< .001
Atypically developing	76	5	203	0.199	0.040	0.119	0.279	<.001
EF Domain (4-7)								
Working memory capacity	36	7	459	0.259	0.035	0.188	0.330	< .001
Working memory updating	2	1	37	0.344	0.144	0.057	0.632	0.019
Interference control	8	2	63	0.309	0.074	0.160	0.458	< .001
Behavioral Inhibition	18	5	185	0.153	0.049	0.055	0.251	0.002
Shifting	12	4	211	0.292	0.054	0.183	0.400	< .001
Planning	7	2	122	0.372	0.075	0.222	0.522	< .001
EF Domain (8-15)								
Working memory capacity	39	11	2248	0.232	0.032	0.168	0.297	< .001
Working memory updating	12	1	40	0.135	0.087	-0.036	0.307	0.120
Interference control	52	4	1295	0.228	0.044	0.139	0.317	<.001
Behavioral Inhibition	14	4	269	0.292	0.048	0.197	0.387	< .001
Shifting	30	6	339	0.205	0.043	0.119	0.291	< .001
Planning	17	4	177	0.204	0.052	0.101	0.307	< .001
Narrative form (8-15)								
Oral	86	6	266	0.252	0.044	0.165	0.340	< .001
Written	96	10	2349	0.200	0.026	0.148	0.252	< .001
Narrative Competence (4-7)								
Micro-structural	45	8	578	0.209	0.023	0.163	0.0255	< .001
Macro-structural	32	8	527	0.329	0.025	0.278	0.380	< .001
Narrative Competence (8-15)								
Micro-structural	105	12	2476	0.213	0.024	0.164	0.261	<.001
Macro-structural	77	14	1208	0.216	0.026	0.164	0.268	< .001

Note. Italic text indicates the levels of the categorical variables

1.3.5. Research Question 3: Potential moderators of the relation between EF and NC before and after 7 years old.

Table 1.2 shows a summary of the impact of the following moderators on the relation between EF and NC in the two developmental windows considered.

- *Typically vs atypically developing population.* We categorized the sample in typically developing and atypically developing participants based on the presence of a diagnosis (i.e., deafness, SLI, learning disorders, ADHD and ASD). The studies involving children younger than 7 years old (n = 795) indicated that the effect sizes differ between the groups, $F(1, 83) = 4.400$, $p = .039$. The association between EF and NC was almost twice in atypically developing children ($z = .436$) than in typical peers ($z = .249$), unless both effects are significant.
  On the contrary, in the subsample of studies involving children older than 8 years old (n = 2615), the analysis indicated that the effect size was the same for typically ($z = .211$) and atypically ($z = .196$) developing populations, $F(1, 180) = 0.132$, $p = .715$.
  The number of studies involving atypically developing populations of children, however, was relatively small in both subsamples: we found only four studies, with a total of eight different effects and 143 atypically developing children younger than 7 years old; only five studies, with a total of seventy-six different effects and 203 atypically developing children older than 8 years old.
- *EF domains.* Looking at EF, we investigated if, before and after 7 years old, effect size differs on the type of EF domains taken into consideration (i.e., interference control, behavioural inhibition, working memory capacity, updating of working memory, shifting, planning). Results showed that before 7 years, the effect size did not statistically differ on the type of EF domains, $F(5, 77) = 2.069$, $p = .109$. At this stage, EF domains are equally significantly associated with NC. However, in the subsample of studies involving participants older than 8 years old, variance in the effect size was significantly explained by EF domains, $F(5, 162) = 3.399$, $p = .006$. In line with the age effect previously discovered, the relation between NC and the majority of the EF processes decreases, with the exception

of behavioural inhibition. The effect size of the association between behavioural inhibition and more general NC is larger than those observed in younger children.

Also, the association of shifting and planning with NC remain significant in older children, even if it is lower. As regards working memory dimension, the measures addressing its capacity remains similarly associated with NC, whereas those addressing updating processes decreased significantly in older children.

- *Narrative Competence.* Looking at characteristics of NC, we next compared studies on children before and after 7 years old, analyzing if micro versus macrostructural levels of narratives moderates the effect size of the relation between EF and NC. Results referring studies on participants younger than 7 years old indicated that the effect size was higher for macrostructural (z = .329) than microstructural (z = .208) competences, $F(1,75) = 12.23$, $p < .001$, unless both the effect were significant ($p < .001$). After 7 years old, however, no significant difference emerged for the comparison between micro and macrostructural aspects, $F(1, 180) = 0.074$, $p = .784$.

  Next, we questioned if, in the subsample of studies with children older than 8 years old, the relation between EF and NC differs on the type of narrative tasks (i.e., written versus oral form). Results indicated that the type of narrative task did not explain variance in the effect size, $F(1, 180) = 1.36$, $p = .243$.

1.4. DISCUSSION

EF and NC are two widely investigated dimensions of human cognitive development, but our understanding of their relationship is limited. For instance, we do not know if these dimensions are related over time or if this relationship changes across development. We do not know much about this relationship, especially in atypically developing children and adolescents, although we know that these areas are usually impaired in such populations. In general, few studies have investigated this relationship. Mostly, these studies involved small samples used cross-sectional design and produced mixed results. The aim of this meta-analysis is not to answer these questions according to the studies published so far. It intends to raise some points that can guide future research on these topics, such as which age range needs further consideration by scientists. We claim, as of right now, that more studies in general – and specifically more longitudinal studies – are needed to shed light on the relationship between these dimensions over time in typical and atypically developing individuals.

So, the first purpose of the present meta-analysis was to establish if, overall, EF and NC are transversally – not longitudinally – associated.

As expected, the studies collected showed great heterogeneity within and between themselves. However, the multilevel meta-analysis showed that – overall – a positive but small relationship between EF and NC exists (r = .236). It means that the researches selected provide evidence that – in general – individuals who performed well at EF tasks are also good narrators and vice versa. The result obtained reflects the high variability there are between the studies included. Nine studies reported an average effect size below 0.20, but most reported moderate (0.30-0.49) effect sizes. Inspection for publication bias reveals that the results obtained are similar in the published and unpublished literature, so the probability of overestimating the magnitude of this relationship is remote.

The second purpose was to examine if the relation between EF and NC changes over time and at which point it starts to change significantly. In order to fulfill this aim, we considered the mean age of participants in the studies. Results showed that the relation between EF and NC changes over time and decreases over development. The plot of the association between NC and EF across development (Figure 2) showed that the transversal association increases during the preschool years, when both NC and EF dramatically develop, peaking in the early elementary school years and then starting to decrease significantly after 7 years old.

Different factors might explain the turning point we can observe at this age.

We speculate that a key role might be played by literacy acquisition to which the early years of elementary schools are dedicated. During these years, children develop effective decoding skills (Seymour et al., 2003). Specifically, children speaking languages with shallow syllabic complexity and orthographic depth (e.g., Italian, Spanish, German, Greek) become accurate and fluent in foundation reading before the end of the first school year. In contrast, children speaking languages characterized by deep orthographies (English, French, Danish and Portuguese) – the majority of children involved in the studies selected for this work belong to this group – become fluent at nearly 8 years old (Seymour et al., 2003).

Research on the development of reading and writing suggests that the development of these skills is deeply interrelated and that, especially during elementary school years, reading contributes significantly to the quality of narrative composition (Abbott & Berninger, 1993; Olinghouse, 2008), especially from a macrostructural point of view (i.e., better structured and cohesive narrations). It is possible that after literacy acquisition, the role of EF in narrative production is downgraded by other factors that contribute to NC development, such as reading skills. Of course, this is a speculative interpretative hypothesis. To our knowledge, there are no studies that had taken into consideration the role of both EF and reading skills on the development of NC.

Changes in exposure to narratives could also explain the decrease in the association between EF and NC. The amount of this exposure may play a role in the development of NC and downgrade the association between EF and NC. It is true that narratives are cross-culturally used in children rearing systems, and children are exposed to them from very early in life to a greater or lesser extent. However, during preschool and the first years of elementary school, children are exposed to narratives and narration is widely used as an educational strategy in school. Narratives create a pleasant and creative learning environment and a more general constructive and enjoyable atmosphere for the children (Nanson, 2021). Moreover, the use of narrative in education attracts the interest of the children and assists in the better understanding of the information obtained through this. Often, story grammar becomes part of the school curricula and children are taught to become good narrators, so it is possible that when the development of good NC becomes formal learning, NC may progressively be less associated with or dependent on EF.

It seems that the two dimensions are more associated early in childhood, the period in which EF and NC – taken singly – dramatically know rapid and qualitative changes (Anderson & Reidy, 2012; Berman & Slobin, 2013). We have discussed the possibility that EF may become less relevant for supporting NC over development, but it is also possible that NC supports EF development across time, becoming less essential nearly 8 years old. There is evidence that language skills support EF development, especially across preschool age, and narrative language could be considered a "naturalistic" way to investigate children's language in connected speech. (Schraeder et al., 1999). Therefore, it is possible that the practice of constructing causally coherent true narratives could help children in initiating and regulating behaviour – as demonstrated in language research (Vallotton & Ayoub, 2011; Zelazo & Jacques, 1996) – and that narrative language may have a mediating role in EF performance, as there is evidence that language skills have this role in both deaf and hearing children (Botting et al., 2017).

However, it is still possible that increasing cognitive demands associated with the transition to elementary school and the development of other competencies play a more significant role than NC in the development and reorganization of EF. The role of NC – and language – in EF development can be progressively nuanced by the other increasing competencies in this period, which could be responsible for the decrease observed in their association. It should point out that the argument that the magnitude of the relation between NC and EF seems to decrease over time applies only to the transversal relationship between them. One competence may relate longitudinally with the other and vice versa. For instance, NC and EF may be weakly related at 9 years old, but EF at 5 years old is significantly associated with NC observed at 9 years old. However, there is insufficient data in the literature to answer this question with a meta-analysis.

The third purpose of this work was to try to understand some moderators responsible for the heterogeneity observed between studies in the magnitude of the association between EF and NC.

Since the magnitude of the transversal relationship between EF and NC changes over time, we analyzed the role of these moderators in two different time windows: before and after 7 years old. We found that, before 7 years, the association between EF and NC is stronger in children with atypical development, such as ASD, ADHD or SLI. However, later in development, the strength of the association fades. After 7 years, results suggest the strength of the association appears similar in typical and atypical development unless only in the latter is statistically significant.

As mentioned above, NC and EF are skills that predict important life outcomes and are trainable (Abel et al., 2015; Diamond & Lee, 2011). They are frequently impaired in children with ASD, ADHD or SLI (Gooch et al., 2016; Craig et al., 2016) and our results may suggest that in such populations the impairment on EF could somewhat impair NC, or vice versa, between 3 and 7 years of age. In literature, several training programs aimed at improving EF or NC have been described (e.g., Spencer et al., 2018; Thorell et al., 2009), showing promising results in preschoolers (Petersen, 2011; Scionti et al., 2020). There is also evidence that the training effects are higher in children with developmental risks or psychopathological traits (Melby-Lervåg & Hulme, 2013; Scionti et al., 2020; Wass et al., 2012).

Establishing if two dimensions are associated across the development is the first necessary step to hypothesize that training one could foster the development of the other. Currently, researches aimed to study the effectiveness of EF or NC training did not take into consideration possible far transfer effects on them. In the same way, there are no studies that implement integrated interventions targeting both NC and EF or studies that verify their effectiveness.

The results of this meta-analysis could be read as a first step in the direction to project research on integrated interventions and plans to verify the effectiveness of single EF intervention on NC and vice versa. Based on our results, we could do some speculative hypotheses related to the fact that – if a far transfer between NC and EF is possible - the chance to observe it on the non-directly trained skills would reduce after 7 years. Following this reasoning, according to our results, only training programs aimed at improving single specific competencies might be effective in older children showing impairments both in NC and EF. This is consistent with research that unanimously agrees that intervention is likely more effective and pervasive when provided earlier in life rather than later (Centers for Disease Control and Prevention, 2022).

Moderation analyses also explained part of the heterogeneity in the effect size between and within different studies depending on different EF domains and NC levels assessed.

We found that before 7 years the association between EF and NC is stronger if we considered the macrostructural level of NC, which includes several important story characteristics such as the quantity of information, story structure and cohesion. This is not unsurprising, as in the transition from preschool- to school-age this competence shows a remarkable increase (e.g., Roch et al., 2016, Zanchi & Zampini, 2021). For instance, analyzing the stories produced by children aged 4 to 8, Schneider et al., (2006) showed a significant increase in the quantity of relevant information included in the narrations as children's age increased. In addition, as children grow and develop their NC, they gradually move from non-goal-directed sequences toward complete episodes. From preschool to elementary school, children go from producing stories that include few causal connections between events to being able to conceive an overall plot with most of the story grammar elements and following a logical progression of events in their stories, which make them appear more cohesive and well structured. It is possible that EF play a significant role in this progress and that this progress may support EF development.

It seems that later in development, the strength of the association between EF and macrostructural NC fades. After 7 years, results suggest the strength of the association appears similar at the microstructural and macrostructural levels. However, children's ability to tell stories continues to develop during primary and secondary school. Older children indeed include more events than do younger ones (Stein et al., 1979); they correctly use a broader range of conjunctions (Shapiro & Hudson, 1991) and more advanced anaphoric strategies (e.g., pronouns were used to maintain a reference to characters, whereas nominals were used to switch a reference) that make the stories appear more cohesive. Also, EF shows an increase in late childhood and adolescence, but its development may be less involved in NC and vice versa. As argued before, a more significant role in NC increase at this time may be played by reading skills consolidation or other competencies.

As well, heterogeneity in the effect size seems cannot be explained by the narrative form (oral vs written) used in the articles collected. This is consistent with results found by Bigozzi and Vettori (2016) who showed that in the transition from oral to written code, typically developing children who master writing preserve their oral narrative skills. There is evidence that difficulties in written over oral narrative form may be observed in atypically developing children who struggle with handwriting. Unfortunately, our sample size was not adequate to investigate the interaction of the two moderators (i.e., population and narrative form) in the subgroup analysis. In the subgroup of studies involved in the second meta-analysis (children older than 8 years) atypically developing children represent only 8% of the sample.

As regards EF domains, we found that the strength of the association between EF and NC appears similar for different EF domains before 7 years old. After 7 years, results showed a general decrease of the strength of the relation, even if some differences from medium overall effect size emerge by different EF domains.

More specifically, in preschoolers and first and second graders, the contribution of EF to NC appears statistically equal across EF domains. This could be because at this age EF tend to be more related and less differentiated from each other (Monette et al., 2015; Scionti & Marzocchi, 2021; Usai et al., 2014; Wiebe et al., 2011), so any attempts to connect the various tasks to one distinct EF domain at this age may be artificial. For this reason, specific patterns between EF domains and NC could be challenging to observe in this time window. Also, a technical consideration may explain the absence of evidence. Studies included in the first meta-analysis showed substantial between-study heterogeneity within the EF domains, which decreased the pooled effect's precision (i.e., increased the standard error). Yet, when the EF domains effect estimates are imprecise, their confidence intervals will have a large overlap, as in some of our cases (e.g., working memory updating CI index: 0.057, 0.632). Consequentially, this might make it harder to find a significant difference between subgroups – even if this difference could exist.

Anyway, specific patterns in the relation with NC may emerge after 7 years old, when EF domains are more differentiated and distinguishable (Lehto et al., 2003).

In general, the contribution of all EF domains to NC seems to decrease after 7 years, with the notable exception of behavioural inhibition. This domain refers to the ability to suppress a dominant but inappropriate response or prevent impulsive motor response, according to Nigg's definition (2017). Togheter with interference control, behavioral inhibition may be critically involved over development to monitor the production of extraneous comments and derailments while telling a story or inhibit semantic competitors while producing words. As well, NC may be involved in inhibition tasks. Narrative language may indeed be used to exert control over attention and inhibit inadequate response and interferent representation.

Like inhibition, also working memory capacity, shifting and planning appear to be involved in NC at this age. Working memory capacity could be required to keep in mind ideas before translating them into linguistic representations, as well as to recall episodic contents for an accurate organization of temporal sequences in the story. Shifting could be required in the generation of complete episodes and in the ability to monitor the communicative flow. Instead, planning may play a coordinating role in the story organization, e.g., putting in the correct sequence all the story elements (Khan, 2013). These results are in line with studies reporting working memory, shifting and planning are correlated with text generation in older children (Drijbooms et al., 2016; Puranik, 2006) and adolescents (Swanson & Berninger, 1996b). Other domains, instead, seem to be significantly less associated in this period with NC than in the previous time window, such as updating of working memory. This is consistent with previous findings in Swedish (Tonér & Nilsson Gerhold, 2021) and Canadian preschoolers (McNiven, 2010).

1.4.1. Study limitations

Finally, we would like to discuss some limits of the present work. As claimed above, the current meta-analysis cannot respond definitively to some questions about the relationship between NC and EF because of its limits. The first limit was related to the fact that few studies investigate this relationship with a longitudinal design. Therefore, even if our results clearly show that a relationship between NC and EF is definitively positive, we know it is just transversal. We cannot say something about how and if these dimensions are related longitudinally across time if there is one point at which one predicts the other and vice versa because there is not enough research addressing this issue. Future research should investigate if these domains are predictive of each other and establish the direction(s) of their development. A second limitation concerns the time variable used in this meta-analysis to answer the question of whether the relationship changes over time: the mean age of participants. Some studies included in the present meta-analysis involved participants of a large range of ages (e.g., 7-12; 7-14, see Table C1.2), so it was hard to classify the studies by age stage (e.g., preschoolers; school-aged; adolescents). We preferred not to exclude these studies from the analysis and chose to consider the mean age of the participants collecting – where available – the effect size adjusted for the effect of age. The time effect is one of the most interesting issues for a developmental psychologist. Even if the praxis to analyze the impact of time/age over a phenomenon in meta-analytic developmental psychology research is consolidated, it should be kept in mind that using aggregate information – such as the mean age of participants – may produce ecological-biased results (Piantadosi et al., 1988; Thompson & Higgins, 2002). Therefore, any conclusion around the relationship between EF and NC changes should be taken cautiously and considered just orientational. Aggregating data suggest that a turning point in this relationship is around 7-8 years old, but studies covering this age range also include 6- and 9 years old participants. Furthermore, studies covering this age range in the sample of articles selected from the meta-analysis are few ($k = 4$). Meta-analytic research led to summarizing results from different studies, which potentially may offer a comprehensive picture of a phenomenon. In this case, what we can draw is that the relationship between EF and NC seems to decrease over time, even if we cannot be sure of the exact time point in which it starts to drop, but it seems that it places around the first three grade of elementary school. Future studies should keep more in exam this period than preschool.

Another limitation concerns the intrinsic multidimensionality and complexity of EF construct examined and the large variety of instruments used to capture the construct across development. We based the instruments' classification on the scientific literature (Nigg, 2017; Diamond, 2013; Henry & Bettenay, 2010; McCormack & Atance, 2011) in order to clarify which task assesses specific component, but we are aware of the "task impurity problem", a phenomenon in which one task assesses various EFs components beyond the one it aims to evaluate, which is frequently in young children. So far, we invite the reader to take cautiously into consideration findings about the specific pattern of relationships between various EF domains and NC since this may depend on the classification we used.

Finally, the last limitation we mention is that NC and EF are two dimensions that in real life are related to many other dimensions of human development that could mediate or explain their relationship. One of these among all is theory of mind, which is associated with both dimensions (Ketelaars et al., 2012; Perner and Lang, 1999). In certain circumstances, speculatively, these variables might be responsible for the presence or the lack of association between EF and NC across the studies. Studies included in this meta-analysis consider the account of potentially confounding variables (e.g., age) on the correlation between EF and NC, to various degrees and differently. They used to control their effects by reporting partialized correlation coefficients of the relation between EF and NC. Unless this operation is fundamental to provide a reliable measure of the association between EF and NC, it increases the between-study heterogeneity. For this reason, another limitation in interpreting our results is that we cannot be sure that this relationship is direct. Further investigations are necessary for this scope.

1.4.2. Conclusions

In conclusion, despite these limitations, this work suggests that, over time, the domains of EF and NC are associated and may depend on each other. This seems to be especially true in young, atypically developing children and for macrostructural elements of NC. However, in general, the relation between EF and NC that is stronger in early childhood is bound to decrease over development. Since these competencies are usually impaired in children with atypical development, but they can be effectively trainable, we stressed that good practice might be to introduce small groups intervention to support one or both competencies at the end of preschool and in the first two grades, i.e., at the time EF and NC appear more related.

Furthermore, the results provided in this meta-analysis and their limitations suggests some orientational consideration for future research:

- Previous research has focused more on these domains taken singly than on their relationship. However, to understand human development and support it with effective intervention, we should also focus on connecting its parts. NC and EF are promising domains because they predict many life outcomes and seem trainable. We should know much about their relationship, especially in atypically developing people and in longitudinal ways. This is to understand when and how it is better to intervene to be effective.
- Previous research on EF and NC focused mainly on two age bands (i.e., 3-6 and 9-12) and considered large age ranges. This makes it hard to understand the development of the relationship between EF and NC across time. Even if results provided by single studies are frequently controlled by age differences, it would be insightful observe the correlation in more homogeneous groups of age. Furthermore, since the strength of the relationship seems to decrease over time, and a turning point in this sense may be represented by the first two grades of elementary school, studies focused on this particular time window – that has been more neglected – should be encouraged to understand better what happens at this specific stage and if we can use it to support child development.

CHAPTER 2

Can Narrative Performance be a valid ecological measure of Executive Function in Preschoolers?

2.1 INTRODUCTION

Several authors see narrative evaluation as a "naturalistic" approach in the study of language development because narratives represent a real and contextualized request for children (Schraeder et al., 1999). Furthermore, telling a story is a multi-componential competence that requires the integration of multiple abilities (Boudreau & Chapman, 2000; Karmiloff & Karmiloff-Smith, 2002) as:

1. An adequate linguistic competence, i.e., the correct use of words to communicate the meanings intended and the appropriate syntactic structures, prepositions and adverbs needed to make a story logical and cohesive (Orsolini, 1990);
2. A good level of socio-cognitive expertise, i.e., the knowledge about the world, people, and the potential motivations that can lead the characters to act while simultaneously maintaining psychological distance from the present situation and considering others' points of view (Fernández, 2013; Gamanossi & Pinto, 2014).
3. Adequate EF supporting the organization of the narrative discourse around a typical structure (i.e., the story's grammar), which includes an introduction, provision of the setting and description of the characters in the story, a problematic situation that shapes the protagonist's goal, attempts to solve the problem, and a conclusion (e.g., Pinto, 2003; Stein, 1988; Stein, Glenn & Freedle, 1979).

Narrative difficulties may reflect linguistic, cognitive or pragmatic/social difficulties. However, psycholinguistic and clinically-based research provides evidence that oral narratives capture more subtle language difficulties, including pragmatics, and are a good predictor of long-term language and academic skills (Fazio et al., 1996; O'Neal et al., 2004). Narratives are a promising tool for identifying pragmatic problems in children with language impairment and a valuable tool to assess communication skills and determine the severity of deficit, as well as potential treatment candidacy options in adult patients with neurogenic discourse impairments (Cannizzaro & Coelho 2002; Ylvisaker et al. 2008).

Due to the theoretical involvement of EF in storytelling, some authors claimed that narrative evaluation and EF tasks might potentially represent a larger related construct of goal-directed thinking/behaviour referred to as managerial knowledge (Cannizzaro & Coelho, 2013). According to these authors, story grammar, and especially the story's episodic structure (goal – attempt – outcome), represent a type of goal-directed thinking and action similar to other knowledge stored and processed by adults in the prefrontal cortex, sometimes referred to as structured event complex (SEC) knowledge (Grafman 1995; Mar, 2004; Ylvisaker et al. 2008), sometimes referred with the general meaning of EF. Indeed, EF is an umbrella term encompassing a set of higher-order cognitive abilities implied in goal-directed behaviour. These functions enable us to alter overlearned behavioural patterns when they become unsatisfactory, understand complex or abstract concepts, solve new problems, and manage relationships (Van der Linden et al., 2000). Despite its importance, EF abilities have been traditionally quite difficult to define and measure. No single behaviour can be directly tied to EF, and there are no "pure measures" tapping EF skills. One challenge of psychological science is understanding behaviour in the context of real-life experience, especially for clinical purposes. This notion is even more relevant when assessing behaviour related to EF due to the elusive nature of the construct, its resistance to assessment, and the possible impact exerted on everyday function (Delis et al. 2001; Royall et al. 2002). The problem of poor ecological validity is relevant in EF assessment. Performance-based tests seem unable to capture children's EF processes relevant to real-world outcomes. EF tasks are well structured and generally assessed under optimal conditions, and individuals with problem behaviours may sometimes perform adequately on these tasks (Gioia et al., 2008). A common way to address the ecological validity of EF is by relating them with informant-based rating scales.

However, studies found a low systematic association between these measures suggesting they do not measure the same construct. In addition, the rater who fulfils the rating scale chooses which instances from children's everyday lives to map onto the questions probed. This may be fraught with issues related to informants, such as context effects and differences in how different informants judge child behaviour (Barkley, 2006).

Twenty years ago, Grafman and Litvan (1999) proposed that experimental tasks may be the most promising tools for exploring complex behaviours (e.g., EF) due to their close association with real-life activities. In this light, the assessment of narrative competence (NC) might represent an ecologically valid way to explore EF working in children's storytelling. As mentioned above, Cannizzaro & Coelho (2013) found that in 46 neuro-typical adults, narrative discourse structure is closely related to EF ability. The authors used two tasks to elicit narratives, looking for the presence of organizing story grammar elements and administered several tasks tapping both linguistically and visually-based EF. Performing a factor analysis of story structure and EF variables, they found two factors they interpreted as the expression of the output-fluidity and organizational-efficiency of the performance. Therefore, they suggested that narrative structure and EF might represent aspects of goal-directed knowledge that are not bound by a traditional linguistic and non-linguistic division.

As regards childhood, empirical data supporting this hypothesis is provided by a recent meta-analytic study (see Chapter 1). The authors, summarizing the results from a pool of 30 empirical studies, found that the association between EF and NC is moderate and increases in early childhood until seven years of age. However, our knowledge about which specific NC aspects correlate with EF is still limited. Due to inherent limitations of the meta-analysis method, the research presented in Chapter 1 merely showed that (a) macrostructural indices of NC are more associated with EF than microstructural, and (b) EF abilities are equally moderately associated with general NC in early childhood.

Moreover, results found on correlations between NC and EF might be biased across studies because both dimensions are related to many other dimensions of human development (e.g., theory of mind

[TOM]). Since not all studies considered possible confounding variables within their models, we do not know if other variables could further explain the correlations between EF and NC. To understand if the assessment of narratives could be a valid mean to explore EF, we need to increase our knowledge on which precise and fine-grained processes of EF and NC are eventually related and control for the effect of other confounding variables that across development might explain part of the variability shared by EF and NC. This could be remarkably complex to do with EF due to the measurement impurity issue. Measurement impurity is a ubiquitous problem in all areas of cognitive assessment, including but not limited to the assessment of EF abilities. Task impurity derives from the involvement of various executive and non-executive processes within EF tasks, making it hard to know what the task is actually measuring. Because EF abilities necessarily manifest themselves by operating on other cognitive processes, any EF task strongly implicates other cognitive processes that are not directly relevant to the target EF. *M*easurement impurity is not a problem that can be definitively "solved". However, it is a characteristic of performance-based tasks that may be managed through the explicit measurement of cognitive processes contributing to EF task performance (Willoughby et al., 2018).

In this study, we attempted to address the measurement impurity issue by (a) deriving indices from the performance at the EF task as precisely as possible; (b) using both "visually-based" and "linguistically-based" EF tasks that differ in language knowledge and involvement required to complete and (c) administering multiple control trials to control for other cognitive and perceptual skills that potentially contribute to the performance.

Following Cannizzaro and Coelho (2013), the present study would lay the groundwork for a series of studies aimed at identifying specific indices related to EF in children's narrative evaluation. This would be a little step to understand how and if we could look at indices of narrative performance as an ecologically relevant measure of EF abilities in preschoolers.

2.1.1. The assessment of narrative skills and their association with executive function abilities

Many studies have investigated children's narrative production to identify its cognitive and linguistic bases (e.g., Fiorentino & Howe, 2004; Norbury & Bishop, 2003; Orsolini, 1995; Westerveld & Gillon, 2010). Generally, NC is evaluated using story generation, story retelling, and storytelling tasks. The task that seems to investigate narrative competence better is storytelling. Here, the child is asked to tell the story represented in a picture storybook. In this way, it is possible to exclude the influence of the examiner's narrative model on children's production. The storytelling procedure is straightforward: the child is asked to look carefully at the pictures and then to tell the story (looking again at the picture book). This method was used in many works on NC (for a review, see Strömqvist & Verhoeven, 2004). The advantage of this procedure is that using pictures reduces the children's memory effort so that they have "only" to tell the story they are looking at (D'Amico et al., 2008).

The task paradigm used to assess NC influences the children's performance. Roch and collaborators (Roch et al., 2016) compared story retelling and storytelling performance in a group of bilingual children. The results showed a better performance in story retelling than in storytelling in all indices of narrative production considered. In line with what was suggested by Lever and Sénécal (2011), the authors claimed that storytelling relies more upon constructive processes, while story retelling relies on reconstructive processes. Undoubtedly, EF abilities are theoretically implied in both constructive and reconstructive processes.

Generally, two levels have been considered in analyzing NC: microstructure and macrostructure. Microstructure refers to the linguistic competence used in narration, including, for example, the lexical variety, mean length of the utterance (MLU) and syntactic complexity (e.g., Justice et al., 2010; Mäkinen et al., 2014). Instead, the macrostructure level refers to global narrative characteristics, such as the ability to produce a well-structured coherent, and cohesive story. During development, children's narrative production shows a remarkable increase at the macrostructural level (e.g., Castilla-Earls et al., 2015), particularly between 4 and 7 years of age.

The macrostructural level is the most influenced by EF (Scionti et al., submitted). This level includes several indices: quantity of information, story structure, coherence and cohesion.
The amount of relevant information included in the stories has been frequently considered an NC index (e.g., Fiorentino & Howe, 2004). Analyzing the stories produced by preschool and school-aged children, Schneider, Hayward and Dubé (2006) found a significant increase in the quantity of relevant information included in the narrative production as children grow. There is evidence that the quantity of information reported by preschool children in retelling tasks is associated with their performance at card sorting and phonological working memory span tasks (Tonér & Nilsson Gerholm, 2021; Khan, 2013). This supports the idea that children who retold more information in their stories show a higher working memory span and shifting abilities. In addition, a study investigating the EF processes underlying written story composition skills reported a significant association between the quantity of information and the performance at a card sorting task in a sample of 186 typically developing school-aged children (Aran-Filippetti et al., 2015). Based on our knowledge, no published studies have explored the association between EF and information quantity in preschoolers using storytelling or story-generation paradigms.

Considering story structure, Orsolini (1995), analyzing the narratives produced by 70 preschool children, identified four developmental levels: (1) chronicles that are narrative production in which events are only temporally linked without mentioning the problematic situation; (2) incomplete episodes, that are characterized by the presence of a problematic situation and the lack of intermediate events (e.g., the attempts to solve the problems); (3) minimal episodes, that are narratives in which both the problematic situation and the intermediate events are included; and, finally, (4) complete episodes, characterized not only by the problem-attempts-consequence sequence but also by the presence of reactions and emotional responses of the characters. As children's age increase, they gradually move from non-goal-directed sequences towards complete episodes.
In contrast, Berman and Slobin (1994) referred to Stein and Glenn's (1979) schema to evaluate the

global structure of narratives produced by children. Three main components define a complete episodic structure: *Goal* (G) resulting from internal responses related to the problematic situation, the *Attempts* (A), and the *Outcome* (O), i.e, the direct consequences of achieving or not a goal (GAO units; Liles, 1987; Merritt & Liles, 1987). If one or two components are missing, the episode is incomplete. Analyzing the development of story structure in preschoolers, the authors found that most 5-year-old children produced incomplete episodes, omitting the attempts to solve the problem and the outcomes.

Another type of episodic analysis focuses on the hierarchical organization of narratives. It is similar to the episodic analysis based on Stein and Glenn's (1979) story grammar but differs because the goal is separated from the Internal Response. After identifying each story grammar component (i.e., GAO), the Goals are identified as either superordinate or subordinate. Superordinate goals are the ultimate goals that a protagonist wants to attain, and subordinate goals are subgoals used to obtain the ultimate goals. Therefore, this analysis examines two devices children use to organize their narratives: complete GAO Units and hierarchical goal structures. The completeness of the GAO Units and the use of hierarchical goal structures are essential for producing well-structured narratives. In typically developing children, the number of complete episodes and embedded episodes is significantly increasing with their age (Roth & Spekman, 1986; Trabasso & Nickels, 1992). There is evidence that episodic structure analysis could be a sensitive measure of children's narrative organization skills (Luo & Timler, 2008; Park, 2014). Park (2014) analyzed the episodic structure of narrative productions made by children with hearing impairments. The author found that these children showed difficulties presenting the logical relations of episodes in a story generation task, which indicates they may not fully understand the temporal and causal relationships between characters and events. A positive correlation was found between the hierarchical episodic structure and the children's performance at the Tower of London and the Digit Span task, supporting the idea that problems in narrative organization are associated with lower working memory and planning skills.

The notion of structure is often confused with the coherence of narration. As seen, the notion of structure concerns the macrostructure of text organization composed of unique rules and guiding principles. Following these rules, a narrator is expected to include an introduction, characters, and a sequence of events developing and leading to the problem's solution and conclusion (Halliday & Hasan, 1976; Ripich & Griffith, 1988). In contrast, coherence refers to *how* the components of a story and the events are interrelated and organized in a meaningful way (Louwerse & Graesser, 2005; Shapiro & Hudson, 1991). To produce a coherent narration, the children need to use a scheme to organize the content that helps the listener to understand the characters and the temporal and logical course of events. Indices lacking coherence are generally considered repetitions and incongruences children produce within their narrations. For instance, Marini et al. (2019) consider errors of global coherence the utterances that merely repeat previously introduced concepts without providing any additional information or those utterances that derail from the flow of discourse or include ideas that are conceptually incongruent with the images shown in the picturebook. Pinto et al. (2016) consider a different index of incongruence in analyzing children's story coherence, i.e., the contradictory utterances produced by children (for instance, the child said, "the monsters want to make peace", and then she contradicts herself by saying, "but the monsters wanted to attack"). To produce a coherent narrative, children must correctly represent the story events they look at in the picture book and their timeline. The study of Veraksa et al. (2020) on the relationship between EF and narrative production in 5-6-year-old children showed that story coherence indices were associated with both visual and verbal working memory.

Regarding cohesion, different indices have been used to assess this variable in children's narratives, including the use of conjunctions, articles and pronouns. Halliday and Hasan (1976) identified interclausal connectives (e.g., temporal and causal connective words) and referencing expressions (i.e., the correct use of nominal and pronominal forms to refer to characters across the story) as devices to make narratives cohesive. Children can create cohesion relatively early, although their ability to correctly use a range of conjunctions increases with age (Shapiro &

Hudson, 1991). Few studies have investigated the association between EF and cohesion obtained through connectives. Artico and Penge (2016) investigated NC in children with learning disorders and found that the number of connectives produced within the narration was associated with SH and PL skills in these children. Concerning pronominal references, whereas preschool-age children can use pronouns and articles, they have problems re-introducing and marking given referents appropriately in narration tasks (Hendriks et al., 2014; Orsolini et al., 1996).

Regarding the introduction of characters within the story, Bamberg (1987), analyzing the narratives of children ranging in age from 3 to 10 years, found that new referents were frequently introduced using a definite noun phrase and that only 50% of the oldest children used an indefinite noun phrase at the first occurrence of a new referent. The connection between referential cohesion and EF was more extensively studied than other macrostructural competencies. For instance, McNiven (2010) found that from kindergarten to grade 2, referential cohesion in maintaining and re-introducing given referents across the story was significantly correlated with updating. Kuijper et al. (2015) investigated some EF skills possibly implied at different moments of narrations. They found that in typically- and atypically-developing children, the ability to re-introduce a character correctly was associated with working memory and inhibitory processes.

Concerning the microstructural level, researchers usually refer to the sentence level productivity and complexity (Justice et al., 2006; Liles, Duffy, Merrit & Purcell, 1995). Typical measures of productivity are the number of communication units, which are defined as the main clause and its subordinate clauses (Hughes, McGillivray & Schmidek, 1997), the total number of words, and the number of different words (Justice et al., 2006), this last also considered as a measure of lexical diversity (Heilmann, Nickerts & Miller, 2010). Indices of syntactic complexity can be the MLU or the clausal structures used in narratives, especially the number of subordinate clauses (Bishop & Dolan, 2005; Justice et al., 2006). Considering narratives' productivity, Westerveld and collaborators (Westerveld, Gillon & Miller, 2004) found differences among 5- 6- and 7-year-olds, but not between 4- and 5-year-olds, in the total number of different words and the total number of words. Justice and collaborators (2006) conducted a study to develop a clinical tool that accounts for microstructural aspects of narrative production for children. Their results highlighted a general linear increase in all the microstructural measures considered through the age of 10 years, with a plateau in performance evident at this later age. In a recent study by Mäkinen et al. (2014) on a group of 172 Finnish children aged between 4 and 8 years, the results showed that older children produce longer (in terms of the number of communication units, number of different words and the total number of words) and syntactically more complex (considering MLU and clausal density) stories than younger children. Microstructural competence was frequently studied in association with EF abilities, but evidence on the correlation with specific EF abilities is mixed. Some authors found a significant moderate association between morphosyntactic complexity in storytelling tasks and interference control skills, whereas others found an association with working memory. For instance, Marini et al. (2020) found a moderate association between morphosyntactic complexity indexed by the number of subordinate clauses produced by children and accuracy rate at a Stroop-like task, whereas other authors found a weak but significant association with digit span and n-back tasks (Drijbooms et al., 2015; 2016; Peristeri et al., 2020), commonly used to measure working memory span and updating. Most studies on the association between EF and NC in storytelling tasks did not find a significant correlation with lexical productivity. Among the studies included in the meta-analysis presented in Chapter 1, for lexical productivity, a significant association in school-aged children was found only with the performance at inhibition tasks like Go/NoGo test and a Walk/Don'tWalk test (Artico & Penge, 2016; Drijbooms et al., 2015). As regards retelling, a significant correlation was found in school-aged children between lexical productivity and performance at verbal working memory tasks (Puranik, 2006).

2.1.2 The present study

The literature on narratives and EF in children leaves many open issues related to the association between narrative indices and EF skills at preschool age. These issues need to be addressed to understand if indices of narrative performance could be helpful in measuring EF skills working on storytelling. Many of these issues are derived from the almost exclusive use of the retelling paradigm and the great variety of tasks used to tap EF abilities in most studies investigating the relationship between EF and NC. In the retelling paradigm, the child's narrative production is influenced by the examiner's narrative (Roch et al., 2016). Despite EF abilities probably implied in retelling and storytelling, even if differently, storytelling is the best way to assess preschool children's NC (Zanchi & Zampini, 2020).

Instead, the vastity of tasks used to address children's EF skills and the variety of indices of the performance (e.g., accuracy, efficiency, speed) considered by various authors makes it particularly hard to have an overview of which EF mechanisms are implied in telling stories. Moreover, most previous studies (e.g., Tonér & Nilsson Gerholm, 2021) have been conducted using "linguistically-based" EF tasks, requiring children to use their linguistic knowledge to complete them. As a result, we do not know much about whether and how the correlation between EF and NC is due to language involvement. A few studies have used "visually-based" tasks that did not require children to use linguistic knowledge to complete tasks (e.g., Peristeri et al., 2020), but the results were inconclusive. Therefore, the aim of the current study is to explore the association of indices of macrostructural narrative competence with preschoolers' EF abilities, using – in contrast to previous studies – a storytelling task to evaluate children's NC and tasks that tapped both verbal, non-verbal and mixed EF abilities. This is to answer the question whether NC can be used as a proxy of EF in preschoolers. Based on the findings of previous studies, we choose to assess the EF domains of working memory span, interference control, shifting, and planning skills. All the tasks administered follow paradigms commonly used in EF assessment with preschoolers. However, we manipulated the requests and added some experimental and control phases to the original tasks in order to control the influence of other executive and non-executive processes that potentially may contribute to the performance (see Methods section). This allows to obtain more precise indices of EF abilities.

Furthermore, we included in the present investigation two original indices of narrative performance that we hypothesized could be related to EF. These indices are the number of scenes the child describes within a single utterance and the number of anticipations of story events the child provides across the story. We hypothesized the former could have been related to child's planning skills, whereas the latter could have been related to working memory.

In summary, the present study investigated in a sample of Italian-speaking preschoolers the association between EF and the following indices of NC:

- Macrostructural competence: coherence (repetitions, incongruence, contradictions); cohesion (ratio of connectives used); story structure (Stein and Glenn's story grammar; GAO episodic structure; hierarchical goal structure)
- Cognitive indices: number of scenes described per utterance; number of anticipations.

Since previous findings on their relationship are confused, and evidence from meta-analytics is not sufficient to make specific hypotheses on the patterns of association between macrostructural indices, and EF domains, the aim of this study is explorative in nature.

2.2 METHODS

2.2.1 Participants

Participants were monolingual children, all native speakers of Italian.

In total, 41 typically developing children who attended kindergarten classes in public school (21 girls and 20 boys, mean age 4.58 years, range 4.0-5.0) were tested.

They were recruited by contacting two public schools in the province of Lecco (Northern Italy) and one in Salerno (Southern Italy). The sample is composed of children who have a different socio-economic backgrounds, with 40% of children living with at least one parent educated to a degree level and 33% living with at least one parent with a high school diploma and the remaining 26% living with a parent that has a middle school diploma or a professional qualification. 50% of the sample had an income below the national middle-class income of $ 35,608 (Kochhar, 2017) at the time of the data collection. Differences in educational level and income are partly explained by the study location, where on average higher income and educational level are observed in participants attending the school located in Lecco, $F(1,34) = 4.672$, $p = .038$; $F(1, 40) = 6.902$, $p = .012$, respectively. However, the children from the low-middle class are equally distributed between the two locations, $F(1, 19) = 0.776$, $p = .390$.

The inclusion criteria for the children were no history of language impairment or hearing loss and no significant exposure to any other language than Italian. To exclude children with possible neurodevelopmental disorders, parents completed the Strengths and Difficulties Questionnaire ([SDQ] Goodman et al., 2000), which is a brief behavioural screening questionnaire, also standardized for Italian preschool populations (Baldo et al., 2017). The questionnaire served also to collect information reported above on socio-economic background of the participants.

Two participants (one male and one girl, both 5 years old) reported Z scores below the norm values adjusted for age and sex, so they were excluded from further, leaving 39 participants in the sample.

2.2.2. Materials and procedure

Children were tested individually on a single day in a quiet room at school by two female research assistants trained by the author of this thesis.

Children were asked to sit at a table facing the examiner and presented the tasks in a predefined order. Half of the children first performed the interference control and shifting tasks, followed by the narrative production task, the working memory tasks, the planning tasks and the TOM task. The other half of the participants received the tasks in reversed order.

Narrative production. We used the 'Narrative Competence Task' ([NCT], Zanchi and Zampini, 2020), a storytelling task created to assess narrative competence in Italian children aged 3-8. The NCT is a wordless 18-picture book created to elicit children's narratives. It was developed with respect to typical story grammar (presentation of the characters, problematic situation, attempts to solve the problem, solution, and conclusion of the story).

The story is about two children, a boy and a girl, who go to the park accompanied by their grandfather and mother, respectively. The children meet and start playing with a ball together. Unintentionally, they throw the ball into a tree and then try to get it back in different ways. At the end of the story, a policeman helps the children to get the ball back, and they resume playing. The described situation is familiar to children because the events included represent what can occasionally occur in a park, and the images are sufficiently simple to be immediately clear (see https://doi.org/10.1027/1015-5759/a000569 for several examples of the pictures).

During the task administration, the children were asked to look through the illustrations to familiarize themselves with the story. They were then invited to tell the story keeping the book on the table in front of them. The situation is a joint attention situation, where the examiners can see the storybook on the table, and the children can spontaneously use language and gestures in telling the story. The examiners could not interfere with the narration. They only supported the children with positive feedback ("Good!" and "Well done!"). If queried by the children, the examiners had to keep their answers as brief as possible and encouraged them to continue with their stories. The task was audio- and video-recorded and later transcribed into CHAT format according to the CHILDES transcription conventions (MacWhinney, 2000) by a native Italian-speaking trained transcriber.

Interference Control. Two tasks were administered to assess interference control: the Fruit Stroop (FS) task (Archibald & Kerns, 1999) and the Fish Task (FT; adapted from Viterbori et al., 2012), which are adaptations of the Stroop and Flanker paradigms, respectively.

The FS task serves to measure a child's ability to handle perceptual interference with the lexical-semantic system. The task is composed of three pages of stimuli. The examiner provided a training trial before each page to reduce working memory involvement in reminding the instructions across the task. In the training trial, the examiner ensured that the child understood and memorised the instructions and provided feedback to the child.

The first page consisted of rows of 15 appropriately coloured fruits (i.e., yellow bananas and red strawberries) arranged pseudo-randomly. The child is asked to name the colours of the fruits as quickly as possible. The second page presented the same fruits in the same positions as on page 1, only devoid of colour. The participant must name the colours that the fruits *should be* as quickly as possible. These two phases are a baseline for controlling colour naming and semantic access for naming purposes.

Page 3 presented the same fruits, arranged differently as on pages 1 and 2, only now were coloured incorrectly (red bananas and yellow strawberries). The child is again required to name the colours that the fruits *should be* as quickly as possible. The instruction given is purposely worded. Since it is the same provided by the examiner on Pages 2 and 3, working memory and shifting demands are more limited in this version of the Stroop task than in other traditional adaptions of the Stroop paradigm for children (i.e., Day/Night task).

Accuracy (i.e., the total number of correct responses, range: 0-15) and the total time of the responses on each page are collected.

A summary measure of the interference score (IS) was calculated following the approach for the Stroop. We first calculate the efficiency of the performance in all three phases by dividing the accuracy rate from the total response time spent on each page (ACC/Time). Then the efficiency score on the third page was subtracted from the average of the baseline efficiency scores (pages 1 and 2) to obtain the IS_{stroop}, where positive values indicate more interference (i.e., fewer linguistically-based IC skills).

$$\mathrm{IS}_{stroop} = \frac{(\mathrm{ACC}_{page1}/\mathrm{Time}_{page1}) + (\mathrm{ACC}_{page2}/\mathrm{Time}_{page2})}{2} - \frac{\mathrm{ACC}_{page3}}{\mathrm{Time}_{page3}}$$

The FT measures a child's ability to handle visual interference.

We used the version of Viterbori et al. (2012), adapting it to be parallel to the FS.

There are three phases: the first phase contains 15 trials with the stimuli are arrows, all oriented in the same direction; the second phase contains 15 trials with the target fish, and the interfering fish are oriented in the same direction; the third phase contains 15 trials with all the target fish, and the interfering stimuli are oriented in the opposite direction.

In all three phases, the child is asked to turn their head where a *centrally located fish* is oriented so that the instruction does not change across the task, reducing shifting and working memory demands. The first two phases serve as a baseline, while the last phase measures the child's ability to inhibit the visual interference provided by the presence of the interfering fish oriented in the opposite direction.

Before each phase, a training trial was provided by the examiner. In the training trials, the examiner ensured that the child understood and memorised the instructions and provided feedback to the child.

Accuracy (i.e., the total number of correct responses, range: 0-15) and the total response time are collected in each phase.

A summary measure of the interference score ($IS_{flanker}$) was then calculated as done for the FS task. We first calculate an efficiency score of the performance in each of the three phases by dividing the accuracy rate from the total response time spent on each phase (ACC/Time). Then the efficiency score in the third phase was subtracted from the average of efficiency scores in phase 1 and phase 2, to obtain the $IS_{flanker}$, where positive values indicate more interference (i.e., fewer visually-based IC skills).

$$\mathrm{IS}_{flanker} = \frac{(\mathrm{ACC}_{phase1}/\mathrm{Time}_{phase1}) + (\mathrm{ACC}_{phase2}/\mathrm{Time}_{phase2})}{2} - \frac{\mathrm{ACC}_{phase3}}{\mathrm{Time}_{phase3}}$$

Shifting. We administered two additional phases at the FS and FT to measure shifting.

Since isolating the verbal component from shifting processes is laborious, these EF abilities should be considered *mixed*, i.e., both linguistically and visually-based.

Shifting at the FS was measured by presenting the child with a page with eight fruits depicted in incorrect colours and seven depicted in the correct colours. When the colour is correct, the child is asked to name the colour of the fruit (i.e., red or yellow); when the colour is incorrect, the child is asked to name the object (i.e., strawberry or banana). Therefore, across the task, the child is asked to shift the response accordingly to the colour of the stimuli. The task did not require interference control skills because when the stimuli are incongruent, the child must label the object name of the fruit. There is well-documented evidence that there are no interference costs in colour-object interference tasks when the child is asked to name the objects, even if they are coloured incorrectly (Prevor & Diamond, 2005). This is because children's prepotent tendency is to say what the object was.

As for the other phases of the FS, the examiner provided a training trial to ensure that the child understood and memorised the instructions and provided feedback to the child.

A score of shifting efficiency ($shifting_{stroop}$) was computed by dividing the total number of correct responses at the incongruent stimuli (i.e., fruits depicted incorrectly, range: 0-8) by the esteem of response time that occurred in those trials.

Also, the FT includes a parallel phase to measure shifting. Shifting was measured by presenting the child with seven trials with the target fish, and the interfering fish were oriented in the same direction, and eight trials with all the target fish and the interfering stimuli were oriented in the opposite direction. When all the fish looked in the same direction, the child was asked to turn their head where the centrally located fish was oriented. When the target fish and the interfering stimuli were oriented in the opposite direction, the children were asked to turn their heads in the direction pointed by *the flankers.* Therefore, the child is asked to shift the response to the stimuli's spatial features across the task. The task did not require IC skills because when the stimuli are pointed in the opposite direction, the child is asked to look in the direction pointed by the flankers. No visual interference costs in the flanker's tasks are supposed to be when the child is asked to point in the direction indicated by the majority.

As for the other phases of the FT, the examiner provided a training trial to ensure that the child understood and memorised the instructions and provided feedback to the child.

A score of shifting efficiency ($shifting_{flanker}$) was computed by dividing the total number of correct responses at the incongruent stimuli (i.e., fish pointed in opposite directions, range: 0-8) by the esteem of response time that occurred in those trials.

Working Memory. Two tasks were administered to assess phonological and visual working memory: the Not this! (NT) task (Howard, 2017) and the Self-ordered pointing (SOP) task (Cragg & Nation, 2007).

The NT is based on the Direction Following Task (Im-Bolter, Johnson, & Pascual-Leone, 2006) and requires children to carry out auditory instructions of increasing complexity. Instructions ask participants to point to a stimulus that *is not* of a particular colour, shape, or size (or some combination of these). The requirement to find a shape that is not of a particular quality is essential to minimise the opportunity to chunk these auditorily presented features, and the instruction is given by the examiner when the child has in front a blank sheet.

The task consists of two trials at each level of complexity (levels 1-8; 16 total trials), the difficulty of which is aligned with the number of stimulus features that must be concurrently held in memory. For instance, at level 1 the examiner may ask the child to "point a shape that *is not* green" (a single feature – green – to hold in mind), whereas at level 3 the examiner may ask the child to "point a shape that is not big, not red and not a circle" (three features to hold in mind – big, red, circle). The task continues until the earlier of completion (at level 8, eight characteristics to remember) or failure to accurately complete both trials within a level.

Since the task posits many demands on other cognitive and perceptual processes, we administered four tasks that serve as a baseline before the real task.

Firstly, we administered a colour-shape-dimension recognition task to ensure that the preschool children could recognise and distinguish between the figures. Secondly, we control for the comprehension of the negative clause using items from the Test for Reception of Grammar (TROG-

2), where the child is required to point to the picture that matches phrases as as 'point the star that *is not* red', from an array of four options. Thirdly, we control for visual perception and scanning abilities, administering a task where the child must find and circle 8 small butterflies distributed in an 8x18 grid on a paper sheet containing 136 small distractors (frogs). The task has a time restriction of 60 seconds.

Forty, we administer a short-term phonological memory task, where the child is asked to point to the figure that matches with the auditory instruction provided (e.g., point to the butterfly that is purple and small – three features to hold in mind). As in the NT task, here, the difficulty is aligned with the number of stimulus features (from three to eight features) that must be concurrently held in short-term memory.

Performance was indexed by the total number of correct responses the child gave (0-16). The phonological span corresponds to the maximum number of features the child could hold in mind. So, for instance, if the child collects 1 point at each level, reaching level 8, the span computed is 8.

The Self Ordered Pointing (SOP) is a traditional working memory task (Petrides & Milner, 1982). In this version, the children were shown a set of pictures of abstract designs and were required to point to a different picture on each trial until all the pictures had been pointed out once. Set sizes of 2, 4, 6, and 9 pictures were used, with a unique set of pictures for each set size. The pictures are pseudo-randomly arranged on the grid sheets (i.e., sometimes in the same and sometimes in different spatial arrangement) to prevent the use of dumb strategies that would allow outperforming the task (e.g., point always the same location or point always a different location). The children completed all conditions of the task in a fixed order. The participants were first shown a demonstration using three pictures of abstract design, which serves as training trials, where the examiner provides feedback.

The use of abstract designs that are hard to put into words was chosen to control the mediation of language and obtain a more precise measure of visual working memory (Hongwanishkul et al., 2005; Joseph et al., 2005). As in the NT task, the difficulty of the SOP is aligned with the number of pictures (1-9) that must be concurrently maintained in memory. The task continues until the earlier completion (at a set of 9 pictures) or failure to accurately point to two pictures within a set. Performance was indexed by the child's total number of correct responses at each set of pictures (0-21). The visual span corresponds to the maximum number of pictures the child could hold in mind. So, for instance, if a child points correctly to two pictures in the first set, four pictures in the second set, five pictures in the third set and then, in the last set, they point correctly to three pictures and then makes two mistakes, their span is five (i.e., the maximum number of pictures they could correctly hold in mind).

Planning. Truck Loading (Carlson et al., 2004) measures planning abilities in preschoolers. In this task, children must load and deliver party invitations using a toy truck while adhering to 4 rules: (a) the street is one-way, (b) one can drive around the block only once, (c) the colour of the invitation must match the colour of the paper house, and (d) invitations must be taken only off the top of the pile from the back of the truck. The experimenter first demonstrated the game using two houses and two invitations; then, the children completed a warmup by delivering two invitations with the examiner's help. After a rule check, the task begins. One new house was added for each successive level of difficulty, ending with five houses and four levels of difficulty. For each level, children received three trials condition: the normal condition is the one just described (truck loading $_{normal}$); the non-verbal condition, where the task must be performed while the child must say repetitively "ba-ba-ba" (truck loading$_{art}$); the double-task condition, where the task must be performed while the child must stomp their feet (truck loading$_{double}$). In truck loading$_{art}$, the request served to suppress the private speech to have a measure of the task performed, limiting language mediation. Therefore, the performance was considered an index of visually-based planning skills. In truck loading $_{double}$, the request served to monitor and exclude the cost of the double request in truck loading $_{art}$. In truck loading $_{normal}$, we could not exclude the involvement of both visual and verbal processes; therefore, the performance at this phase was considered an index of *mixed* planning skills.

For each level, the children had to pass the trial in the normal condition to continue to the next level. For each trial, in every three conditions, children obtained a score of 1 if they stacked the invitations in the correct order and a score of 0 if they failed. The time taken to load the invitations onto the trucker was also collected. Self-corrections were permitted only during the loading phase. For each condition, the dependent variable was the efficiency score, computed by dividing the number of correct orders by the total time taken.

Theory of Mind (TOM). TOM was assessed by administering the Sally and Anne test (Baron-Cohen et al., 1985), widely used in the literature on child development. In this task, the child is presented with a story about two characters: Sara and Anna (the original name "Sally" and "Anne" used by Baron-Cohen et al. were modified in the present study). First, Sara put a ball into her box. Then she left the scene, and Anna transferred the ball into her box, wanting to play a trick on Sara. After that, Sara came back, and the experimenter asked the child the critical belief question, "Where will Sara look for the ball?" followed by two control questions ("Where is the ball really?" and "Where was the ball in the beginning?"). If the child pointed to the ball's former location and correctly answered the control questions, they could represent Sara's wrong belief and pass the test. The picture of the story was retrieved and adapted from Bolander (2014).

The child obtained a score of 1 if they passed the test and a score of 0 if they failed.

2.2.3 Coding of narratives

Both macrostructural and microstructural characteristics of children's narrative competence were considered. A detailed description of each index is reported in the following paragraphs.

Macrostructural Level. Each child's storytelling during the NCT was videotaped, transcribed, and then coded following the coding system described below:

Quantity of information. The the number of "events", the things that occurred in the story (e.g., "la bambina salta con la corda" [the girl is jumping with the rope]) and the number of "agents", the number of characters that were performing an action in the story, were counted. Descriptions (e.g., "c'era un triciclo" [there is a tricycle]) were not coded as events, and characters who were named but did not perform any action were not coded as agents. This allowed for a distinction between children telling a story and children only describing or labelling the pictures in the book. A maximum of 49 events and 7 characters were counted based on adults' narrations examined in previous studies (see Zanchi & Zampini, 2020). Adding the number of events and agents, a child could have a raw score ranging from 0 to 56 points on the quantity of information index. The percentage of the quantity of information was calculated as follows: [(events + agents /56)∗100].

Structure. To evaluate the story structure, three indices were computed: (1) Stein and Glenn's story structure; (2) GAO Episodic Structure; (3) Hierarchical goals structure.

The first index is based on Stein and Glenn's story grammar. We adapted the procedure developed by Norbury and Bishop (2003), giving 1 point to each key passage included in the narrative (i.e., initiating event, problem, attempts to solve the problem, turning point, solution, and conclusion). A maximum of 8 key passages was counted based on adults' narrations examined in previous studies (see Zanchi & Zampini, 2020). Each child received 1 point for each of the key elements included in their narrative. Therefore, children could have a raw score ranging from 0 to 8 points on the story structure index. The percentage of elements included in the story structure was calculated as follows: [(total key elements mentioned/8)∗100].

The second index considers the number of episodes containing a Goal-Attempt-Outcome (GAO) unit. The number of episodes composing the story structure and the GAO units were identified based on adult narrations examined in previous studies (Scionti et al., submitted; Zanchi & Zampini, 2020). Goals refer to statements which describe the character's plan for addressing the ongoing events (e.g., "lui voleva prendere la palla" [he wanted to retrieve the ball]). Attempts refer to statements which describe the character's actions to achieve the goals ("sale sul triciclo" [he climbs on his tricycle]. Outcomes refer to statements which describe the consequences of the attempts ("ma non riesce" [but he could not reach the ball]). A maximum of four episodes was counted. We give a point of 1 for each complete episode and 0 for an episode missing one or more components of a GAO unit. Therefore, children could have a raw score ranging from 0 to 4 points on the GAO episodic structure index. The percentage of complete episodes was calculated as follows: [(complete episodes /4)∗100].

The third index considers the goal-plan hierarchy of children's narratives, analysing the presence of superordinate and subordinate. The superordinate goal is the goal that the protagonists in a narrative ultimately want to achieve (e.g., start playing again with the ball). The subordinate goal refers to the sub-goal used to obtain the superordinate goal (e.g., retrieve the ball that ended up on a tree). The goal plans of the NCT were analysed using protocols from Trabasso and Nickels (1992). In each narrative, superordinate and subordinate goals were identified and coded as G1 and G2, respectively. When a narrative has both superordinate and subordinate goals, 1 point was given. 0 points were given when the narrative had only a superordinate goal or only a subordinate goal.

Coherence. Using an adaptation of the procedure employed in the study by Marini et al. (2019), we counted the number of errors in global coherence.

Global coherence errors included the production of utterances that: (1) repeat previously introduced concepts without adding any new information (e.g., "questo qua andava in triciclo / poi ho visto questo qui andava in triciclo [this one was riding a tricycle / then I saw this one was riding a tricycle]); (2) include ideas that are conceptually incongruent with the images shown in the picturebook (e.g., "un giorno la bambina stava giocando col triciclo" [one day the girl was playing with the tricycle]); (3) conceptually incongruent utterances that contradict what was previously told (e.g., la palla cade / non cade [the ball fell down / the ball does not fall down]). Adding the number of repetitions, incongruences, and contradictions made by a child, a general score of global coherence errors was obtained. The percentage of global coherence errors was calculated as follows: [(global coherence errors / total number of utterances)∗100].

Cohesion. We adopted the use of connectives and the anaphoric use of the article within the narration as indices of cohesion.

For the anaphoric use of the article, we considered the passage from the indefinite article, typically used when a character is introduced in the story, to the definite article appropriate for an already given object or character. A maximum of 8 passages from indefinite to definite articles were counted based on adults' narrations examined in previous studies (see Zanchi & Zampini, 2021). A child received 1 point for each of these passages included in the narratives. The percentage of passages from indefinite to the definite article was calculated as follows: [(total score of passages produced/8)∗100].

As regards connectives, we use an adaptation of the procedure employed in the study by Pinto et al. (2015), counting both causal and temporal linguistic connectives:

- Causal: So, then, thus, consequently, because, therefore, it follows that, it comes out that, to this aim, in that case, it turns out that, as things stand, as things do not stand, for this reason (e.g., "sono tristi perchè la palla è lassù" [they are sad because the ball is up there]).
- Temporal: Then, after, afterwards, subsequently, right at that moment, before that, in the end, in origin, at the beginning, beforehand, in conclusion, at the end, suddenly, soon, in the meantime, until, at this moment, in the first place, until now, from now on, (e.g., "poi ci prova lei con la corda" [then she tries with the rope]).

Then, the percentage of total connectives was computed over the total number of words produced during the storytelling: [(total connectives/ total number of words)∗100].

Microstructural Level. To control for child's linguistic competence, we considered the microstructural level (i.e., the language used in the narration) of narration produced, as was done in the study by Blom and Boerma (2016). The following measures were considered:

Lexical diversity. To assess children's lexical skills, we used the D index, which was obtained using the VOCD command of CLAN (Malvern, Richards, Chipere, & Purán, 2004). The measures taken into account by the D index are: (1) Type, that is the number of different words used by the child (e.g., "bambino, bambini" [child, children] are computed as one type); (2) Token, which is the total number of words produced during the storytelling (e.g., "bambino, bambini" [child, children] are computed as two tokens). D index is based on a mathematical model of how the type/token ratio varies with token size; therefore, it is not a function of the number of words in the sample, at least for texts between 100 and 400 tokens (McCarthy & Jarvis, 2007).

Mean Length of Utterance (MLU). The MLU measures grammatical complexity (e.g., Devescovi et al., 2005; Rice, Redmond, & Hoffman, 2006); indeed, more morphosyntactically complex sentences are generally longer than simple ones. This value is computed by considering the mean number of words per utterance (MLU = tokens/ total number of utterances).

Syntactic complexity. For each child, the total number of subordinate clauses was counted. Both implicit subordinates, which are characterised by the presence of a verb in indefinite mode (infinitive, gerund, past participle), and explicit subordinates, which are characterised by the presence of a verb in finite mode (indicative, subjunctive, conditional, imperative), produced by children, were included. When multiple subordinates were present within the same utterance, all subordinates produced were counted. Coordinate sentences were not included in the calculation of subordinates (e.g., in "the girl jumps with her rope and the boy goes on his tricycle", there are no subordinates present).

Cognitive indices in the narratives

Number of scenes per utterance. The number of scenes the child describes within a single utterance was counted. The utterance was defined following a prosodic criterion, that is, the lowering of intonation (Bolinger, 1978) and slow in speech rate (Klatt, 1975) or "any understandable change in the conversational turn" that is perceived as an interruption of the discourse (Suttora et al., 2017). For each utterance, the number of images it referred to was identified and coded as "pag1" if the utterance refers to the visual content of one page; "pag2" if the utterance refers to the visual content of two pages, "pag3" if the utterance refers to the visual content of three pages. Most utterances produced by children referred to one page at a time; none produced utterances referring to more than three pages at a time. Therefore, we choose to compute the percentage of utterances referred to more than one page over the total number of utterances produced [(pag2 + pag3 / total number of utterances)∗100].

Anticipations. The number of anticipations of story events the child provides across the story was counted. Anticipations are identified and coded as “ant” each time the utterance produced by a child refers to events that will happen the child could remember and choose to anticipate to the listener. For instance, looking at the image in which the girl is trying to retrieve the ball with her rope, the narrator says, “lei cerca di prenderla con la corda ma non riesce” [she tried to catch it with the rope but she failed]. Here, the narrator anticipates the listener that the protagonist’s attempt will fail, which will be only shown in the subsequent picture. Anticipations are partly similar to incongruencies because they provide conceptually incongruent ideas with the images shown in the picture book. However, contrary to incongruencies, anticipations did not reflect an incoherent representation of the story. Children who produce anticipations must have a clear and coherent representation of the story events in memory. The reported content is incongruent with the image shown at that moment but not with what happened in the storyline. We compute the percentage of anticipations over the total number of utterances produced [(anticipations / total number of utterances)∗100].

2.2.4 Reliability

The number of utterance, MLU, and lexical diversity, were counted automatically in CHAT based on the transcripts. The first author coded the transcripts for the macrostructural and cognitive indices. To assess the inter-coder reliability, a random 20% of these transcripts were coded independently by a second coder. For all the measures, the percentages of agreement on the detection of elements were determined as follows: Quantity of information 91% (Events 89%, Agents 93%); Story structure 87% (Stein and Glenn’s story structure 92%; GAO Episodic Structure 89%; Hierarchical goals structure 80%); Coherence 95% (Repetitions 100%, Incongruences 87%, Contradictions 98%); Cohesion 88% (Anaphoric use of article 90%, Connectives 85%); Number of Scenes per utterance 86%; Anticipations 89%.

2.2.5 Data Analyses

Means, ranges, and standard deviations for NC and EF measures are reported in Table 2.1. In summary, we tested for the association between the following indices of NC:

- Macrostructural competence: coherence (repetitions, incongruence, contradictions); cohesion (ratio of connectives used); story structure (Stein and Glenn's story grammar; GOA episodic structure; hierarchical goal structure)
- Cognitive indices: number of scenes described per utterance; number of anticipations.

And the following indices of EF abilities:

- Linguistically and visually-based working memory span
- Linguistically and visually-based IC without shifting demands
- Shifting without linguistically and visually-based IC
- Mixed and visually-based PL

Controlling for the influence of:

- Some non-executive processes potentially implied in the performance on EF tasks (i.e., TOM)
- Demographic variability of the sample (age, sex, socio-economical status [SES])
- Linguistic competence shown in narratives (microstructural competence: lexical variety, morphosyntactic and grammar skills).

A series of stepwise regression models was performed to explore the relationship between NC and EF. Initially, all NCT variables and other variables (e.g., TOM and non-executive processes implied in EF performance) that could influence the relationship between NC and EF were included in the models. Then a backward elimination approach was adopted to estimate the best NCT indices related to EF performance. We considered changes in the variance explained by models (R^2) and the effect size of the estimates to determine the best model fit.

The regression approach was preferred over correlation analysis because it allows controlling for potential confounding variables of the association between EF and NC. However, this analysis aims not to determine the impact of one variable on the other but only the strength and the sign of their relationship.

Table 2.1 Means, ranges, and standard deviations for NC and EF measures

	4 years old	**5 years old**	**All sample**	
	M (SD)	M (SD)	M (SD)	Range min-max
N	16	23	39	–
% Female	68	43	54	–
Raven's Matrices	66 (21.82)	62 (28.15)	62.97 (25.50)	16 – 98
SES	4.73 (1.033-)	5.23 (0.75)	5.03 (0.89)	2 – 7
SDQ	-0.454 (0.71)	-0.038 (0.70)	-0.21 (0.73)	-1.71 – 1.44
Executive Function				
Fruit Stroop (Interference Score)	0.38 (0.239)	0.19 (0.224)	0.27 (0.25)	-0.211– 0.758
Fruit Stroop (Shifting Score)	0.27 (0.143)	0.29 (0.156)	0.28 (0.15)	0 – 0.750
Fish Task (Interference Score)	0.11 (0.095)	0.09 (0.076)	0.10 (0.08)	-0.045 – 0.291
Fish Task (Shifting Score)	0.21 (0.060)	0.23 (0.065)	0.22 (0.06)	0.087 – 0.328
Not This! (verbal span)	3.19 (0.981)	3.00 (1.567)	3.08 (1.34)	1 – 6
Self-Ordered Pointing (visual span)	5.06 (1.289)	5.83 (1.696)	5.51 (1.57)	3 – 9
Truck loading (normal condition)	0.13 (0.086)	0.13 (0.094)	0.13 (0.08)	0 – 0.355
Truck loading (articulatory suppression)	0.10 (0.058)	0.13 (0.095)	0.12 (0.08)	0 – 0.421
Truck loading (double task)	0.11 (0.060)	0.13 (0.084)	0.12 (0.07)	0 – 0.296
Narrative Competence				
NCT - MLU	6.34 (2.00)	7.33 (1.39)	6.92 (1.71)	3.75 – 10.11
NCT – d index	40.26 (7.78)	39.39 (8.18)	39.75 (7.93)	25.78 – 50.90
NCT – syntactic complexity	7.69 (3.62)	8.83 (4.08)	8.36 (3.89)	0 – 16
NCT – Number of Utterances	23 (7.18)	21.43 (6.19)	22.08 (6.57)	14 – 42
NCT – % story structure	63.28 (21.63)	67.93 (15.45)	66.02 (18.12)	25 – 100
NCT – % GAO episodic structure	32.81 (28.45)	50 (23.83)	42.95 (26.87)	0 – 75
NCT – Hierarchical Goal Plan	0.44 (.512)	0.57 (0.51)	0.51 (0.51)	0 – 1
NCT – % quantity of information	38.39 (10.94)	38.04 (10.25)	38.18 (10.40)	17.85 – 62.50
NCT – % variety of connectives words	28 (9)	30 (21)	29 (17)	10 – 100
NCT – connectives over the total words produced	9.489 (4.68)	8.27 (2.94)	8.77 (3.74)	1.66 – 20.87
NCT – % anaphoric use of article	17.18 (18.18)	21.19 (15.73)	19.55 (16.67)	0 – 62.50
NCT – % coherence errors	8.36 (9.25)	4.658 (5.30)	6.17 (7.31)	0 – 27.27
NCT – % images per utterance	2.97 (4.43)	4.65 (6.92)	3.72 (5.58)	0 – 27.77
NCT - % anticipations	8.75 (7.53)	9.92 (7.40)	9.44 (7.38)	0 – 25

Note. SES = Socio-economic status; SDQ = Strengths and Difficulties Questionnaire (Preschool Version); NCT = Narrative Competence Task; MLU = Mean Length Utterance; GAO = Goal-Attempt-Outcome.

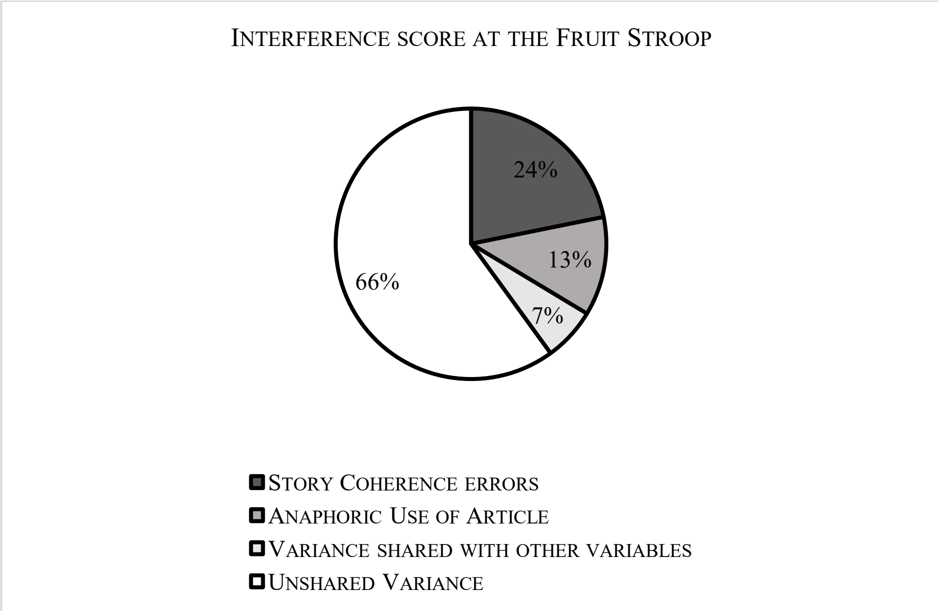
2.3 RESULTS

2.3.1 Narrative Performance Indices and Linguistically-based/Mixed EF abilities

The relationship between NC and linguistically-based EF indices was examined by computing multiple stepwise univariate regressions with interference score$_{stroop}$ shifting$_{stroop}$ working memory$_{not\ this}$ as the outcome, and all the collected narrative indices as independent variables. For the relationship involving mixed EF indices, we computed univariate stepwise regressions with $shifting_{stroop}$ $shifting_{flanker}$ truck $loading_{normal}$ as the outcome and the same narrative indices as independent variables. TOM, age, SES and baseline sessions of the related outcomes were added as covariates to capture the possible effects of these variables on the relationship. To control for type I error a *p* value of less than .003 (.05/13 = .003) was adopted across the computed regressions of interference $score_{stroop}$ $shifting_{stroop}$ $shifting_{flanker}$ and truck $loading_{normal}$ a *p* value of less than .002 (.05/17 = .002) was computed for regression on working $memory_{not\ this}$. Medium and large effect sizes are also reported. The models excluded NCT indices that did not reach a statistical significance step by step, revealing some notable patterns of association between narrative performance indices and linguistically-based EF abilities.

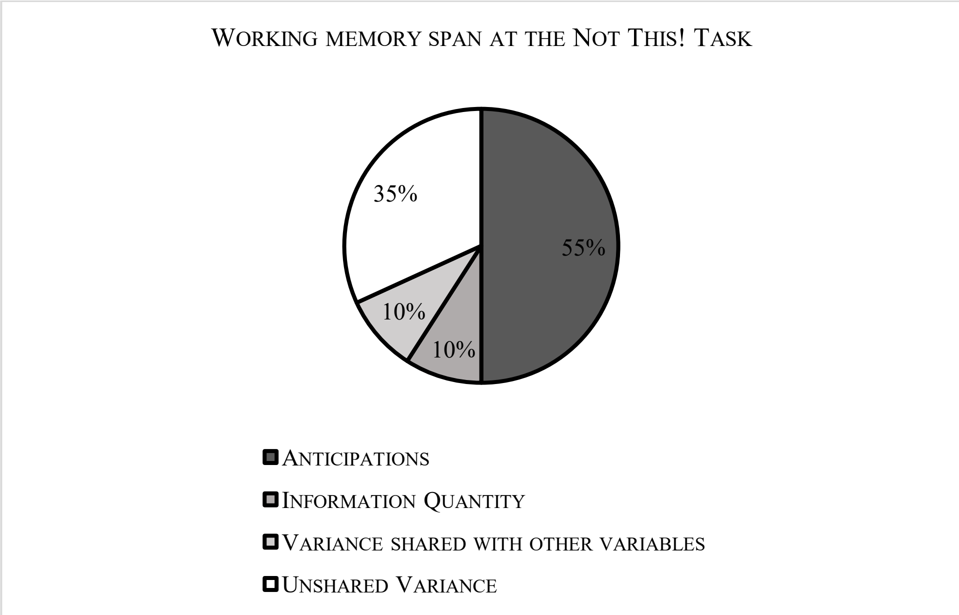
As far as interference $score_{stroop}$ is concerned, the full model were all variables are included, shared the 57% of the variance of the performance, $F(13, 25) = 4.295$, $p < .001$). After 13 iterations removing all the indices and the variable that did not significantly associate with IS_{stroop}, the final model shared the 61% of variance of the dependent variable ($F(3, 35) = 17.294$, $p < .001$). This model included as independent variables the following variables: story coherence errors ($\beta = 0.502$, $p < .001$, $sr^2 = 0.237$); anaphoric use of article ($\beta = -0.365$ $p = .002$, $sr^2 = 0.130$); anticipations ($\beta = -0.279$ $p = .016$, $sr^2 = 0.073$). In accordance with the significance levels set to control for type I error ($p < .003$), only the indices of story coherence errors and anaphoric use of articles was significantly associated with interference $score_{stroop}$. The two indices, controlling for the other variables, together shared 37% of the variance of interference $score_{stroop}$ performance (see Figure 2.1).

Figure 2.1 Best NCT-indices of Interference Score at the FS task.



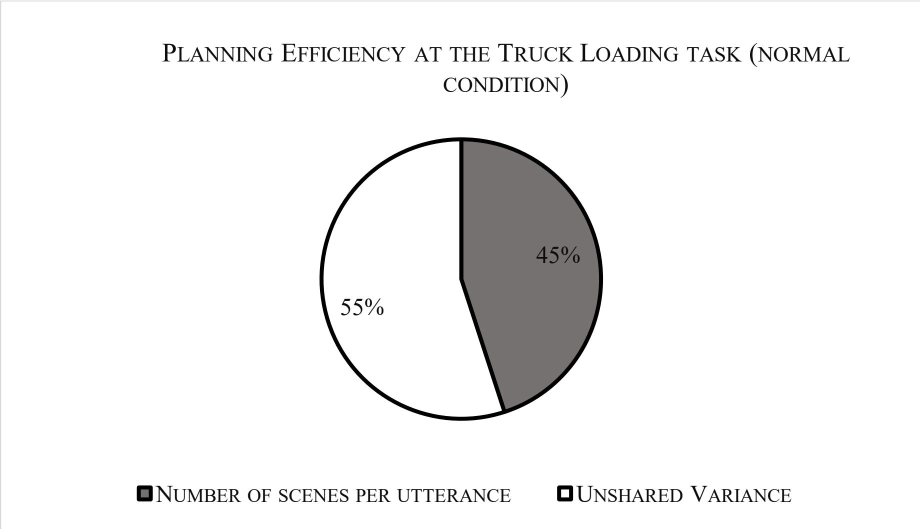
As far as working memory$_{\text{not this}}$ is concerned, the full model where all variables plus processes eventually implied at the NT Task (i.e., colour-shape recognition, visual scanning, negative clause comprehension and visual scanning) are included, sharing the 67% of the variance of the performance, $F(19, 19) = 5.12$, $p < .001$). After 16 iterations removing all the indices and the variable that did not significantly associate with working memory$_{\text{not this}}$, the final model still share the 67% of variance of the dependent variable ($F(4, 34) = 19.84$, $p < .001$). This model included as independent variables the following variables: information quantity ($\beta = 0.338$, $p = .002$, $sr^2 = 0.097$); number of scenes per utterance ($\beta = 0.262$, $p = .017$, $sr^2 = 0.055$); anticipations ($\beta = 0.725$, $p < .001$, $sr^2 = 0.551$); TOM ($\beta = 0.229$, $p = .035$, $sr^2 = 0.042$). In accordance with the significance levels set to control for type I error ($p < .002$), only the indices of anticipations and information quantity were significantly associated with working memory$_{\text{not this}}$. The two indices, controlling for the other variables, together share 65% of the variance of working memory$_{\text{not this}}$ performance (see Figure 2.2).

Figure 2.2 Best NCT-indices of Verbal Working memory span at the NT task.



As regards mixed EF abilities, performance in the truck loading$_{normal}$ that may involve language mediation in the performance resulted being associated with the number of scenes described per utterance, $\beta = 0.674$, $p < .001$ with a moderate effect size ($sr^2 = 0.454$). The index alone shares the 45% of variance of truck loading$_{normal}$ performance, $F(1, 37) = 30.74$, $p < .001$ (see Figure 2.3). This model was estimated after 15 iterations that excluded step by step all the variable included in the full model, $F\ (15, 23) = 2.38$, $p = .030$.

Figure 2.3 Best NCT-indices of Planning Efficiency at the TL task.

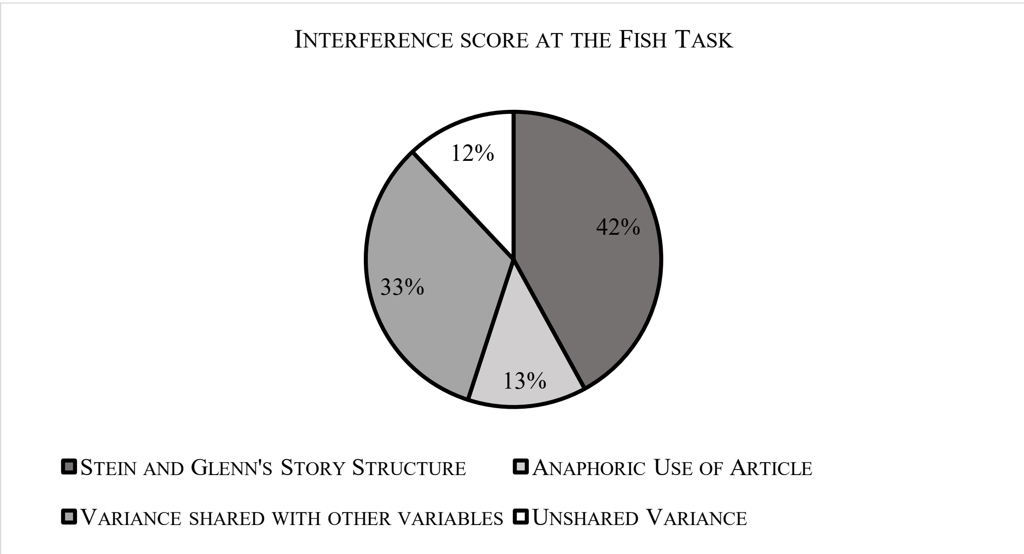


Shifting processes measured (both at the FS and FT) did not appear to be associated with any of the narrative indices collected.

2.3.2 Narrative Performance Indices and Visually-based EF abilities

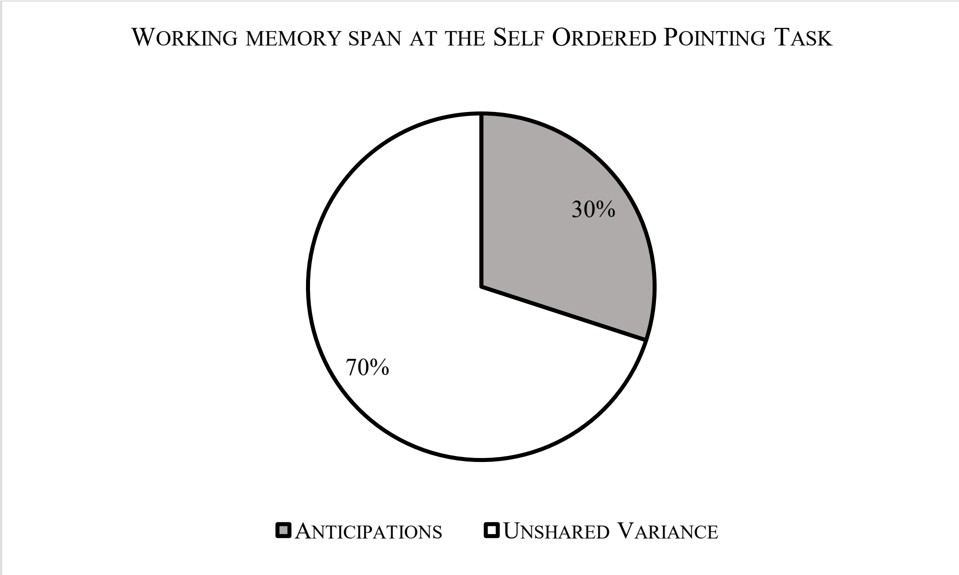
The relationship between NC and visually-based EF indices was examined by computing multiple stepwise univariate regressions with interference $score_{flanker}$ working $memory_{sop}$ truck $loading_{art}$ as outcome and all the narrative indices collected as independent variables. TOM, age, sex, SES and baseline sessions of the related outcomes were added as covariates to capture possible confounding effects of these variables on the relationship. To control for type I error across the computed regressions a *p* value of less than .003 (.05/14 = .003) was adopted. Medium and large effect sizes are reported in the text. The models excluded NCT indices that did not reach a statistical significance step by step, revealing some notable patterns of association between narrative performance indices and visually-based EF abilities. As far as interference $score_{flanker}$ is concerned, the full model, where all variables are included, shared the 53% of the variance of the performance, $F(15, 23) = 3.903$, $p < .001$). After 11 iterations removing all the indices and the variables that did not significantly associate with interference score $_{flanker,}$ the final model shared 62% of the variance of the dependent variable ($F(5, 33) = 8.77$, $p < .001$). This model included as independent variables the following variables: Syntactic complexity ($\beta = -0.348$, $p = .024$, $sr^2 = 0.066$); anaphoric use of article (β = -0.432 p = .002, sr^2 = 0.128); MLU (β = -0.661 p < .001,sr^2 = 0.167); number of scenes described per utterance (β = -0.371 p = .008, sr^2 = 0.094); Stein and Glenn's story structure (β = -0.928 p < .001, sr^2 = 0.423). In accordance with the significance levels set to control for type I error (p < .003), among macrostructural indices, only the indices of anaphoric use of article and Stein and Glenn's story structure were significantly associated with of interference $score_{flanker}$. These indices, controlling for the contribution of the other variables, share 55% of the performance variance (see Figure 2.4).

Figure 2.4 Best NCT-indices of Interference score at the FT.



As regards working $memory_{sop}$, the full model, where all variables are included, shared 21% of the variance of the performance, F(15, 23) = 1.665, p = .132). After 15 iterations removing all the indices and the variable that did not significantly associate with working $memory_{sop,}$ the final model was composed of only one index that shared the 28% of the variance of the dependent variable, F(1, 37) = 15.805, p < .001 (see Figure 5). This index is the number of anticipations made by children within narration (β = 0.547, p < .001, sr^2 = 0.299).

Figure 2.5 Best NCT-indices of Visual Working memory span at the SOP.



Controlling for the effect of demographic differences and double task demands, truck loading$_{art}$ did not result in associating with any of the narrative performance indices collected. However, when the double task condition was not included in the model, truck loading$_{art}$ was significantly associated with the index of hierarchical goal structure (β = 0.557, p < .001, sr^2 = 0.223), suggesting that the capacity to perform the double request of the task could be involved, not the planning process per se.

2.4 DISCUSSION

The present study's purpose was to better define the relationship between oral narrative production and EF abilities. This purpose was defined as a first step that would lay the groundwork for the study of NC as a potential ecologically valid way to explore the workings of EF across development. Previous studies on the elder adult population conducted by Cannizzaro and Coelho (2013) have suggested the assessment of NC may provide a balanced, ecologically valid measure of goal-directed behaviour under the larger construct of SEC knowledge. The authors found that narrative discourse structure (i.e., story grammar) was strongly related to performance on both linguistically and visually-based measures of EF and that their variability is captured by the same latent factor. They claimed that story grammar, just as EF performance, may represent a type of goal-directed thinking and action similar to other knowledge stored and processed by adults in the prefrontal cortex (Grafman 1995; Mar, 2004; Ylvisaker et al. 2008). The challenge of the present study was to understand if this could be valid also for preschool age, which represents a critical moment for NC and EF development. Previously, a meta-analytic study showed that EF abilities are significantly associated with macrostructural NC in early childhood more than in late childhood and adolescence. In line with that research, our results revealed some strong patterns of association between EF and NC that might suggest that some indices of narrative performance could be very useful to address EF abilities in an ecological way.

Both visually and linguistically-based tasks used to measure interference control resulted associated with the index of anaphoric use of articles. This index refers to the children's ability to mark the newness of new characters introduced in the story using the indefinite article and then, once the characters are introduced, refer to the same characters with a definite article. Children in the sample performed a few correct passages from indefinite to definite articles to refer appropriately to the already given characters. This is not because they did not refer to given characters with a definite article but because they used the definite article from the beginning to refer to a new entity and then kept referring to it using definite articles or pronouns within the narration. The finding is in line with previous studies that analysed the narratives of children ranging in age from 3 to 10 years and found that new referents were frequently introduced using a definite noun phrase and that only 50% of the oldest children used an indefinite noun phrase at the first occurrence of a new referent (Bamberg, 1987). The number of passages correctly performed seems to be an index of interference control skills in children. Regardless of the type of interference and response elicited by the task, the process of controlling an interference appear similar to the process of inhibiting the pronoun/definite noun phrase for referring to a new entity. The reason why definite noun phrases/ pronouns could be compared to interference is grounded on the grammar constraint that less informative forms (i.e., pronouns) are preferred from the speaker's perspective to explicit forms, such as indefinite noun phrases (Hendriks et al., 2008; Gundel et al., 1993).

Pragmatic rules showed that speakers use pronouns to refer to a character when a pronoun is informative enough in the context of use for the listener to identify the intended referent. Otherwise, if the pronouns are unrecoverable from the listener's perspective, an explicit form (i.e., definite noun phrase) must be used. The situation where the child is asked to tell the story may explain why children with poor interference control introduced new characters and objects with pronominal forms or definite noun phrases more frequently. In the assessment of narrative skills, we asked the child to look at the picture book alone before starting to tell the story. Thus, the children already knew the story they told. Maybe, children's previous knowledge of the story created an interference at the moment to introduce the characters – the new entities were not so "new" in our children's minds – making it more difficult for those with poor inhibition skills to select an indefinite noun phrase instead of a definite noun phrase/pronoun.

Furthermore, the situation of mutual knowledge in which the children are asked to tell the story could make it harder for children with poor IC to discard the pronoun/ definite noun phrase. If the listener shares the attention on the picture book with children, a child with poor interference control skills might easily choose to rely on shared knowledge and present the character not as something new that needs to be signaled with an indefinite article. The extra cognitive effort posited by this kind of situation was handled only by children with higher interference control skills.

As regards the specific ability to handle perceptual interference with the lexical-semantic – and not visual interference – we found an association with story coherence. Story coherence is a global index that refers to errors the children make in their production when they repeat without adding new information, contradict previously introduced concepts, or include ideas that are conceptually incongruent with the images shown in the picture book and with the story plot. Among these error types, 83% of these were incongruence errors. Children who made these errors reported contents that were not coherent with what happened in the storyline or what they could observe on the page. For instance, in the story plot, the boy finds a ball in the story and then asks the girl to play with him with the ball. Looking at these scenes, a 4-year-old child tells the protagonists have shown the ball to their parents, and their parents say to them, “who left the ball there?” and the children answer, “we don’t know, we have found it”. Then, in another scene where the protagonists are playing with the ball, the child says that the girl says to the boy that he has beautiful curly hair. At a certain point in the storyline, the boy throws the ball at a turtle’s shell. The turtle gets scared, and the boy sneers. Looking at this scene again, the child said that the boy was sorry for throwing the ball at the turtle. At the story’s end, the picture shows the children returning to play with the ball. Here the child makes up that the children come home.

Children who make these kinds of errors tell a story that is not the one illustrated. They follow their imagination, add details and interpret facts following the personal and imaginative representation of the story they have created. These children also showed less efficiency in controlling interference with the lexical-sematic system at the FS task. Here, the children are presented with a grid of fruits that are depicted in the wrong colour, and they are asked to name as quickly as possible the colour the fruit had in reality, not the colour the fruits are depicted. The ability to produce the correct response (i.e., name the colour the fruit had in reality) without the intrusion of other relevant but unintended target behaviour (e.g., naming the colour is shown) is similar to telling story events without incongruent information. Children who struggle in controlling interference also struggle in controlling their representation of the story events. In both cases, the task demands the child to control their verbal behaviour and act oppositely to what they used to do (i.e., naming the colour / telling the events they see). Success depends on the focused and correct completion of target verbal behaviours, while interferent representations are inhibited. This association also found consistency in previous literature. In Cannizzaro and Coelho’ (2013) study, there is an index referring to non-episodic information, personal insights and tangentially related information reported by adults within narrations. Although the authors did not use the term “story coherence” to refer to it, it is similar to what we considered as indices of story coherence in our study. Notably, the authors found a significant relationship between the number of non-relevant details/non-episodic information the adults reported and the performance at the Stroop task. The two measures aligned onto the same factor the authors define as "Organization-Efficiency" because "both measures possess qualities of organization and the efficient structuring of information within a given framework without the inclusion of unnecessary information or information that is not salient in reaching the stated goals of the task" (p. 543).

The child's ability to handle visuospatial interference resulted associated with the general index of story structure. The index is based on Stein and Glenn's story grammar and reflects the child's ability to mention all the key passages in the narrative (i.e., initiating event, problem, attempts to solve the problem, turning point, solution, and conclusion). The task we used to measure children's ability to handle visuospatial interference required them to turn their heads to the direction pointed by a *centrally located fish*, ignoring the visual interference provided by the presence of the flankers fish that are oriented in the opposite direction. The task is manipulated to be as parallel as possible to the FS. In both tasks, the conflict is perceptual, but in the FT the conflict is given by visual flanker's effect (not a conflict with semantic representation) and the target behaviour required is a motor – not a verbal behaviour. There are no shifting demands across the task because in all the trials administered, the centrally located fish and the flankers look in opposite directions. Thus, the child is asked to keep the focus on the fish at the centre, avoiding distractions.

Children who tell well-structured stories are those who can keep focused on the FT, avoiding distractions provided by the flankers. Notably, the association does not have to do with the number of events or agents (information density) the children mention within the story but with the number of key events mentioned. Children more able to handle visual interference, staying focused on the target fish, are the same that, when looking at the picturebook to tell the story, were able to focus their attention and tell the listener the key events. Crucially, "events" are not merely descriptions of the pictures but refer clearly to what is happening in the story. This distinction is fundamental to conceiving story events according to the story's grammar perspective. For instance, to state the problem situation, the utterance "the ball ends up above a tree" or "the ball goes to the tree/ goes there" is valid as an event. Instead, the utterance "it's on the tree" is not valid as an event. Again, in the final scene of the picturebook, the utterance "the children play with the ball" is not a valid event, while "the children return to play" is. Analysing the story structure of children with low visual interference control (i.e., lower than the median score of the performance), we noticed that 68% omitted to mention the initiating event (i.e., "the boy asks the girl to play with him, and she accepts"). These children skip the part the protagonists decide to play together. After introducing the two characters playing separately with their toys, these children tell they are playing with the ball / they are throwing the ball around. 56% of the stories produced by children with low visual interference control skills miss the conclusion, and 87% miss at least one of the attempts made by the protagonists to retrieve the ball. Therefore, we observed these children reported the events as descriptions of single pictures without giving them a structure or skipping the pages omitting to tell the key events. This observation is in line with the results provided by Berman and Slobin (2013). The authors found by age 5, most children describe the problem event, but only 50% of them include attempts to solve the problem and the conclusion within their narrations.

The significant association of the FT with the more general story structure index and not with the indices of GAO episodic structure and hierarchical goal structure might indicate that the core of this association is not on the goal structure. Goal structure captures the child's ability to understand and organise the narration around the characters' goals and their goal-oriented actions and reactions. For instance, a child who had a clear representation of the goal of the protagonists could tell the listener, "the girl tried to recover the ball (goal) with their rope (attempt), but she failed (outcome) because the rope was too short". Instead, a child who reports the same episode saying, "the girl used the rope to catch the ball, but the rope fell down", may not have a clear representation of the story from the goal structure perspective. The actual outcome of the character's action is the failure related to the goal, in the character's goal perspective. "The rope fell down" is an event but not the outcome of the character's action from the goal structure perspective. Goal structure reflects a complex structured knowledge of the story events, probably more related to planning or organization skills than interference control. Instead, the child's ability to tell essential story events without skipping them or reporting them as mere images description may need simpler EF processes, like those involved in the FT.

The performance at the tasks we used to assess verbal (NT) and visual working memory (SOP) was associated with the NCT cognitive index of anticipation. In this work, we wanted to collect an index of verbal and visual working memory performance as precisely as possible, reducing the impurity caused by perceptual processes involved that could potentially influence working memory function. For this purpose, we tested for child's ability to (1) easily recognise stimuli's features we used (i.e., colours, dimensions and geometric shapes), (2) comprehend negative clauses, (3) hold a small amount of verbal information in an active, readily available state for a short interval of time and (4) efficiently, quickly, and actively look for information relevant to the environment**.** All children performed correctly in over 50% of control trials. Anyway, we included the indices of performance at these tasks in the regression model, as well as other variables that could be at least in part responsible for the variance shared by working memory and NCT index of anticipations (e.g., TOM) or be moderators of such relationship (e.g., age, SES). The anticipations index was developed based on observations that, when they tell stories they already knew, preschool children do not tell the events of the page they are looking at but sometimes anticipate to the listener some elements or events of the story that have yet to happen. When looking at the picture book for the first time, the child is asked to look carefully at the illustrations to familiarise themselves with the story. At this point, children seem to construct a representation of the story events structure they kept in mind during storytelling. This representation may manifest externally through anticipations of story contents. In the sample, 32 out of 39 children produced at least one anticipation in their narrations. On average, children produced 2.65 anticipations, with older preschoolers using them more frequently than younger ones. Analysing a posteriori the contents of anticipations qualitatively, we observed that the contents and events more often anticipated by the children within narrations are:

- The event that the ball ends up on the tree, which represents the problem of the story
- The failure of the attempts made by the two protagonists
- The arrival of the policeman that comes to help the children, which represents the turning point of the story

These represent some of the most important key passages of the story plot. However, data showed that children who produced more anticipations were not the same who included more key passages in the narratives (r = .140) or more information (r = .112). Indeed, among children who produced a small percentage (0-7%) of anticipations (n = 17), eleven children included over 60% of key events in the narrations. Anticipation is a spontaneous behaviour that can occur or not in telling stories, but it does not associate with story grammar. Instead, it is associated with working memory, with a large effect size for both linguistically and visually-based working memory processes. Previous studies on the structure of working memory across childhood investigated whether verbal and visuospatial working memory are supported by a common resource pool, or whether they are maintained by separable cognitive resources (Baddeley, 2000; Shah & Miyake, 1996). Recent studies provide evidence that supports more a multi-component model of working memory that includes a domain-general processing component and separate domain-specific storage components (Gathercole et al., 2004; Alloway et al., 2006). The tasks administered to measure working memory differ in storage components, but the performance of the tasks is statistically associated (r = .41), which supports evidence that both the tasks tapped the working memory processing component and that this component is associated with the anticipation index.

Children look at the illustrations in the book, store and process them supported by domain-general working memory. Then, they could manifest the content stored and processed in working memory through anticipations when telling the story. This suggests that when we have the chance to listen to anticipations produced by children in narrations, this may be interpreted as an index of children's high general working memory competence. Even if the relationship is bidirectionally significant, a low rate of anticipations within children's narration should not be interpreted as an index of low verbal working memory span. Being careful is necessary for several reasons. Firstly, the occurrence of anticipations within narratives is quite a rare event. Despite 82% of the sample producing at least one anticipation per story, only 20% of children produced more than 4 anticipations in their narrations. Secondly, contrary to macrostructural indices of NC (e.g., story coherence and cohesion), the absence of anticipations within children's narration did not reduce the quality of the narration or affect the global structure. Thirdly, a rate of anticipations has not been reported in previous research on NC of adults or children, and we did not have data to compare to our results. Further studies are necessary to understand better this index's value for working memory assessment in both child and adult populations.

In addition to anticipations, information density was revealed to be an index associated with working memory – but this time, only verbal working memory is concerned. The result is interesting because the setting of storytelling did not require the child to remember story events and agents as in retelling paradigms. In storytelling, the children flip through the pages and tell the listener what they see. However, as seen previously, when looking at the picture book for the first time, children construct a representation of the story that may be kept in mind during storytelling and interfere in several ways with narrative production. Although the task does not directly trigger children's memory, some children, during storytelling, rely on the representations of events and characters they have stored in memory. Since performance at NT and SOP are correlated, and only the performance at NT results are associated with the Information Quantity index, the EF process associated with the quantity of information reported by children refers specifically to the linguistic representation stored. It is possible that children who showed more difficulties in storing and processing verbal information did not store and process representations related to events and characters occurring within the story during the first view. Therefore, they might report less information than children who successfully rely on information stored and processed in their minds during the first view of the book. At the same time, when children had good verbal working memory skills, they could keep in mind representations of events and characters they saw, and when telling the story, they used to report these events.

As regards shifting processes, the performance at the tasks we used to assess them was not associated with any of the collected NCT indices. Previous studies on children aged 8 to 11 found that shifting measured by card sorting tasks was significantly associated with story structure indices. Some authors claim shifting processes might be involved in the generation of complete episodes within a narrative discourse and in the ability to monitor the communicative flow (Mozeiko et al., 2011). Indices of NCT considering story structure, episodic structure, or discourse planning seem to be more related to interference control and planning skills in our sample. Comparing these results with other studies may be contrived since few efforts have been done in past studies to isolate executive (e.g., interference control) and non-executive components from the shifting tasks. As for planning skills, it is hard to isolate the verbal component from shifting processes. Therefore, we had to consider the processes as mixed EF abilities, with a difference in the output required: a verbal response for the shifting $_{stroop}$ and a motor response for the shifting $_{flanker}$. In both tasks, we tried to control the interference control demands, asking the child a response that did not require handling a Stroop or Flanker effect but only required a shift from the previous set. In this way, we wanted to reduce the inhibition component frequently involved in complex shifting tasks. It is possible that previous studies (see, for example, Khan et al., 2013), finding a significant association between shifting and episodic structure, tap the inhibition component involved in shifting tasks and other functions.

Finally, the performance at the tasks used to assess planning skills associated with the number of scenes described per utterance. In order to monitor the role of language in planning skills, we administered the task in normal conditions – leaving the child to use private discourse to perform the task – and in conditions of articulatory suppression – to limit the role of language.

Despite the average performance efficiency across the conditions being similar, only the performance at the task in the normal condition was associated with the aforementioned NCT index.

This suggests that, in narratives, the number of scenes described per utterance was not an index sensitive to a child's general planning skills, but it could be helpful to address specific discourse planning aspects. Instead, when the linguistic component was isolated from planning (truck $loading_{art}$), no NCT indices resulted significantly associated with the EF performance. Moreover, the fact that performance at truck $loading_{art}$ was associated with hierarchical goal structure only when the performance efficiency at the double task condition was not included in the model reveals that the executive component related to the NCT index was not planning *per se*. It is the multitasking component of the two conditions they have in common. This means that children who map the protagonist's goals into their stories, specifying the subordinate and superordinate goals related to their actions, are those more able to execute more tasks at a time. This ability probably refers more to children's organizational skills.

2.4.1 Limits and conclusions

The present work had some relevant limitations that must be taken into account. The first limitation refers to the small sample size involved in the study. A small sample size may cause a problem of underpowered statistics. This means that a relationship that may exist in reality is actually not captured and detected by statistical analyses. Maybe in the present study, some indices of narrative performance that do not relate to EF abilities are actually related to them.

Secondly, since our study wants to explore possible indices of narrative performance that could be useful to address EF abilities ecologically, we collected and coded many indices within 39 narrative productions. This increases the chance of observing a false positive effect, especially with a small sample size. To prevent this, we adopted a more severe confidence interval to reject the null hypothesis (99.7 – 99.9% CI). However, given such limitations, further studies and replications of these results are needed. The final goal of this study is ambitious: understand which indices of narrative performance could be helpful to observe the online working of EF abilities from an ecological perspective. Different studies have established a connection between the development of

storytelling and EF abilities. We do not know if the development of one influences the development of the other unidirectionally or bidirectionally. However, we know they are correlated across development, especially at preschool age and before seven years of age. Studies on the adult population shed light on the fact that impairments in narrative performance and EF abilities are profoundly interrelated and could be seen as expressions of the same cognitive dimension.
Studies on the children population are further behind. With this study, we discovered some precise indices of narrative performance that could be useful to tap EF abilities. This is important because we need tools others than questionnaire for addressing EF in an ecological way.

We found that the anaphoric use of the article within narration could be a valuable index to address children's IC skills regardless of the nature of the stimulus that has to be inhibited or the type of response required. The number of anticipations children make within their narrations could be a valuable index to address children's high working memory skills. These indices may lead to observing the working of interference control and working memory as domain-general EF processes.
Looking at domain-specific EF processes, we found that story coherence errors made by children could be a specific index to observe the child's ability to handle interference with the lexical-semantic system. Instead, the number of events relevant from the story grammar perspective cited by children within their narrations may be helpful to observe their ability to handle visual interference and stay focused on the critical elements. The number of scenes described per utterance could be valuable to observe the aspects of planning related to discourse - but not planning skills in general. These associations between specific narrative indices and EF abilities are relevant because they are still significant when controlling for variables generally hypothesized to be a responsible part of the variance shared between NC and EF processes. This makes the pattern of association we found pretty solid, despite the limitations discussed above.

CHAPTER 3

Keeping track of characters: cognitive mechanisms underlying referential choices in pre-schooler's oral narratives.

3.1 INTRODUCTION

Reference is a fundamental function of language. This function allows us to talk about the world surrounding us and allows us to indicate to others what things in the world we are talking about. What makes reference in language far from trivial is the fact that speakers can choose between various forms when referring to a particular referent in the world: cross-linguistically, there are complex referential systems based on large varieties of linguistic expressions. For example, when referring to a toy, English speakers can use the indefinite noun phrase (NP), *a toy,* the definite description *the toy,* the demonstrative pronoun *this* or *that,* or the neuter pronoun *it.*

Referencing is an interesting aspect of narrative production because it indexes the children's ability to tie sentences together at the local level and draws on the children's pragmatic knowledge of the listener's needs (Norbury and Bishop, 2003). Listeners must select the referent intended by the speaker from a range of potential referents. Referring expressions can be highly ambiguous and refer to various referents, depending on their context. There is no simple or "one-to-one" correspondence between one referent and one expression. The same female character, for instance, can be mentioned in very different ways, going from highly informative forms, such as the full indefinite NP "*a little girl with a rope,*" to less explicit forms, such as the pronoun "*she*". Moreover, during storytelling, the cognitive status of the referents varies in accessibility as the narration progresses. For instance, it is influenced by the presence and prominence of other character entities, leading the narrator to re-adjust their referential choices depending on the level of cognitive accessibility of the targeted referent. This paper will examine how Italian-speaking preschool children *versus* young adults adjust their referential choices when referring to more or less accessible character entities in a complex narrative context. Adult narrators use referring expressions that are unambiguous with respect to the narrative discourse (Hendriks et al., 2014). They predominantly use an anaphoric strategy of using pronouns to maintain a reference and nouns to (re)introduce references. They avoid pronouns in situations where the use of a pronoun may result in a non-intended interpretation (e.g., when the entity referred to is not topical or syntactically prominent in the previous clause). In contrast, young, typically developing (TD) children often struggle with appropriate referring expression choices (Orsolini et al., 1996).

Why do young TD children have difficulties in their choice of referring expression, and, more generally, which cognitive skills are needed to come to a correct decision about which form to use? The present study aims to elucidate the cognitive mechanisms involved in reference choices.

Most referential theories assume that the choice of any particular referential expression is closely connected to the accessibility that the referent is assumed to have at a given moment in the discourse representation (e.g., Givon, 1983; Ariel, 1990, 2001; Gundel et al., 1993; Chafe, 1994; Gundel, 1998). Factors that can affect the accessibility of an entity are numerous and heterogeneous, resulting from different linguistic and non-linguistic sources (Ariel, 2001; Arnold, 2010). Different discourse features have been identified at the linguistic level to affect reference accessibility. In particular, a referent that is *given* (i.e., previously introduced in the discourse) and *topical* in the discourse (i.e., mentioned in a syntactically prominent position, like the subject position) is generally more accessible for the listener (Ariel, 1990; Gordon et al., 1993; Chafe, 1994; see also Arnold, 2010; Arnold et al., 2013). While the topical referent is considered to be the most salient in the discourse (Grosz et al., 1995), “given” referents may not always be highly accessible, leading to a gradient of “givenness.”

In the narrative context, the accessibility level reflects the level of knowledge the narrator has shared with the listener: the narrator must provide the level of information that coincides with the level of cognitive accessibility of the referent for the listener (Ariel, 1996; Cornish, 1999). Not all story characters are in the narrative foreground at any given point (Brown & Yule, 1983; Gundel, Hedberg, & Zacharski, 1993). Thus, the appropriate level of information will depend on whether the speaker is referring to a new character for the first time (*introduction*), linking to an immediately previously mentioned character (*maintenance*, [- topic shift]), or referring to a previously mentioned character that is not topical or syntactically prominent in the previous clause (*re-introduction;* [+ topic shift]).

Cross-linguistically, there is agreement about the more accessible a referent is, the more reduced the expression used by the narrator, and conversely, the less accessible a referent is, the more explicit the referential expression should be (Arnold, 2010) in order to emphasise the entity's role and to disambiguate the reference.

Italian speakers can rely on a system of referential expressions where the phonologically and semantically weak forms to refer to an object are null subject pronouns – signalled by person/number inflection on the verb – and clitic object pronouns, followed by overt subject pronouns and strong object pronouns that occupy an intermediate level, and then indefinite and definite NPs that occupy the top levels (Berretta, 1990). Adjusting the theory (Ariel, 1990) to the Italian referential system, null and clitic pronouns are more expected to refer to highly accessible referents in subject and object position, respectively. In contrast, a large range of "intermediate" expressions, including overt subject, strong object pronouns, demonstratives, and definite NPs, are more expected for referents with a low accessibility level.

Overall, salient and semantically rich linguistic forms as NPs are used to introduce new referents and re-introduce old ones. Overt subject pronouns, along with other stressed pronoun forms (e.g. demonstrative pronouns), may sometimes be used in referent re-introductions to signal a topic shift or when the referent's role needs to be emphasised (Orsolini et al., 1996). On the contrary, null subject pronouns and clitic object pronouns would be predominantly used in referent maintenance.

3.1.1. Development of referential choice per discourse functions in children

As early as the age of two, children show sensitiveness to whether a referent is available in the preceding discourse context, producing null and overt subjects in a pragmatically appropriate way (Serratrice, 2005). However, until late childhood, children show some difficulties in the choice of referring expressions, using forms that sometimes are unrecoverable for a listener. These difficulties occur differently depending on referential discourse function (introduction, maintenance, re-introduction).

Signalling to the listener that a new character is introduced in the story or that the discourse on a previously introduced character is continuing are basic pragmatic functions reflecting the speaker's and listener's sharing of the story context representation. Considerable literature focused on developmental changes in referent introduction and showed that as early as age 2, children used to select salient and explicit forms as NPs to refer to new entities in the discourse context. (Serratrice, 2005). Across languages, research indicates that introduction adequacy increases from preschool to the early school years and is generally high (above 60%) in six-year-olds children (Chen & Pan, 2009; Schneider & Hayward, 2010). At preschool age, most errors in referent introduction corresponded to the overuse of definite NP, which incorrectly signalled given entities. Other studies comparing children's performance in the presence and the absence of mutual knowledge (Hickmann, 2003; Kail, 1998; Kail and Sanchez y Lopez, 1997) showed that children do not reliably use indefinite determiners for referent introduction and definite determiners for referent maintenance until approximately 7 years of age. It is generally around the age of 9 that the articles system is consistently used adequately for referential purposes within narrations.

Evidence on referent maintenance showed that this referential function is the easiest to produce for children. The literature indicates that, from as young as four, children predominantly use reduced pronominal forms to maintain reference, showing early sensitivity to the information status of the referent in question (Bamberg, 1987). In null-subject languages such as Italian, children predominantly use null forms to maintain the reference to a character across the story (Orsolini et al., 1996; Hickmann and Hendriks, 1999).

A study on Cantonese-speaking children aged 3, 5, 7, and 12 years also showed that children of different age groups do not differ in the adequacy levels obtained for referent maintenance. The same study showed that adequacy levels became stable for introduction from age 7 and increased between all age groups for re-introduction. Therefore, longitudinal studies supported the hypothesis that maintenance adequacy develops first, followed by introduction, and lastly by re-introduction. Re-introduction is the most demanding referential function for children. Evidence on children's adequacy in re-introduction is more mixed. Allowing the listener to unambiguously assign one of the entities introduced in previous discourse to the current referential expression seems to be a particularly complex pragmatic skill. It involves (a) the ability to monitor the narrative discourse, integrating the current clause in the preceding discourse context, and (b) the capacity to take into consideration the listeners' access to referents in the story context so that if referents are not easily accessible for them, the use of more explicit referential forms is needed.

In children, the selection of a referential expression has been shown to rely, at least partly, on a "subject thematic strategy" whereby the main character of the story tends to be pronominalised, while secondary characters are more likely to be re-introduced by a NP (Karmiloff-Smith, 1981, 1985). A more recent cross-linguistic study of narratives in Chinese, English, French and German (Hickmann, 2003) showed that local co-reference significantly affected the morpho-syntactic form of the referential expression chosen by children in all age groups (four to five-year-olds, seven-year-olds, ten-year-olds). In non-coreferential contexts (e.g., re-introducing non-topical character), the children were significantly more likely to use a nominal rather than a pronominal form. In a study on the re-introduction of referents in narrative production of Italian 4-to-10-year-olds children, Orsolini et al. (1996) reported that regardless of age, full NPs were the most frequent forms for referent re-introduction, followed by null subject pronouns and object clitic pronouns. Furthermore, an analysis of the inferrability of the referent's antecedent based on the semantics of the verb and the structure of the preceding discourse context revealed that in some cases, the use of reduced forms (e.g., a null subject or clitic object pronoun) was still adequate. However, the results of this study also showed, as far as preschool age is concerned, children tend to exploit the importance of a character in the story plot or the visual availability of the referent in the non-verbal context as properties that make an entity accessible enough to prevent the narrator from using explicit referring expressions such as full NPs. Unlike older school-age children, pre-schoolers failed more often in judging when null or clitic forms were appropriate in the re-introduction, using null forms irrespective of the referent's accessibility.

The study of Bamberg (1987) showed that children aged 3-4 are driven in re-introduction by the referent's topicality, using pronouns to re-introduce reference to the main character and NP for the other character. Instead, older children are less influenced by these aspects and use full NPs to re-introduce the reference irrespective of the character's importance in the story plot. Although already at first grade (6 years), children show to master the distinction between maintenance and re-introduction functions, some studies revealed that even at the age of 11, children are not fully aware of possible referential ambiguities they did.

3.1.2. Cognitive mechanisms

The present study aims to investigate different mechanisms that may underlie referential choices made by preschool children. Literature suggests that referential choices made by children are affected differently by discourse function (introduction, maintenance, re-introduction of characters). Especially as far as preschool age is concerned, children struggle with referents' introduction and re-introduction, whereas they show adult-like competence in referents' maintenance.

Referential competence seems to rely on at least two abilities:

(a) the ability to monitor the narrative discourse, integrating the current clause in the preceding discourse context.

(b) the capacity to take into consideration the listeners' access to referents in the story context so that if referents are not easily accessible for them, the production of more explicit referential forms is needed.

In cognitive terms, the first ability relies on working memory (WM) capacity. WM capacity is considered one of the core executive functions (EF) in Diamond's model (2013). It requires the ability to hold information in mind and mentally work with it (e.g., relating one thing to another using information to solve a problem, Diamond, 2013). For a correct representation of the discourse, the child needs a sufficient WM span. Instead, the second ability clearly refers to Theory of Mind (TOM) and Inhibitory processes. For taking into consideration the listener's access to a referent in the story context, the child has to mentalise the other's perspective, so Theory of Mind (TOM) processes are expected to be necessary. Moreover, besides TOM, to produce a referential expression that is optimal and recoverable for the listener, the child must also inhibit the optimal form from the speaker's perspective (i.e., the pronoun). Due to the constraint that pronouns are preferred to full NPs (see Hendriks et al., 2008), children tend to select pronouns to refer to a given referent. However, if a child can consider others' perspectives, the child may calculate that the pronoun may not be interpreted correctly by the listener. If so, if the child can also inhibit the pronoun, they would discard the pronominal form and select an explicit form instead. Suppose, for example, that the child wishes to refer to a non-prominent referent, e.g., *the girl*, that occupies the object position in the previous clause (1):

(1) Il poliziotto prende in spalla *la bambina* [the policeman puts *the girl* on his shoulders]

(a) E *ø* riesce a prendere la palla [and *ø* reachs the ball]

(b) E *la bambina* riesce a prende la palla [and *the girl* reaches the ball]

(c) E *lei* riesce a prendere la palla [and *she* reaches the ball]

If the child uses a null pronoun (a), the listener will interpret this pronoun as referring to the most prominent referent in the discourse, i.e., the policeman, in accordance with the constraints of Italian grammar (Berretta, 1990). As this is not the referent intended by the child, the child must discard the null pronoun (a) and use the more explicit full NP, as in (b), or an overt pronoun, as in (c). So, this step involves the child's TOM and interference control (IC). IC indeed refers to the ability to inhibit, suppress, or de-activate competing internal representation to focus attention on goal-relevant information (Nigg, 2017).

Although previous studies have suggested that IC, WM and TOM may be involved in referential adequacy, few studies have directly investigated the relation between these mechanisms and the choice of referring expression made by children.

Hendriks et al. (2014) investigate referential choice in children, young adults and elderly adults to determine whether different speakers egocentrically base their referential choice on the preceding discourse or also take into account the perspective of a hypothetical listener. They found that, on average, 4–7-year-old children, in contrast to adults, have trouble considering the listener's perspective, so they tend to overuse pronouns to refer to all given referents, resulting in massive production of ambiguous pronouns that are unrecoverable for a listener. In contrast, the elderly adults showed a clear sensitivity to the listener's perspective in narrative production. Still, they appeared to lack the necessary cognitive capacities (i.e., WM) to keep track of the prominence of discourse referents, producing more potentially ambiguous pronouns than young adults, though fewer than children. The results of this study suggest that adequate referential choice depends on perspective-taking in language and cognitive capacity, such as the ability to keep track of the prominence of discourse referents. In another study, Kuijper et al. (2015) examined the cognitive mechanisms underlying the choice of referring expression at different discourse moments in a sample of 6–12-year-olds, typically and atypically developing children. They found support for the view that children who re-introduce given referents with NP are those who passed TOM tasks and showed higher WM capacity.

3.1.3. The present study

Narratives evaluation is a "naturalistic" approach to studying language development because they represent a real and contextualised request for children (Schraeder et al., 1999). Moreover, considering the several competencies needed to tell a story, narrative evaluation allows a multi-componential approach to studying language development. The experimental situation we used in this study is where the narrator and listener share their attention on a picture book, reflecting the joint attention situation that children and adults generally live when telling a story.

The study aims to investigate preschooler's ability in constructing a referentially cohesive oral narrative production and possible cognitive mechanisms that underlie this skill.

In narrative productions of children and adults, consistently with previous studies, we distinguished three moments corresponding to three referential functions: 1) Introduction of new referents, 2) Maintenance of references, and 3) Re-introduction of a previously mentioned referents that are not the discourse topic (or the most syntactically prominent referents) at that moment.

We are interested in how children use specific and less specific referring expressions in these three moments. We compare the referential choice made by children *versus* young adults throughout the same narrative in these three different referential functions.

As found by previous studies, the choice of referring expression is expected to vary during these three particular moments in discourse. We expected a higher proportion of NPs in referent (re)introductions and a higher proportion of null pronouns and clitic pronouns in referent maintenance. Furthermore, in line with previous studies, we expect to find significant differences between adults and children in referential choices made to introduce and re-introduce a character in the narrative discourse but not to maintain referents.

Since preschool children are still developing adequate referencing skills, we hypothesized that the referential choice made especially in (re)introduction might be related to the general linguistic competence they showed in narrative production.

Then, we investigate the role of TOM, WM and IC in explaining the difference between adults and children in the referential choice made in different moments across the story. Since the need to keep track of referents and take the listener into account may vary across referent functions (introduction, maintenance, re-introduction), we expect that TOM, IC and WM may have different effects on referential choice at the three discourse moments. However, we chose to investigate the impact of these mechanisms on referential choices only at moments of the discourse in which children's choices significantly differ from adults. In line with the previous findings of Kuijper et al. (2015), we hypothesized that, since sufficient WM span is needed to keep referents activated in mind, WM span would not be so related to the choice of referring expression to introduce a new character. Furthermore, since a speaker can rely solely on the discourse to introduce a new character, we do not hypothesise significant associations between the choice of referring expression and TOM or IC. Previous studies have found no significant relations between TOM, IC, WM and the referring expressions chosen for referent maintenance. Furthermore, we do not expect significant variations in the referential choice made by children vs adults, so we do not expect a significant role of these cognitive mechanisms at such a moment.

Regarding referent re-introduction, speakers generally cannot rely solely on the discourse and use the more economical pronoun but need to consider the listener's perspective and select a more specific form. Therefore, in line with Kuijper et al. (2014), we predict TOM and IC are related to the choice of NPs when re-introducing a referent. WM is also hypothesized to be associated with the choice of NP in referent re-introduction since narrators should keep track of the accessibility of the referents within the discourse.

In summary, the present study will provide a detailed examination of the choice of referring expression made by Italian pre-schoolers *vs* young adults. Furthermore, it will provide insights into linguistic competence and the cognitive mechanisms implied in constructing a referentially cohesive oral narrative production.

3.2. METHODS

3.2.1. Participants

Participants were monolingual children and young adults, all native speakers of Italian.

In total, 41 typically developing children who attended kindergarten classes in public school (21 girls and 20 boys, mean age 4.58 years, range 4.0-5.0) were tested.

They were recruited by contacting two public schools in the province of Lecco (north Italy) and one in Salerno (south Italy). The sample is composed of children who have different socio-economic characteristics, with 40% of children living with at least one parent educated to a degree level and 33% living with at least one parent with a high school diploma and the remaining 26% living with a parent that has a middle school diploma or a professional qualification. 50% of the sample had an income below the national middle-class income of $ 35,608 (Kochhar, 2017) at the time of the data collection. Differences in educational level and income are partly explained by the study location, where on average higher income and educational level are observed in participants attending the school located in Lecco, $F(1,34) = 4.672$, $p = .038$; $F(1, 40) = 6.902$, $p = .012$, respectively. However, the children from the low-middle class are equally distributed between the two locations, $F(1, 19) = 0.776$, $p = .390$.

The inclusion criteria for the children were the following: no history of neurological, language impairment or hearing loss, and no significant exposure to any other language than Italian. In order to exclude children with possible neurodevelopmental disorders, parents completed the Strengths and Difficulties Questionnaire ([SDQ] Goodman et al., 2000), which is a brief behavioural screening questionnaire, also standardised for Italian preschool populations (Baldo et al., 2017). Two participants reported Z scores below the norm values adjusted for age and sex, so they were excluded from further, leaving 39 participants in the children group.

The young adult group included 22 participants from the province of Lecco and Salerno (13 women and 9 men, mean age 25 years, range 20-33 years). Inclusion criteria were being Italian native speakers, absence of neurological or hearing problems. 70% of the sample have a high school diploma, and the remaining 30% have a degree diploma.

3.2.2. Materials and procedure.

Children were tested individually on a single day in a quiet room at school by two female research assistants that were trained by the first author of this study. Children were asked to sit at a table facing the examiner and presented the tasks in a predefined order. Half of the children first performed the IC task, followed by the narrative production task, the WM task and the TOM task. The other half of the participants received the tasks in reversed order.

Interference Control (IC). The Fruit Stroop (Archibald & Kerns, 1999) task was administered to assess IC. The task measures a child's ability to handle perceptual interference with the lexical-semantic system. It is composed of three pages of stimuli. In order to reduce WM involvement in reminding the instructions across the task, the examiner provided a training trial before each page. In the training trial, the examiner ensured that the child understood and memorised the instructions and provided feedback to the child.

The first page consisted of rows of 15 appropriately coloured fruits (i.e., yellow bananas and red strawberries) arranged pseudo-randomly. The child is asked to name the colours of the fruits as quickly as possible. The second page presented the same fruits in the same positions as on page 1, only devoid of colour. The participant must name the colours that the fruits *should be* as quickly as possible. These two phases serve as a baseline for controlling child's ability of colour naming and semantic access for naming purposes.

Page 3 presented the same fruits, arranged differently from pages 1 and 2, only now were coloured incorrectly (red bananas and yellow strawberries). The child is again required to name the colours that the fruits *should be* as quickly as possible. Since the instruction provided by the examiner on Pages 2 and 3 is the same, WM and shifting demands are more limited in this version of the Stroop task than in other traditional adaptions of the Stroop paradigm for children (i.e., Day/Night task). Accuracy (i.e., the total number of correct responses, range: 0-15) and the total time of the responses on each page are collected.

A summary measure of the interference score (IS_{stroop}) was calculated following the approach for the Stroop. We first calculate the efficiency of the performance in all three phases by dividing the accuracy rate from the total response time spent on each page (ACC/Time). Then the efficiency score on the third page was subtracted from the average of the baseline efficiency scores (pages 1 and 2) to obtain the IS_{stroop}, where higher positive values indicate more interference (i.e., poorer IC skills).

$$\mathrm{IS}_{stroop} = \frac{(\mathrm{ACC}_{page1}/\mathrm{Time}_{page1}) + (\mathrm{ACC}_{page2}/\mathrm{Time}_{page2})}{2} - \frac{\mathrm{ACC}_{page3}}{\mathrm{Time}_{page3}}$$

Working memory (WM). The Not this! Task (Howard, 2017) was administered to assess phonological WM. The task is based on the Direction Following Task (Im-Bolter, Johnson, & Pascual-Leone, 2006) and requires children to carry out auditory instructions of increasing complexity. Instructions ask participants to point to a stimulus that *is not* of a particular colour, shape, or size (or some combination of these). The requirement to find a shape that is not of a particular quality is important to minimise the opportunity to chunk these auditorily presented features, and the instruction is given by the examiner when the child has in front a blank sheet.

The task consists of two trials at each level of complexity (levels 1-8; 16 total trials), the difficulty of which is aligned with the number of stimulus features that must be concurrently held in memory. For instance, at level 1, the examiner may ask the child to "point a shape that *is not* green" (a single feature – green – to hold in mind), whereas at level 3, the examiner may ask the child to "point a shape that is not big, not red and not a circle" (three features to hold in mind – big, red, circle). The task continues until the earlier of completion (at level 8, eight characteristics to remember) or failure to accurately complete both trials within a level.

Since the task posits many demands on other cognitive and perceptual processes, we administered four tasks that serve as a baseline before the task. These four tasks were used to ascertain that the performance at the WM task was not biased by deficit in other perceptual processes involved. Firstly, we administered a colour-shape-dimension recognition task to ensure that the preschool children could recognise and distinguish between the figures. Two children struggle with shape recognition, pointing out less than 50% of the shapes correctly. However, they were included in the final sample for the analysis because, in the experimental phase, they could recall 6 and 7 shapes correctly each. Secondly, we control for the comprehension of the negative clause using items from the Test for Reception of Grammar (TROG-2), where the child is required to point to the picture that matches phrases such as 'point the star that *is not* red', from an array of four options. Thirdly, we control for visual perception and scanning abilities, administering a task where the child must find and circle 8 small butterflies distributed in an 8x18 grid on a paper sheet containing 136 small distractors (frogs). The task has a time restriction of 60 seconds.

Fourthly, we administer a short-term phonological memory task, where the child is asked to point to the figure that matches with the auditory instruction provided (e.g., point to the butterfly that is purple and small – three features to hold in mind).

As in the Not this!, here, the difficulty is aligned with the number of stimulus features (from three to eight) that must be concurrently held in short-term memory.

Performance was indexed by the total number of correct responses the child gave (0-16). The phonological span corresponds to the maximum number of features the child could hold in mind. So, for instance, if the child collects 1 point at each level, reaching level 8, the span computed is 8.

Theory of mind (TOM). Theory of mind was assessed by administering the Sally and Anne test (Baron-Cohen et al., 1985), which has been widely used in the literature on child development. In this task, the child is presented with a story about two characters: Sara and Anna (the original name “Sally” and “Anne” used by Baron-Cohen et al. were modified in the present study to be more familiar to Italian children). First, Sara put a ball into her box. Then she left the scene, and Anna transferred the ball into her box, wanting to play a trick on Sara. After that, Sara came back, and the experimenter asked the child the critical belief question, “Where will Sara look for the ball?” followed by two control questions (“Where is the ball really?,” and “Where was the ball in the beginning?”). If the child pointed to the ball’s former location and correctly answered the control questions, then they showed the ability to represent Sara’s wrong belief and passed the test. The picture of the story was retrieved and adapted from Bolander (2014). The children obtained a score of 1 if they passed the test and a score of 0 if they failed. This score was used in the model estimate.

Narrative production. We used the ‘Narrative Competence Task’ ([NCT], Zanchi and Zampini, 2020), a story-telling task created to assess narrative competence in Italian children aged 3-8. The NCT is a wordless 18-picture book designed to elicit children’s narratives. It was developed with respect to typical story grammar (presentation of the characters, problematic situation, attempts to solve the problem, solution, and conclusion of the story).

The story is about two children, a boy and a girl, who go to the park accompanied by their grandfather and mother respectively. The children meet and start playing with a ball together. Unintentionally, they throw the ball into a tree and then try to get it back in different ways. At the end of the story, a policeman helps the children to get the ball back, and they resume playing. The described situation is familiar to children because the events included represent what can occasionally occur in a park, and the images are sufficiently simple to be immediately clear (see https://doi.org/10.1027/1015-5759/a000569 for several examples of the pictures).

The main characters of the story are the girl and the boy who appear in almost all the pictures of the storybook (girl: seventeen pictures; boy: sixteen pictures), whereas the mom, the grandfather, the turtle and the policeman are secondary characters who appear into four, three, two and three pictures, respectively.

The task elicits topic shifts and topic maintenance at several points within the story. On page 1, the girl and her mom are introduced, and then on page 2, the boy and his grandfather appear. On page 3, the mom and the girl appear again so that a topic shift is elicited.

Then, the story continues with the children present together in most of the scenes so that they represent a referential entity apart (‘the children/they’) that the participants have to maintain across the narrative discourse). The children play with the ball, throwing it at each other (pages 4-5). Then on page 6, a turtle is introduced, as the ball lands on it. The reference to the turtle is also re-introduced on the next page, as the turtle, frightened, retracts into the shell, and the ball ends up on it. Then a topic shift is elicited again when the girl retrieves the ball, and the children go back to play together (pages 8-9).

At this point, the problematic situation of the story occurs: the ball ends up in a tree (page 10), and there follows a series of attempts that the children make – individually and together – to retrieve the ball. The succession of pictures in which the boy tries and fails (pages 11-12), the girl tries and fails (pages 13-14), and then the children try together and fail (page 15) elicits a series of topic shifts in the narrative discourse, where the narrator is required to re-introduce the three entities. Finally, on page 15, the policeman is introduced. On page 16, the children ask him for help, which elicits the maintenance of the reference to the policeman in the object position. On the next page, however, the role of the policeman becomes active since he gets the girl on his shoulders so she can finally reach the ball. The active role played by the policeman in this scene elicits a topic shift in which his character is re-introduced in the subject position within the narrative discourse. Thanks to the policeman's help, on page 18, the children return to play together. In the background, while the children are playing, the grandfather and the mother are shown, still intent on the bench doing their thing, eliciting one last topic shift in the discourse.

During the task administration, the children were asked to look through the illustrations to familiarise themselves with the story. They were then invited to tell the story keeping the book on the table in front of them. The situation is a joint attention situation, where the examiners can see the storybook on the table, and the children can spontaneously use language and gestures in telling the story. The examiners could not interfere with the narration. They only supported the children with positive feedback ("Good!" and "Well done!"). If queried by the children, the examiners had to keep their answers as brief as possible and encouraged them to continue with their stories. The task was audio- and video-recorded and later transcribed into CHAT format according to the CHILDES transcription conventions (MacWhinney, 2000) by a native Italian-speaking trained transcriber.

3.2.3. Coding

Referential cohesion. To examine children's ability to produce a coherent and cohesive discourse adequately, we first focused on referential coherence: how children refer to characters during their narrative. The analyses included the following referents: the boy, the girl, the mom, the grandfather, the turtle, the policeman and the ball. We chose to include also the ball because it is in sixteen pages of the storybook and was referred to by children and adults in subject and object positions within the narrations.

Following Norbury and Bishop's (2003) guidelines, narrative productions are segmented into syntactic units. A syntactic unit was defined as a main clause and all subordinate clauses belonging to this main clause if any. For example, complex sentences with three subordinate clauses (e.g., "Camilla decide allora di provare ad usare la corda per recuperare la palla [Then Camilla decides to try using the rope to retrieve the ball]") were counted as one syntactic unit. Coordinated clauses (e.g., "il triciclo si muove e Fabio cade per terra [The tricycle moves and Fabio falls to the ground]") were counted as two syntactic units. We coded coordinated sentences with a null subject in the second clause as two syntactic units, too (e.g., "I bambini notano la palla e ø decidono di giocarci [the children notice the ball and Ø decide to play with it]").

Direct speech and asides addressed to the examiner were not included in the analyses.

We distinguished between three situations: (a) introduction of reference to a story character for the first time, (b) maintenance of reference to a character that is referred to in the previous syntactic unit, and (c) re-introduction of a referent that is not mentioned in the previous syntactic unit. Less specific forms are generally expected to be used to maintain a reference to a character. These may be null subject and clitic object pronouns if the character is highly prominent. However, when more characters are present in the discourse, speakers tend to use more full NPs to maintain reference than when only one character is present (Arnold & Griffin, 2008). For the referent's re-introduction, generally, full null phrases are used. However, overt subject pronouns, which always signal a topic shift in Italian, can be felicitously used to re-introduce story characters, especially if the active referents differ for gender and/or number.

Each referent was coded for morpho-syntactic form (null subject pronoun, overt subject pronoun, strong object pronoun, clitic object pronoun, noun phrase), and discourse function (referent introduction, re-introduction, and maintenance).

Referents could be expressed through various morpho-syntactic forms. In the example in (1a-1b) the ø symbol stands for a null subject pronoun. Person/number information related to the referent is provided by the person/number inflexion on the verb.

1a *ADU: i bambini la notano e **ø** decidono di giocarci.

%eng: "the children notice it and **ø** decide to play with it".

1b *CHI: il bambino lascia il triciclo e **ø** si mette a giocare a pallavolo.

%eng: "the boy leaves the tricycle and **ø** starts to play volleyball".

In (2a) and (2b), we have two examples of singular and plural third-person overt pronouns: lei, "she", and loro, "they", respectively. In some clauses, demonstrative (e.g., "questa" [this] and numeral (e.g., "i due" [the two]) pronouns are used to refer to story characters in subject position. In order to avoid the over-proliferation of subcases, we coded them as overt pronouns.

2a *ADU: poi comincia a fare un tentativo **lei**.

%eng: "then **she** starts to make an attempt".

2b *CHI: e **loro** hanno detto.

%eng: "and **they** said".

Strong object pronouns are typically rare and used to express contrastive focus in Italian. An example is provided in (3):

3 *ADU: il bambino chiede a **lei** di giocare.[1]

[1] Unlike the other examples provided, Example 3 is made up. In the corpus of narratives produced by our sample, only one strong object pronoun was produced in third dative plural person (i.e., "a loro" [to them]). Since there is no

%eng: “the boy asks her to play”.

Pronominal object clitics are more frequently used, as exemplified in (4a) and (4b).

4a *ADU: i bambini **la** notano e ø decidono di giocar**ci**.

%eng: “the children notice **it** and ø decide to play with **it**”.

4b *CHI: dopo ø cerca di prender**la**.

%eng: “then she tries to catch **it”**.

Full NPs are shown in (5a-5b). In Italian, the definite article must agree in gender and number with the noun.

5a *ADU: la palla rimbalza sul guscio del**la tartaruga.**

%eng: “the ball bounces on the shell of **the turtle**”.

5b *CHI: ad un tratto trovarono **una tartaruga**.

%eng: “suddenly they found **a turtle**”.

Then, all referents were coded for discourse function: referent introduction, re-introduction, and maintenance. Regarding the referent introduction, we included all the referring expressions used on the first mention of the referent. For example, in (5b), the NP una tartaruga, ’a turtle’, was also coded for introduction because this was the first mention of this referent. Subsequent mentions of a referent were divided into two categories: re-introduction and

corresponding clitic for this form in Italian, for clarity, we chose to not report this example in this section, although it has been coded as a strong object pronoun.

maintenance. A referring expression was coded for re-introduction if it expressed a subject or an object argument that was not mentioned in the immediately preceding syntactic unit and/or if it expressed a subject argument whose immediate antecedent was in object position (in accordance with Serratrice, 2007). Consistently to previous studies (Orsolini et al., 1996; Azar & Özyürek, 2016), when there is a shift from plural to singular (e.g., the children to the girl), and vice versa (e.g., the boy to the children), the referring expression was coded as referent re-introduction. In all other cases, a subsequent mention of a referent was coded for maintenance, as shown in examples 6-13.

6 SubjA – SubjB reintroduction

*ADU: **il bambino** cerca di recuperare la palla salendo sul triciclo
mentre **la bambina** lo osserva abbastanza perplessa.

%eng: "**the boy** tries to retrieve the ball by getting on the tricycle
while **the girl** watches him quite puzzled".

7 ObjA-ObjB re-introduction

*CHI: poi il bimbo mette **la bimba** in spalla e ø provarono a prendere **la palla**.

%eng: "then the boy put **the girl** on his shoulders and they tried to catch **the ball**".

ObjA-SubjA re-introduction

8 *CHI: allora il vigile prese sulle spalle **la bimba** e **la bimba** recuperò la
palla che così ø ricominciarono a giocare.

%eng: "then the policeman took the girl on his shoulders and the girl retrieved the
ball so that they began to play again".

9 Plural-Singular re-introduction

*CHI: poi **i bambini** hanno ripreso la palla e poi **la bimba** l' ha tirata talmente forte che è finita su un albero.

%eng: "then **the children** retrieved the ball and then **the girl** pulled it so hard that it ended up in a tree".

10 Singular-Plural re-introduction

*ADU: ma ad un tiro troppo alto di Camilla **Fabio** non riesce a prendere la palla che finisce su un albero.

%eng: "but at a shot too high by Camilla **Fabio** fails to catch the ball that ends up in a tree".

*ADU: dispiaciuti **i bambini** cercano di recuperarla.

%eng: "feeling sorry, **the children** try to recover it".

11` SubjA-SubjA maintenance

*ADU: i **bambini** la notano e **ø** decidono di giocarci.

%eng: "the **children** notice it and **ø** decide to play with it".

12 ObjA-Obj-B maintenance

*CHI: ø prese **la bambina** e ø **la** mise in spalla.

%eng: "he picked the **girl** up and ø put **her** on his shoulder".

13 SubjA-ObjA maitenance

*ADU: si avvicina un **vigile** e i bambini **gli** spiegano il problema.

%eng: a **policeman** approaches and the children explain the problem **to him**.

Regarding the children's narrative productions, information about the picture is crucial for the correct coding of maintenance vs re-introduction discourse function. This is because young children may use pronominal forms to re-introduce the characters. To be sure that the child is maintaining vs re-introducing the reference, we considered the picture referred to.

For instance, in (14), the child is re-introducing the girl with a null subject pronoun. Without seeing the picture, a reader would infer that the character is acting for 'trying to catch the ball with the rope' is the same as the previous clause (i.e., the referent is maintained). Actually, the subject that is trying with the rope is the girl, who is not referred in the previous clause, and here should has been re-introduced with a more explicit form (e.g., the third-person subject pronoun lei, 'she', or the NP la bambina, 'the girl').

14	*CHI:	il bambino cercò di prenderla sopra il triciclo ma ø cadde giù.
	%eng:	"the boy tries to catch it on the tricycle but ø (he) fell down".
	*CHI:	ø cercó prenderla con la corda la palla.
	%eng:	"ø (the girl) tries to catch the ball with rope".

Looking at the pictures, we also coded if the child or the adult referred correctly to the entity within the discourse. On a few occasions, we could find some errors in the referencing production. For instance, in children and adults, we observed some wrong production of the clitic pronouns, as in (15).

15	*ADU:	il bambino **gli** chiese di giocare a palla.
	%eng:	"the boy asks **him** to play with the ball".

Here, the masculine clitic pronoun "gli" is wrongly used to refer to the girl of the story. Indicating the masculine indirect object "gli" for the feminine is very common in Italian informal speech and writing.

Also, in children's productions, we could find errors related to using a plural form to refer to the actions actually performed by one character. For instance, in (16) this child reported that the children tried to recover the ball jumping on the tricycle, but they failed and fell down. The pictures (pages 11-12) of the storybook clearly show that the boy did this attempt, while the girl looks at him, laughing at his failed attempt.

16 *CHI: ø provarono con il triciclo e caddero.
%eng: they tried with the tricycle and ø fell down.

Since the proportion of errors in both groups was low, we chose to exclude them from the analyses.

Language Competence. To assess children's language competence, we measured: (a) verbal productivity; (b) syntactic complexity; (c) lexical diversity. By coding these three linguistic categories, we can provide a broad profile of the language competence of preschool children. Below we discuss how we investigate each category using specific measures:

Verbal productivity. Verbal productivity was measured in two ways: First, we counted the total number of syntactic units per child. Second, we calculated the MLU (mean length of utterance) in words by dividing the total number of words by the number of syntactic units. MLU can also be seen as a measure of grammatical complexity (e.g., Devescovi et al., 2005; Rice, Redmond, & Hoffman, 2006); indeed, more morphosyntactically complex sentences are generally longer than simple ones.

Syntactic complexity. For each child, the total number of subordinate clauses was counted. Both implicit subordinates, which are characterised by the presence of a verb in indefinite mode (infinitive, gerund, past participle), and explicit subordinates, which are characterised by the presence of a verb in finite mode (indicative, subjunctive, conditional, imperative), produced by children, were included. When multiple subordinates were present within the same utterance, all subordinates produced were counted. Coordinate sentences were not included in the calculation of subordinates (e.g., in "the girl jumps with her rope and the boy goes on his tricycle" there are no subordinates present).

Lexical diversity. We calculated for each child the D Index, a measure of lexical diversity that was obtained using the VOCD command of CLAN (Malvern, Richards, Chipere, & Purán, 2004). The measures taken into account by the D index are (1) Type, that is, the number of different words used by the child (e.g., "bambino, bambini" [child, children] are computed as two types); (2) Token, which is the total number of words produced during the story-telling (e.g., "bambino, bambini" [child, children] are computed as two tokens). D index is based on a mathematical model of how the type/token ratio varies with token size; therefore, it is not a function of the number of words in the sample, at least for texts between 100 and 400 tokens (McCarthy & Jarvis, 2007).

3.2.4. Reliability

The number of syntactic units, MLU, and lexical diversity, were counted automatically in CHAT based on the transcripts. For the remaining categories, the transcripts were coded by the first author. To assess the inter-coder reliability, a random 20% of these transcripts (6 stories produced by adults and six stories produced by children) were coded by a second coder. We used intra-class correlation coefficients (ICC) to assess the reliability (95% confidence interval in brackets). Concerning referent function, the ICC was 0.832 (0.521–0.949) for introductions, 0.843 (0.097–0.963) for re-introductions and 0.833 (0.533–0.948) for maintenances.

Concerning the morpho-syntactic form of the referents, the ICC was 0.907 (0.573-0.976) for null subject pronouns, 0.696 (0.024–0.893) for overt subject pronouns, 0.881 (0.652–0.964) for clitic object pronouns, 1.000 for strong object pronouns, and 0.971 (0.761–0.993) for NPs. The average ICCs were 0.891 for the morpho-syntactic form and 0.836 for the function of the referents, respectively. Disagreements were resolved by consensus between the coders.

3.2.5. Data Analysis

The outcome variable was the percentage of referring expressions used (null subject pronoun, overt subject pronoun, strong object pronoun, clitic object pronoun, NP) for the re-introduction, maintenance, and introduction of the referents. We performed analyses separately for the three-discourse function: Introduction, Maintenance and Re-introduction, using IBM SPSS 28. First, we used a non-parametric Mann-Whitney *U* test for differences between children and adults in referential choice at these three moments, controlling for provenience effects. To see whether the speakers' referential choices differ at re-introduction vs maintenance moments, we run a non-parametric Wilcoxon test for repeated measures, separately for adults and children.

Then we used Spearman's rho correlation to test the association between children's referential choice and language competence (verbal productivity, syntactic complexity, lexical semantics). Next, we look at referential choices in which children differ from adults to examine possible cognitive mechanisms underlying the development of referential cohesion. We analyse the impact of TOM, WM, and IC on the referential choice made by children using Generalised Linear Models (GLZM). A log link function was used to accommodate the outcome variables and perform a series of Poisson regressions. Relevant parameters derived from EF and TOM tasks were added one at a time in the model and then included simultaneously in the analysis. The age of children, expressed in months, was included in all models as a covariate. We reported the full model results (i.e., where all predictors are included simultaneously). When divergent from the full model results, we also reported the effects of the single predictors in the model. Interaction effects between TOM and IC were also tested to verify if the mechanisms are involved together or separately in referent (re)introduction and maintenance.

A summary of children's means and standard deviations at the task administered is presented in Table 3.1.

Table 3.1 Mean Scores (Standard Deviations) of age, Raven's matrices, SDQ, cognitive tasks and language competence of children's group.

	M	(SD)	Range
Background variables			
% female	0.54	(0.50)	0 – 1
Age (months)	62.150	(5.542)	50 – 71
Raven's Matrices (percentile rank)	62.974	(25.562)	16 – 97
SDQ (Z score)	-0.209	(0.732)	-1.71 – 1.44
Cognitive Mechanisms			
TOM (Proportion correct response)	0.54	(0.505)	0 – 1
WM span	3.08	(1.34)	1 – 6
IC	0.27	(0.25)	-0.211 – 0.758
Language Competence			
Total number of utterances	25.74	(7.419)	16 - 44
MLU	5.824	(1.056)	3.52 – 7.91
Total number of subordinate clauses	8.360	(3.897)	0 - 16
D Index	39.755	(7.930)	25.78 – 50.90

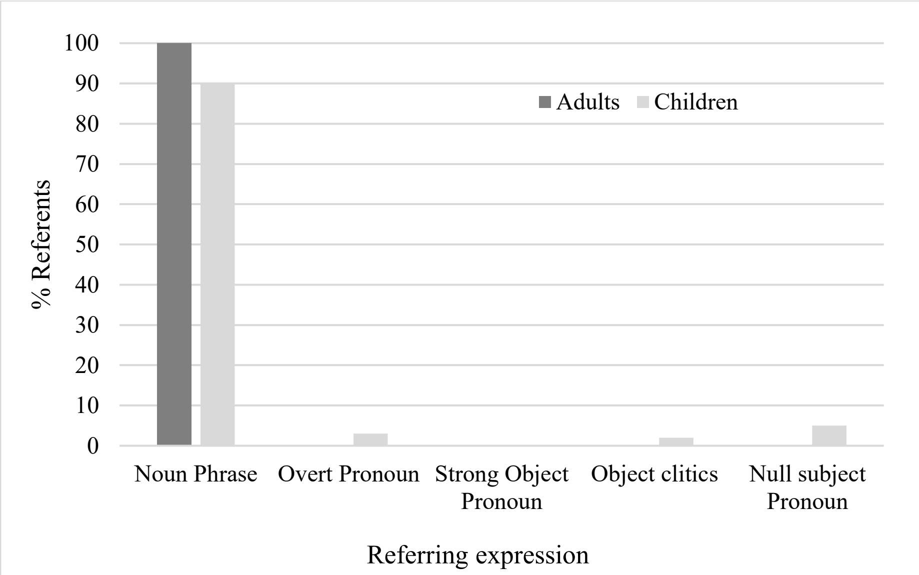
Note. SDQ, Strengths and Difficulties Questionnaires; TOM, Theory of mind; WM, working memory; IC, Interference control skills; MLU, Mean Length Utterance.
N = 39.

3.3. RESULTS

3.3.1. Group effects

Introduction. The percentage of null subject pronouns, overt subject pronouns, strong object pronouns, clitic object pronouns, and NP used in introduction is shown in Figure 3.1.

Figure 3.1 The distribution of referents by morpho-syntactic form in referent introduction



Note. The percentage of each referring expression used to introduce a new character is computed over the total number of introductions produced by participants.

The only type of referring expression used to introduce new characters by adults was NPs (57% indefinite NPs). Children also use NPs (37% indefinite NPs) to introduce new characters 90% of the time. However, a small group of children (n = 11) used pronominal form in referent introduction, as an overt and null subject pronoun to introduce new characters in subject position, and clitic pronoun to introduce them in object position.

As regards overt subject pronouns, five children used demonstrative (e.g., "*questo* va sul triciclo" ["*this* goes on the tricycle"], and personal pronouns (e.g., "*lei* salta la corda" [she jumps the rope"]) to introduce the main characters of the story. One of them disambiguates the referent pointing to it in the picture's book. Four other children used clitic pronouns to refer to secondary characters (policeman n = 3; turtle n = 1) that appeared in the story. Seven children used null subject pronouns to mention characters for the first time: four children to mention secondary characters (i.e., grandfather, n = 1; turtle, n = 3); three to introduce the main characters.

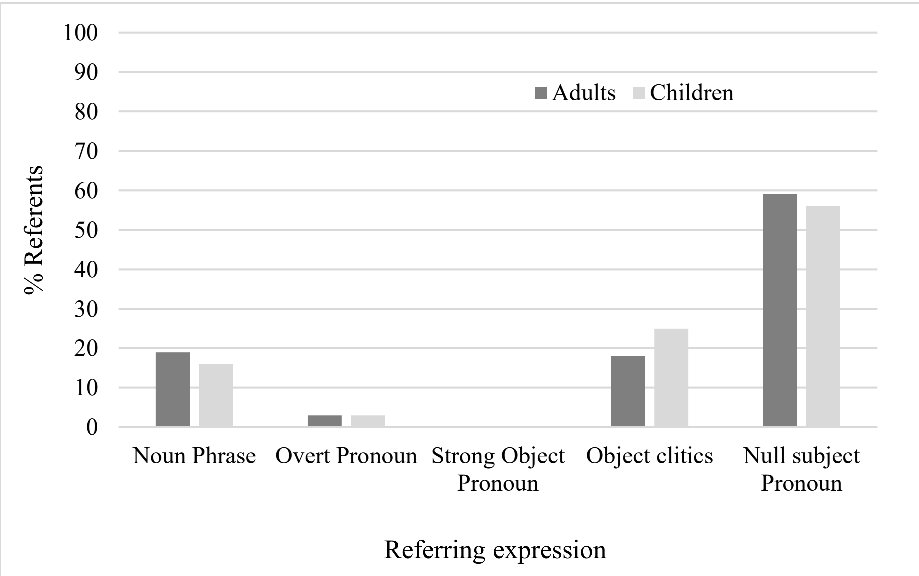
A group (adults, children) by morpho-syntactic form of referring expression (null subject pronoun, overt subject pronoun, strong object pronoun, clitic object pronoun, and NPs) Mann-Whitney *U* test showed statistically significant differences between adults and children only in the frequency of NPs and null subject pronouns in referent introduction, $U = 308.00$, $Z = -2.712$, $p = .007$, $U = 352.00$, $Z = -2.090$, $p = .037$, respectively.

Any pronominal forms (null or clitic pronoun, overt or strong pronoun) are pragmatically infelicitous for introducing new characters. For this reason, we grouped the forms other than NPs used by children into one category: pronominal forms.

Mann-Whitney test showed significant differences between adults and children in the frequency of – more general – pronominal forms in referent introduction, $U = 286.00$, Z = -3.00, $p = .003$. Provenience was not shown to play a significant role in referent introduction with NPs and null pronouns or – more general – pronominal forms for either adults or children (NP: $U = 399.00$, Z = -1.300, $p = .194$, null pronoun: $U = 427.00$, Z = -0.840, $p = .401$, pronominal forms: $U = 403.00$, Z = -1.130, $p = .256$).

Maintenance. The choices for referent maintenance for children and adults are shown in Figure 3.2.

Figure 3.2 The distribution of referents by morpho-syntactic form in referent maintenance



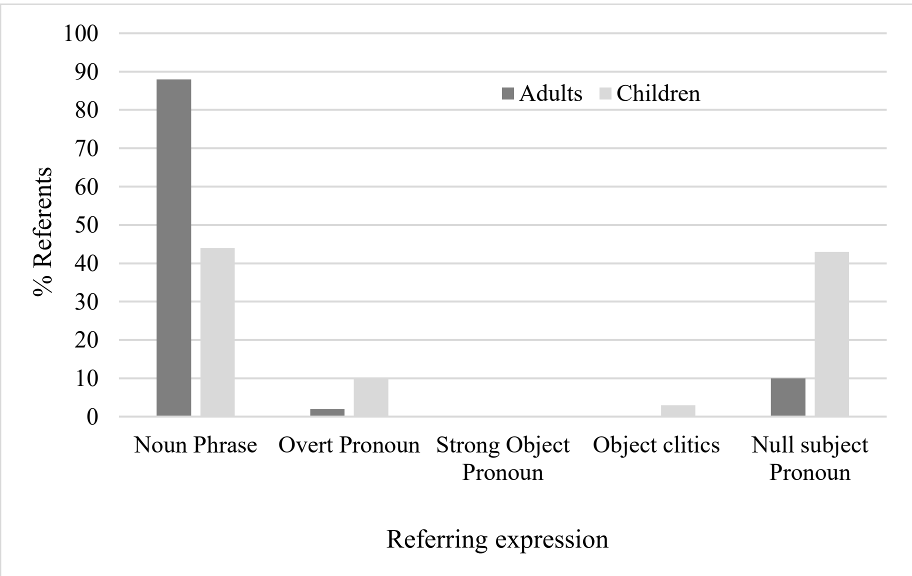
Note. The percentage of each referring expression used to maintain the referent is computed over the total number of maintenances produced by participants.

Children opted for null subject pronouns 56% of the time, followed by clitic pronouns (25%), NPs (16%) and, marginally, overt subject pronouns (3%). The adult group followed the same pattern, but they used proportionally fewer clitic pronouns (18%) to maintain the referents in object position. A group (adults, children) by morpho-syntactic form of referring expression (null subject pronoun, overt subject pronoun, strong object pronoun, clitic object pronoun, and NP) Mann-Whitney *U* test confirms a significant main effect of group in the use of clitic pronouns (U = 290.00, Z = -2.088, p = .037). Children used clitic pronouns significantly more often than adults to maintain the referents in object position, and provenience did not explain this difference (U = 351.00, Z = -1.567, p = .117).

It seems that when children felicitously use a clitic pronoun to maintain the referent, the adults used to be more explicit, preferring a NP. To verify if the pattern observed is statistically significant, we calculate a binary variable: the chance to use NPs over clitics in object position for referent maintenance. We assign the value 1 to speakers who used NPs more often or equally often than clitic pronouns and 0 to those who used clitics more often than NPs. Then, we run a logistic regression with this binary variable as the dependent variable and group as the independent variable, controlling for provenience. Results reveal that the chance of using NPs over clitics did not statistically differ between groups (B = 0.625, SE = 0.526, p = .253).

Re-introduction. In Figure 3.3, we reported the percentages of null subject pronouns, overt subject pronouns, strong object pronouns, clitic object pronouns, and NP used to re-introduce referents.

Figure 3 The distribution of referents by morpho-syntactic form in referent re-introduction



Note. The percentage of each referring expression used to re-introduce the referent is computed over the total number of re-introductions produced by participants.

For adults, NPs accounted for 88% of re-introductions, and the remaining 10% and 2% were null subject pronouns and overt pronouns, respectively. Adults produced null subject pronouns to refer to the "children" entity and only when the active referents in the previous syntactic units were the boy and the girl. In such a situation, the plural marked inflexion of the verb that accompanies the null subject pronoun enables the listener to assign the referent unambiguously. The role of the verb to make the referent inferable in such cases is further illustrated in the example below:

17	*ADU:	**Anna** pensa di poter utilizzare la sua amata corda che utilizza sempre per saltare per recuperare la palla.
	%eng:	"Anna thinks she can use her beloved rope that she uses always to jump to retrieve the ball."
	*ADU:	ma neanche questa sembra una buona idea.
	%eng:	"but it does not look like a great idea either"
	*ADU:	per questo motivo **Tommaso** sorride un po' sotto i baffi.
	%eng:	"for this reason Tommaso smirks off his face"
	*ADU:	allora ø **hanno** un'altra idea.
	%eng:	"then ø [they] **have** another idea".

The verb "hanno" in the concluding sentence of the example has a reference expressed with a null pronoun. Although a null subject pronoun is used to express reference, listeners have no trouble assigning the reference to the most pragmatically plausible characters, Tommaso and Anna (the children). The assignment process occurs with the same speed in these types of items as in the items in which reference is expressed by overt pronominal forms (e.g., "loro" [they]) and full NPs (e.g., "i bambini" [the children]). The crucial information here is the syntactic information marked by the plural inflexion of the predicate in the last clause and the context representation: there are only two active animate referents in the previous clauses. When in the story context the active referents are more than two, as in the following example (18), adults preferred NPs to re-introduce to the children entity:

18 *ADU: la **mamma** accompagna **Tania** al parco dove lei ama giocare con la sua corda.

%eng: "the **mom** takes **Tania** to the park where she loves to play with her rope"

*ADU: lí ogni giorno **ø** incontra **Umberto** il suo amico che viene accompagnato dal **nonno**.

%eng: "every day, ø **[she]** meets her friend **Umberto** who is accompanied by his **grandfather**"

*ADU: **Umberto** ama invece giocare con il suo triciclo.

%eng: "Umberto loves play with his tricycle"

*ADU: d' improvviso **i bambini** si rendono conto della presenza di una palla.

%eng: "all of a sudden **the children** become aware there is a ball on the ground"

To refer to the children entity, adults opted for null subject pronouns 37% of the time and NPs 59% of the time. Overt subject pronouns were used only by three adults: Two used "entrambi" [both] to mark the shift from singular to plural reference (i.e., the children); one used the masculine third-person personal pronoun "he" to re-introduce the reference to the boy in a context where the boy and the girl are the only animate active referents.

The preference for NPs to re-introduce referents is not so pronounced in children. Children opted for NPs and null subject pronouns 44% and 43% of the time, respectively. Then, they opted for overt subject pronouns 10% of the time and for clitic pronouns 3% of the time.

Unlike adults, children used null subject pronouns also to refer to singular entities. For instance, reference to the boy is re-introduced 24% of times with a null pronoun and 54% with a NP.

Reference to the girl was re-introduced 24% times with null pronouns and 62% with a NP. Interestingly, children used null subject pronouns 8.2 times out of 10 to re-introduce the reference to the “children” entity. Conversely, they used NPs to re-introduce the “children” entity in a small percentage (10%) compared to adults (56%). Overt subject pronouns were used more by children ($n = 17$) than adults ($n = 3$).

Children’s overt pronoun in re-introduction mainly refers to the boy and the girl character in a contrastive way (e.g., “*Lui* cade e *lei* ride” [“*he* fell down, and *she* laughs”]. Since the main characters of the story are not of the same gender, and they are only agents in most scenes, the reference expressed by overt pronouns was likely pragmatically predictable from the context and inferable by the listener in most clauses. Two Italian-native speakers, naive to the aim of the study, read the children’s story production without accessing the picture book. They correctly identified 92% and 97% of the overt subject pronouns used by children in these clauses. If taken together, 100% of the overt pronouns produced by children in re-introductions were correctly interpreted by at least one of the two coders. The inter-coders agreement was 89%.

A group (children, adults) by morpho-syntactic form of referring expression (null subject pronoun, overt subject pronoun, strong object pronoun, clitic object pronoun, and NP) Mann-Whitney *U* test yielded significant differences between adults and children in the use of NPs, null pronouns, overt pronouns and clitic pronouns to re-introduce referents (NP: $U = 6.00$, $Z = -6.355$, $p < .001$; null pronoun: $U = 27.00$, $Z = -6.039$, $p < .001$; overt pronoun: $U = 296.50$, $Z = -2.385$, $p = .017$; clitic pronouns: $U = 24.00$, $Z = -6.232$, $p < .001$). None of these differences was explained by the provenience of the sample ($p > .05$).

Looking at Figure 3.2 and Figure 3.3, we can observe that – even though adults re-introduce a character’s reference more often with a NPs than with a null subject pronoun than children do – children used NPs more frequently in re-introduction than maintenance moment and null subject pronouns more frequently to maintain the referent than to re-introduce it. We run a non-parametric Wilcoxon test for repeated measure to compare the use of NPs, clitic object and null subject pronouns by children in referent re-introduction *vs* maintenance. Results confirmed that children change their referential choices between referent re-introduction and maintenance: 66% and 87% of them used, respectively, more null and clitic pronouns to maintain a referent than to re-introduce it (Z = -3.349, $p < .001$; Z = -4.532, $p < .001$, respectively) and 92% used more NPs to re-introduce a referent than to maintain it, Z = -.098, $p < .001$.

It is the same pattern shown by 100% of adults in the use of NPs (maintenance < re-introduction: Z = -4.108 $p < .001$), null and clitic pronouns (maintenance > re-introduction: Z = -4.107, $p < .001$, Z = -3.825, $p < .001$) across referent re-introduction and maintenance.

Table 3.2 Association between children's referential choices and language competence.

		Verbal productivity		**Syntactic Complexity**	**Lexical semantics**
Discourse function	**Referring expression**	**Story length**	**MLU**	**Subordinate clauses**	**D index**
Introduction	Noun Phrase	.214	**.471****	.230	.312
Introduction	Pronominal forms	-.197	**-.464****	-.152	-.242
Maintenance	Noun Phrase	.065	**.532****	.296	.015
Maintenance	Null pronoun	.047	**-.328***	-.001	.081
Maintenance	Overt pronoun	**.413****	.168	.308	.244
Maintenance	Clitic pronoun	-.281	-.207	**.450****	-.069
Re-introduction	Noun phrase	**.355***	**.525*****	.247	**.356***
Re-introduction	Null pronoun	**-.516*****	**-.430****	**-.462****	**-.392***
Re-introduction	Overt pronoun	**.313***	-.199	.294	-.016
Re-introduction	Clitic pronoun	-.290	**-.315***	-.252	-.275

Note. Correlation indices are Spearman's *Rho.* Indices in bold are statistically significant.
N = 39
* = p < .050
** = p < .010
*** = p < .001

3.3.2. The relation between children's reference production and language competence

Next to the differences between groups in reference production, we tested the correlation between reference production and the children's language competence. Table 3.2 shows correlations between referential choices in introducing, maintaining, and re-introducing story characters.

Introduction. The ability to correctly introduce new characters using NPs is only positively associated with children's MLU. Consistently, children using more pronominal forms to introduce new characters produced shorter utterances.

Maintenance. The capacity to maintain a reference to characters within the story in a cohesive way, as seen in Figure 2, is similar between children and adults. Children manage this capacity early, so it is not surprising that language skills still developing in pre-schoolers are not so related to them. We could find that the production of NPs and null pronouns in the maintenance position was positively and negatively related to the MLU but not to the story length. Overt pronoun production correlates with the total number of utterances produced but not with MLU. Clitics production is positively associated with the number of subordinates produced.

Re-introduction. Developmental differences mainly concentrate on the ability to re-introduce characters cohesively. For this reason, we expected that using NPs was positively related to language competence, whereas using null subject pronouns related negatively to children's language skills. Table 3.2 showed that children who use more null pronouns to re-introduce reference produced fewer and shorter utterances in their stories, with fewer subordinates and less lexical variety. On the contrary, overt subject pronouns can be felicitously used in character re-introductions associated only with the story length produced. Using overt subject pronouns to re-introduce referents did not associate with poorer syntactic skills. Children using a higher percentage of NPs in referent re-introduction are those who showed more lexical variety and produced longer stories with longer utterances.

3.3.3. Cognitive mechanisms

Finally, we investigate which mechanisms may underlie the development of referential choice within narrative discourse. Age of children, efficiency of IC, span of WM and first-order TOM acquisition were included as independent variables, both at a time and simultaneously. Both the principal and interaction effects of IC and TOM were tested.

As outcomes, we choose to investigate the referential productions where children, as seen before, differ from adults:

- Introductions made with NPs (children < adults)
- Introductions made with pronominal forms (children > adults)
- Maintenance made with clitics (children > adults)
- Re-introductions made with NPs (children < adults)
- Re-introductions made with null subject pronouns (children > adults)
- Re-introductions made with overt subject pronouns (children > adults)
- Re-introductions made with clitic object pronouns (children > adults)

Introduction. Most children in the sample use NPs to introduce new characters, as in the adult group. However, a small group of children (n = 11) have used pronominal forms as null subject pronouns to introduce characters infelicitously. No significant effects were found for the percentage of introductions made with NPs. Controlling for age effects, children using a higher percentage of NPs in introductions did not differ from others in IC, WM, or TOM, as seen in Table 3.3.

Table 3.3 Estimated effects of cognitive mechanisms on the use of different referring expressions per discourse function in children.

		TOM[a]		**IC[b]**		**WM[c]**	
Discourse function	**Referring expression**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
Re-introduction	**Noun Phrase**	-0.207	0.121	**-0.748*** **	0.162	0.027	0.042
Re-introduction	**Null pronoun**	**0.382*** **	0.104	**0.426** **	0.171	0.031	0.025
Re-introduction	**Overt pronoun**	-0.869	0.511	1.146	0.656	**-0.346* **	0.173
Re-introduction	**Clitic pronoun**	0.142	0.186	.238	0.224	0.040	0.063
Maintenance	**Clitic pronoun**	0.220	0.153	0.429	0.265	-0.012	0.055
Introduction	**Noun phrase**	-0.050	0.098	-0.069	0.073	-0.007	0.198
Introduction	**Pronominal forms**	-0.155	0.638	**1.702** **	0.715	0.059	0.143

Note. TOM = theory of mind task; IC = interference control task; WM = working memory task.
Age expressed in months was included in the model as a covariate (M = 62.15).
[a] Parameters estimated referred to cases when TOM = 0 (meaning children did not pass the tasks), N =39
[b] Interference Score (higher values mean less IC skills), N = 39
[c] Phonological WM span (higher values mean better WM skills), N = 39
* = p < .050
** = p < .010
*** = p < .001

Using pronominal forms, instead, seems related to poor IC skills. The frequency of pronominal forms used in the introduction increases significantly by 5.48% for one point more in the interference score at the IC task. WM and TOM showed no significant effect on using of pronominal form in introducing new characters. Interestingly, if we look at the interaction effect of TOM and IC, we found that the effect of IC statistically decreases when a child did not pass the TOM task (B = 1.274, SE = 0.655, $p = .062$), suggesting that TOM is at least partly involved in the process.

Maintenance. Children showed to use referring expressions in an adult-like way in referent maintanance, except for clitic pronouns. Children produced more clitic pronouns than adults to maintain the reference in object position. No significant effects were found for the percentage of maintanance made with clitics. Controlling for age effects, children using a higher percentage of clitics in introductions did not differ from others in IC, WM, or TOM, as seen in Table 3.3. However, when adding the interaction effects between TOM and IC to the model, we observed that children who failed the TOM task and showed poorer IC skills used significantly 1.61% more clitics in re-introduction (B = 0.915, SE = 0.230, $p < .001$) compared to those who passed the TOM task (B = -0.317, SE = 0.261, $p = .601$). This suggests TOM and IC may be at least partly involved in this referential choice.

Re-introduction. We found a significant effect of IC for re-introducing a referent with a NP. The frequency of NPs produced in re-introduction decreased by 2.91% for each point more at the interference score at the IC task. No significant effects of TOM and WM were found. Results suggest that the correct use of NPs in re-introduction relies mainly on children's IC skills to inhibit the production of pronominal forms. The role of TOM and WM seem marginal. However, children who passed TOM tasks produced M = 44% of re-introductions with NPs versus those who did not pass the task (M = 36%). In the model where all the cognitive mechanisms were added simultaneously, failure at the TOM task did not explain the variance in the percentage of re-introduction made with NPs by children. However, if added alone in the model (or only with age as a covariate), TOM resulted in a significant predictor (B = -0.297, SE = 0.151, $p = .050$): children who failed TOM task produced 0.74% fewer NPs than who passed the task.

In interaction with IC the effect of TOM is marginal. Interaction effect showed the effect of IC is significant for both children who passed (B = -0.598, SE = 0.158, $p < .001$) and failed the TOM task (B = -1.132, SE = 0.336, $p < .001$).

Results on re-introduction with null subject pronouns show partial consistency. Here, both TOM and IC resulted in significant predictors. The frequency of null subject pronouns used to re-introduce characters decreases by 1.53% for each point more in the interference score of the IC task. Moreover, even controlling for IC effects, children who passed the TOM task used significantly fewer null subject pronouns (M = 33%) in referent re-introduction than children who did not (M = 49%), B = 0.382, *SE* = 0.104, $p < .001$. However, the interaction between TOM and IC revealed that when children passed the TOM task, the effect of IC was no more significant (B = 0.148, SE = 0.239, $p = .538$). Children with poor IC who failed the TOM task produced 1.05% more null pronouns than children who passed the TOM task.

Using overt pronoun to re-introduce referents across the story was significantly predicted by WM and age. At increasing age, children tend to use less overt pronouns to re-introduce characters within the discourse (B = -0.101, SE = 0.049, p = .039). Controlling for age effect, children with low WM span are those who produce 0.70% more overt pronoun in referent reintroduction.

Production of overt pronoun was not predicted by TOM and IC, even if interestingly results suggest that children who passed the TOM task but showed poorer IC skills are those who used more overt pronouns in referent re-introduction – even if this pattern is not statistically significant.

Using clitic pronouns in re-introduction has not been predicted by any cognitive mechanisms.

In summary, results showed that the correct use of referring expressions in re-introduction relies mainly on the ability to inhibit the form that is optimal from the speaker's perspective (i.e., a pronoun) and partly also on the ability to consider the listeners' perspective to produce the form that is optimal for them (i.e., NPs or overt subject pronouns). WM span seems less implied in referential choice, with the only exception of overt pronoun's choice for re-introduction purposes.

3.4. DISCUSSION

The first aim of this study was to provide information about the capacity of constructing a referentially cohesive oral narrative production in Italian preschool children, comparing their choice with those made by a sample of young adults of the same provenience.

The experimental situation used in this study differs from that used in previous studies on referential cohesion. Previous studies asked children to tell a story to a hypothetical or naïve listener, usually seated far from the child. Here, children and adults are asked to tell a story while looking at the picture book, having a real listener who shares their attention on the storybook. Such a situation is ecological and reflects the typical situation in which children and adults generally tell a story. We analysed referential expressions used by children and adults to introduce, maintain and re-introduce references to the characters across the story. Developmental differences between children and adults in referential choice were further investigated to address the second aim of our study: to provide insights into linguistic competence and cognitive mechanisms implied in referential cohesion. We discuss the findings at the three moments investigated (introduction, maintenance, re-introduction) in the following three sections:

3.4.1. Introduction

Based on previous studies (Orsolini et al., 1996; Di Domenico et al., 2020), we expected to find a difference in the proportion of NP chosen by children vs adults to introduce new entities.

In general, our findings suggest that as early as 4, children performed quite similarly to adults in referent introduction. All children in the sample used mostly NPs to introduce characters within the story. However, a small group of preschoolers used pronominal forms, such as null (5%) and overt subject (3%) pronouns or clitic object pronouns (2%), to identify new referents. From a pragmatical point of view, introducing a new character within the story with a pronoun is an infelicitous choice.

None of the adult speakers used pronominal forms in the referent introduction. One could hypothesise that the experimental situation where children and listeners shared their joint attention on the picture book may have influenced the production of pronouns in referent introduction by children: listener here could be able to identify the intended referent for the pronoun so that children can rely on the shared context and be justified in using it. However, a previous study (Hendriks et al., 2014) on Dutch-speaking children showed that some children also choose pronouns 6-19% of the time to introduce a character to a *hypothetical* listener and that the frequency of this choice is statistically different from that made by young adults (100% NPs in referent introduction). In line with our results (90% of NP in referent introduction), Dutch-speaking children used an NP, on average, 87.5% of the time to introduce a new character in the discourse. Other studies comparing children's referent introduction in the presence and the absence of mutual knowledge (Warden, 1981; Hickmann, 2003; Kail and Hickmann, 1992; Kail, 1998; Kail and Sanchez y Lopez, 1997) have shown that children acquiring a variety of languages do not reliably use indefinite NPs for referent introduction neither in presence nor absence of mutual knowledge.

Therefore, we think that the infelicitous choice to introduce new characters with pronouns instead of NPs may rely mostly on other individual differences in the children group. For instance, it may be possible that the mutual knowledge situation of our study has encouraged those children who struggle with language to choose pronouns instead of NPs. However, correlation analysis with language competence revealed that the use of NPs was only related to children's grammar skills. Children who, on average, produced longer utterances within their narratives are those who choose NPs over pronouns to signal the newness of the story's characters. Since MLU was the only linguistic feature related to the referential choice made in introductions, it could be possible that the result is an artefact. Children who used NPs (e.g., "the girl"), of course, produced, on average more words within their utterances than those who used a null pronoun (ø) or an overt pronoun ("she"). Indeed, children who chose NPs in referent introduction did not produce longer stories or more subordinates than those who chose pronouns.

Interestingly, our results showed that children who used pronominal forms in referent introductions were those with fewer IC skills. In line with Kuijper et al. (2015), we did not expect that the referential choice made in introductions could have been predicted by any of the cognitive mechanisms investigated, including IC. This is because, to introduce a new character, a speaker can rely solely on the discourse and keep only in mind if a character is new or given within the narrative context. In doing so, neither a high WM span nor the capacity to consider the listener's perspective and inhibit the speaker's optimal choice is required. As found in a previous study (Kuijper et al., 2015), the speaker only needs to be sensitive to the linguistic discourse to perform adequate introductions. However, our results revealed that children who produced a higher percentage of pronominal forms to introduce new characters struggled in suppressing lexical-semantic interference at the Fruit Stroop task.

Interestingly, IC predicts the use of pronouns in referent introduction but not the use of NPs in such a position. In our opinion, this finding may suggest that children's good IC skills are necessary to prevent pronoun use (especially null pronouns) in referent introductions but are not sufficient to produce adequate introductions using NPs.

To our knowledge, only the study of Kuijper et al. (2015) has previously investigated the role of IC in referent introduction, and they found no significant effects of IC in predicting referent introductions. Next, we speculated possible reasons for the divergent effect of IC we found in our study compared to that.

The unexpected role of IC in predicting the use of pronouns in referent introduction might depend on the situation the child is asked to tell the story. In Kuijper et al.'s (2015) study, the children introduced new characters to a listener who could not see the picture and was seated far from them. It is possible that our experimental situation – where a child is asked to tell a story to a listener who shares the attention on the storybook with them – adds an extra and decisive cognitive effort for children with poor IC skills. It is possible that, for such children, mutual knowledge situations make it harder to discard the pronoun. The significant interaction effects we found between IC and TOM corroborate this hypothesis. When the interaction effect was added to the model, we could observe the effect of IC on the choice to use pronominal forms in referent introduction was significant only for children who passed the TOM task and not for those who failed it.

This suggests that poor IC skills mainly impact pronoun choice if the child mentalises the listener's perspective and considers what is already accessible to the listener. For children who acquired TOM, an extra cognitive effort may be necessary to suppress the pronoun for introduction purposes. Instead, if children cannot mentalise the other's perspective, they would not rely on the mutual knowledge shared with the listener, and therefore their poor IC skills would influence less the infelicitous referential choice made, as found by Kuijper et al. (2015).

In addition, another variation in our experimental situation could make it harder for children with poor IC skills to discard the pronoun. In the assessment of narrative skills, we asked the child to look at the picture book alone before starting to tell the story. As specified in the methods section, the children told the story looking at the picture book, so they did not have to rely on their short-term memory, as in Kuijper et al. (2015). However, unlike their study, the children of our sample already knew the story they would tell. Maybe, children's previous knowledge of the story has created an interference at the moment to introduce the characters – the new entities were not so "new" in our children's minds – making it more difficult for those with poor IC skills to suppress the pronoun production.

In addition, there are differences between the task used to measure IC in our study and that used in Kuijper et al. (2015)' study. They used a computerised motor stop-signal task with 70% go and 30% no-go trials, where the stop-signal delay is adjusted to the child's performance by an algorithm to ensure reliable estimates of the stop-signal reaction time, that is, an estimate of the time required for stopping the go response. Instead, we measured IC, another inhibitory function, using a Stroop-like paradigm suitable for preschoolers. We calculated for each child an interference score representing the child's efficiency in suppressing the interference of lexical-semantic internal representation (given by a perceptual conflict as in the original Stroop paradigm) and producing an accurate verbal response, controlled for averaged efficiency in colour naming and semantic access for naming purposes. Maybe the efficiency of this inhibitory process that posits on the verbal system and works to suppress internal representation underlies the infelicitous production of pronouns in referent introduction more than the ability to rapidly suppress a prepotent motor response.

3.4.2. Maintenance

As hypothesized, our results suggest that as early as 4, children showed adult-like adequate referential choice in referent maintenance. Regardless of age, children produced a proportion of null pronouns, NPs, overt subject and strong object pronouns statistically equal to adults. These results are consistent with previous research showing that as young as four, children predominantly use pronominal forms (null forms in null-subject language as Italian) to maintain reference, showing early sensitivity to the information status of the referents in question (Bamberg, 1987; Orsolini et al., 1996; Hickmann and Hendriks, 1999).

The only difference we could find between adults and children was the use of clitic pronouns. Children produced clitic pronouns 25% of the time for referent maintenance, whereas adults used them 18%. In total, 56% of children against 41% of adults in our sample preferred a clitic pronoun over a NP for maintaining the referent in object position.

Previous studies comparing referential choices made by adult speakers vs children showed that adult speakers used NPs more often than children to maintain referents across the story (Arnold & Griffin, 2007, Hendriks et al., 2014). We found this pattern in 59% of adult speakers in our study. However, in contrast to previous studies, we did not find that the chance to use NPs rather than clitic pronouns was statistically higher in adults compared to children.

Previous studies explained that the preference for NPs over pronouns in adult speakers was related to the presence of two or more active referents of the same gender within the story plot (Arnold & Griffin, 2007, Hendriks et al., 2014). This was interpreted as listener-oriented behaviour: young adults are more explicit to make it easier for the listener to identify the intended referent in multi-referent situations.

In the story plot of the narrative task we administered, the main characters are of different gender and the only agents in most scenes. Clitics are marked for person, number and gender. Thus, the use of clitics in referent maintenance where there are only two characters of different gender (i.e., the little boy and the little girl) or where there is only a plural entity (i.e., the children) and a single entity (i.e., the policeman) in the context does not compromise the listener's access to the referent intended by the speaker (see example 19-20, below). Maybe, for this reason, we could not find significant differences in the percentage of NPs used in referent maintenance between children and adults.

19 *CHI: un bimbo passava di lì.

%eng: a little boy was walking there

*CHI: ø vide la bambina.

%eng: he saw the little girl

*CHI: e ø **le** disse.

%eng: and asked **her**

20 *CHI: i bimbi hanno provato a salire uno sopra l'altro.

%eng: "the children tried to climb one on top of the other"

*CHI: ma ø non ci sono riusciti.

%eng: "but they could not reach the ball"

*CHI: il vigile **li** ha guardati

%eng: the policeman looked **at them**.

Since as early as 4, children showed adequate referential choice in referent maintenance, we did not hypothesise possible implied cognitive mechanisms. In line with previous studies, we could not find any significant effect of WM, TOM and IC on the referential choice of TD children in maintenance (Kuijper et al., 2015). However, interestingly, when considering the interaction effects between

TOM and IC, we observed that children who failed the TOM task and showed poorer IC skills were those who used more clitics in re-introduction. In our opinion, this means that even if adequate, clitics production in referent maintenance relies, at least partly, on the inability of children to both consider the listener's perspective and suppress the use of pronouns – that is the optimal form for the speaker. This is in line with previous studies that found that children, compared to adults, behave more egocentric because they preferred pronouns when NPs would have been more suitable considering the listener's point of view.

In addition, we found that children's choice to use a clitic pronoun to maintain topic continuity was associated with the number of subordinate clauses produced, whereas the choice to use NPs in referent maintenance was associated with MLU.

As far as clitic pronoun production is concerned, we interpret this result as evidence that producing clitic pronouns requires sophisticated linguistic competence, especially from a morpho-syntactic point of view. In the Italian pronominal system, clitic pronouns are acquired simultaneously with verb morphology (Berretta, 1986) and emerge later than the other pronominal forms (Antelmi, 1997; Pizzuto & Caselli, 1992). They have a complex internal syntactic structure. They can be analysed as the head of an impoverished determiner phrase and an internal complement of a verb phrase. As the head carries a strong accusative feature, the clitic production requires syntactic checking, triggering a complex movement operation (see Belletti 1999 for a detailed syntactic analysis). The presence of this level of complexity makes the mastery of clitics particularly demanding for children. Thus, it sounds reasonable that children who produced more syntactically complex narratives also produced more clitic pronouns.

As regards the association between NPs production and children's MLU, we thought that the choice of full NPs (always accompanied by articles) – instead of null and clitic pronouns – had increased the average of words produced within utterances, resulting in a significant positive association between them.

3.4.5. Re-introduction

As expected, developmental differences between children and adults mainly concentrate on the ability to re-introduce characters cohesively. Adults and children differ only in the proportion of almost all referential expressions we examined. We could not find differences in strong object pronouns because neither children nor adults produced this pronominal form in their narratives. As mentioned in the Methods section (paragraph Coding), this kind of pronoun is rarely used in Italian. Adults predominantly produced NPs (88%) in referent re-introduction, but in a few cases, they also produced null (10%) and overt (2%) subject pronouns. The production of pronominal forms was limited to contexts when the listener could pragmatically infer these. For instance, they used an overt pronoun to refer to one of two referents of different gender active in the story context or a null form to refer to the protagonists when they are the only two active animate referents in the previous clauses. In the last case, Italian-native speakers can rely on the verb's plural inflexion, leading the listener to access the entity referred to easily.

Contrary to adult speakers, preschool children did not show a marked preference for NPs in re-introduction: they opted for null subject pronouns 43% of the time and for NPs 44% of the time. These results are different from those of Orsolini et al. (1996), who reported that Italian-speaking preschoolers, even if they overused weaker pronouns compared to school-aged children, preferred NPs over null pronouns in referents re-introduction. In our opinion, there are two possible explanations for these divergent results. The first relies on the different definitions of re-introduction we used compared to their study. For Orsolini et al. (1996), the condition to code the discourse function as "re-introduction" was that the reference - previously introduced – "must have been interrupted and not mentioned in the previous clause" (p. 473). Therefore, cases where the reference is previously mentioned in the object position and then mentioned in the subject position were coded by Orsolini as "maintenance" and by us as "re-introduction" (following Serratrice, 2007). In these cases, we observed that preschoolers often infelicitously use null pronouns. Maybe, for this reason, we have observed a higher percentage of null pronouns (43%) in re-introductions than those reported by Orsolini and colleagues (i.e., 10-13%). The second explanation concerns differences in the experimental setting. We used a storytelling task, whereas Orsolini et al. administered a retelling task where children must remind and tell the story to the examiner. In such a context, the children could not rely on shared knowledge with the listener as they might have done in our setting. This may have prompted children in Orsolini's study, sensitive to the listener's perspective, to produce more often NPs than null pronouns compared to the children or our study. There is indeed evidence that preschool children used more null pronouns than NPs because they exploited the visual availability of the referent in the non-verbal context as a property that made the entity accessible enough to prevent them from using NPs (Orsolini et al., 1996).

As the adult group, child participants in our study mainly used null pronouns to refer to the entity "children" within the story, but they did almost twice as often as adults. Moreover, children used null pronouns to re-introduce also the reference to single entities (e.g., the boy, the girl, the ball), while adults never did. Children did that even when null forms did not allow the listener to understand the character they were referring to. This example (21) is taken from a narrative produced by a 61-months child. The participant systematically used null pronouns in referent re-introductions in an infelicitous way.

21 (i) *CHI: poi lui cerca di prenderlo.

%eng: "then he tries to catch it"

(ii) *CHI: ma ø poi cade.

%eng: "but he fell"

(iii) *CHI: poi **ø** prova con la corda.

%eng: "then **ø [: the girl]** tries with the rope"

(iv) *CHI: ma ø non ci riesce.

%eng: "but she failed"

(v) *CHI: poi **ø** provano insieme.

%eng: "then **ø [: they]** try together"

(vi) *CHI: però il vigile li vede.

%eng: “but the policeman sees them”

(vii) *CHI: poi ø lo riesce a prendere.

%eng: “then he can reach it”

(viii) *CHI: poi **ø** giocano assieme.

%eng: “then **ø [: they]** play together”

For instance, after a felicitous re-introduction of the little boy with an overt pronoun (i) and correct maintenance of the reference with a null pronoun (ii), the child used a null pronoun to refer to the girl in clause (iii). Indeed, the little girl is the character that tries to recover the ball with the rope within the story. Following the constraints of Italian grammar (Berretta, 1990), null subject pronouns can be used to refer to a character only if that character is also the subject of the previous clause. In this case, the referent of (iii) differs from the subject of (ii). Thus, the choice made by the narrator is pragmatically infelicitous. A listener who can’t see the book’s picture would infer that it was the boy to use the rope to reach the ball within the story.

Moreover, the infelicitous choice made by the narrator in the girl’s re-introduction made the null pronoun produced in clause (v) ambiguous. If the child had re-introduced the girl with a NP or an overt pronoun, the listener would have inferred that the plural entity the narrator was referring to in (v) was “the children”. This is because the plural inflexion of the verb *provano* (“try”) would have been used when only the boy and the girl were active in the story context. In clause (viii), the use of the null pronoun accompanied by the plural inflexion of the verb *giocano* (“play”) did not lead to retrieving who is playing because there are three active referents in the discourse (a new character – the policeman - was introduced in clause vi). Here, a listener could not determine if the referents playing together were “the children” or “the children and the policeman”.

Therefore, even if the choice of null pronoun in referent re-introduction could sometimes be considered pragmatically adequate, children showed a clear overproduction of null pronouns compared to adults.

Children also used a fewer but relevant percentage of overt subject pronouns (10%) and clitics (3%) to re-introduce given characters in the discourse, compared to adults. In the previous example (21), for instance, the narrator felicitously used *lui* ("he") to refer to the little boy who was not mentioned at all in the previous clause of the child's narrative (i.e., *il pallone finisce sull'albero* [the ball ends up on the tree]). A closer analysis of the occurrence of overt pronouns in re-introductions provide us evidence that the child group used them mainly to refer to protagonists in a contrastive way (e.g., "*Lui* cade e *lei* ride" ["*he* falls down, and *she* laughs"]. As previously explained, since the two protagonists of the story are not of the same gender and they are only agents in most scenes, the reference expressed by overt pronouns was inferable by the listener in most clauses, as ascertained by two native speakers who were naïve to the study.

As far as clitic pronouns are concerned, we found that, unlike adults, all children used this kind of pronoun at least once to re-introduce a referent in object position. This choice could be seen as controversial because there are cases where the use of clitic pronouns did not compromise the access to the referent. For instance, looking at the previous example (21), in clause (vii) the child used the clitic *lo* in the preverbal position to re-introduce the ball, which was previously mentioned in the maintenance position in clause (i), prender*lo*. An Italian-speaking reader would not show doubt in inferring that the clitic refers to *the ball* because the semantics of the verb in the story context led to disambiguating *what* needs to be caught. Furthermore, the explicit parallelism of the Italian verb *prendere* between clauses (i) and (vii) help the listener to infer that the object that needs to be reached/caught is the same (i.e., the ball) across the clauses. As previously found by Orsolini and colleagues (1996), in some cases, using a reduced form is justified instead of a more explicit form because the verb semantics and the structure of the preceding discourse context make the referent easily inferable by a listener. The authors found that school-age children proved to be more successful than preschoolers in judging when a clitic was appropriate in re-introducing a referent *versus* when such reduced forms were not sufficiently informative and a NP had to be used instead.

Children of our sample used only 3% clitic pronouns in re-introduction. Among these, 80% of the time, clitics referred to the ball. In a few cases, they were used to re-introduce references to the children (6%) the girl (10%) or other secondary characters (3%). Regardless of the adequacy of choices made by children, it should be noted that adult speakers systematically preferred to be more explicit (and over informative), even when, as children, they could rely on the preceding discourse or the verb semantics. This evidence is in line with previous studies comparing referencing skills of adults *versus* children (Arnold & Griffin, 2007; Hendriks et al., 2014).

Only two adults used clitic pronouns to re-introduce referents: once to refer to the ball (22) and once to refer to the children (23):

22 *ADU: nella scena c'è una palla dietro un cespuglio.
%eng: "but in the scene there is a ball behind a bush"
*ADU: il bambino la trova
%eng: "the boy finds it"
*ADU: ø scende dal triciclo.
%eng: "ø gets off the tricycle"
*ADU: ø **la** prende.
%eng: "ø takes **it**"

23 *ADU: e [i bambini] non riescono a tirare giù la palla.
%eng: "and ø [: the children] can't retrieve the ball"
*ADU: e intanto adesso anche Francesco ha la sua rivincita.
%eng: "and now even Francesco [: the little boy] has his revenge.
*ADU: e ride mentre lei non riesce a prendere la palla con la corda.
%eng: " and ø [: he] laughs while she can't reach the ball with her rope"
*ADU: e in quel momento passa il ghisa di quartiere che **li** vede.

%eng: “and at that moment, the policeman arrived and saw **them**”

The referent’s antecedent of the clitics used by adult speakers is pragmatically inferable in both examples, but not due to the verb semantics or the structure of the preceding discourse as in children’s narratives. In (22), the clitic pronoun is used to re-introduce the reference to the ball that had been briefly interrupted by an embedded clause. The embedded clause constitutes a syntactic unit apart in accordance with the segmentation system we followed, but the prosodic way it was realised makes the listener perceive that the discourse topic started in the previous clause would continue after that. For this reason, a clitic pronoun is as adequate as a NP. In (23), using a clitic to re-introduce “the children” is justified by the explicit reference to the little girl and the little boy in the previous sentence. There are no other active characters in the discourse, so it is easy for a hypothetical listener to infer that the plural inflexion on the clitic referred to the children.

In our opinion, the differences between adults and children in using clitic and overt pronouns in re-introduction suggest that adults’ approach is totally listener-oriented when they tell a story. We could say that adults used more NPs than children across referential functions, even when a pronoun would have been a more economical but equally informative choice. Adults are susceptible to the possibility that the listener could misunderstand a pronoun, so they more often prefer being explicit. In telling stories, preschoolers behave more egocentrically than adults, but they do not all the time. In line with Orsolini’s (1996) findings, we found that children produced more NPs in referent re-introduction than in maintenance. Therefore, on average, they use more explicit forms when the listener must be considered (i.e., re-introduction moments). Conversely, in referent maintenance, when there is no need to take the listener into account, they produce more null and clitic pronouns than in re-introduction. This is evidence that, as young as 4, children distinguish between these two moments and, on average, they tend to consider the listener. However, in contrast with adult behaviour, children’s approach is not totally listener-oriented. Whenever it is possible, children tend to use a pronoun because this is more convenient for a speaker’s perspective and requires less cognitive effort. Indeed, analysing the mechanisms that implied the choice of null pronouns in re-introduction, we found that children's production of NPs in referent re-introduction relies on their IC skills and, partly, on their TOM. Children who overused null forms in re-introduction are those who failed the first-order TOM task and struggled more in the IC tasks, in line with Kuijper et al., (2015). This confirms the theory that to perform a correct referent re-introduction, a narrator must have the capacity to take into consideration the listeners' access to referents in the story context and, if referents are not easily accessible for them, produce a more explicit referential form (Hendriks et al., 2014). Indeed, for suppressing the production of a null pronoun to re-introduce a character, children need primarily to be able to mentalise and take into account the listener's perspective and, secondarily, need to be good at suppressing internal interferent representation to produce an adequate response. Even if the impact of both mechanisms is significant, results showed that controlling for IC skills, children who failed the TOM task produced much more pronouns than those who passed the task. This suggests both steps are required to avoid the production of null pronouns, but the first – the TOM – is fundamental. Instead, to produce NPs in re-introduction, TOM is no longer sufficient. Children who passed the TOM task produced 12% more NPs than those who failed, but this proportion was not statistically significant. Children with higher IC skills who failed the TOM task showed 22% fewer NPs than those who passed, but the effect is significant in both cases. This suggests that TOM is determinant in suppressing a null pronoun, while good IC is necessary to choose a NP. Results on mechanisms related to the choice of null pronouns differ from those regarding using other pronominal forms in referent re-introduction. Children's use of clitic pronouns in re-introductions was not significantly associated with poor TOM, WM or IC skills. Even if children who failed the TOM task and showed less efficiency at the IC task produced more clitic pronouns than the others, the pattern is not statistically significant. This is probably because the re-introductions they made with clitics mainly referred to the ball and could often be inferred from the discourse features or the semantic verbs. Children who rely on these discourse features and choose clitics in re-introduction might have acquired TOM and be able to suppress the pronoun but decided to use a pronoun because it is more economical from their perspective.

To some extent, we could say the same for the choice of the overt pronoun in re-introductions. Due to the features of the story, using overt pronouns in re-introductions did not compromise 89% of the time the correct access to the referents for the listener. This is interesting because if we look at cognitive mechanisms, we can see that children who produced more overt pronouns in referent re-introduction are younger than the sample mean, passed the TOM task and showed fewer IC and statistically lower WM span. Even if felicitous, the choice to re-introduce characters with overt pronouns in children might rely on the fact that they are generally able to consider the listener's perspective but not so able to suppress the pronoun production or keep the referents (and their name) active in mind or struggle to monitor the discourse. Even if only the effect of WM is significant, we can observe that children who passed the TOM task and those with lower IC skills used 4% and 3.14% more overt pronouns in re-introduction than other children.

Despite being adequate in some cases, using pronominal forms in re-introductions made by children is far from referencing skills shown by adults. Children's use of pronouns for re-introducing given characters related significantly to their linguistic skills. Especially children who used more null pronouns showed a lower score in all the linguistic measures we considered: verbal productivity, syntactic complexity and lexical diversity, while those who used more clitics merely produced shorter utterances. Interestingly, in re-introduction, the use of clitics correlated negatively with the syntactic complexity, while in maintenance, it correlated positively. This might suggest that the syntactic development did not interfere with the pragmatic development in TD preschoolers: children with good syntactic skills who produced more clitics in maintenance did not also produce more clitics in re-introduction.

3.4.5. Conclusions

Our research findings support the evidence that preschool children struggle with appropriate referring expression choices. Children show to distinguish between situations where the reference must be explicit (introduction and re-introduction) and where it could be pronominalized (maintenance). However, compared to adults, they tend to use pronominal forms more frequently, even in situations where the use of a pronoun may result in a non-intended interpretation.

For instance, a consistent but small proportion of children used pronouns to introduce new characters. We found that this infelicitous choice is related to low IC skills. A more consistent proportion of children used null pronouns to re-introduce referent in a proportion almost twice that of adult speakers. We found that children who used more null pronouns in referent re-introduction were those less proficient in verbal productivity, syntactic complexity and lexical diversity. Furthermore, we found that choosing a null form made by these children is related to their TOM, and, to a lesser extent, IC skills. Children who selected adequate NPs in referent re-introduction are those who showed better IC skills, regardless of TOM.

In the situation where the choice of a pronominal form could be adequate (e.g., maintenance with clitics; re-introduction with overt subject and clitic pronouns) but a NP would consent to be sure that the listener could intend the correct antecedent of a referent, children choose pronouns more often than adults. This choice was not related to their linguistic competence, TOM or IC. However, in these children, the choice of re-introducing characters with over pronouns was related to poor WM skills. These appeared to be needed to keep track of the discourse referents and their accessibility. In conclusion, we could say that preschool children showed more egocentric behaviour in choosing the form to express the referents than adult speakers.

Reference is a fundamental function of language, and children with atypical development (e.g., autism spectrum disorders, developmental language disorders) often struggle to choose appropriate referring expressions, making them harder to be intended by their interlocutors. Although concerned with the typical development, our findings may provide helpful insight for clinical speech and language therapists who work with Italian-speaking children that struggle to refer correctly to entities within their discourses:

- Acquiring first-order TOM inference is fundamental before working to reduce the overproduction of null pronouns in re-introductions observed in the preliminary stages of the development of referencing skills in Italian speaking children.
- After TOM has been acquired, training children's IC skills might be necessary to facilitate the proper production of NPs and suppress null forms to introduce and re-introduce characters within the discourse.

Despite these findings, there is a need to replicate these results in a larger sample. To date, only one study (Orsolini et al., 1996) has investigated the referential skills of Italian-speaking preschoolers, and they found higher performance in their sample. Therefore, further studies should replicate our results with other experimental situations. For instance, it could be interesting to examine the children's referencing in situations where the listener cannot access the picture book, in order to verify if the referential cohesion skills observed in our sample are underestimated and which cognitive mechanisms are implied.

General conclusions

23,884. This is the number of studies I have found today (29th January 2023) on Psycinfo with the term "Executive Function" in the abstract.

Using a combination of words like "narrative skill/competence/production", I have found on Psycinfo 5,107 studies on narrative skills.

If I write on the search bar of Psycinfo, "abstract (executive function and narrative)", I can find only 146 papers about narrative and executive skills today.

When approaching human development, researchers often focus on a single aspect of development. I always thought that this monogamy in science between scientists and their object of study did not suit the research in the psychological field. The human mind and its functions are so interconnected and complex that it is strange to focus on only one of its dimensions without looking at the surrounding. If so, the risk is to lose sight of the big picture.

This is the idea that accompanied me through my PhD. I was excited by the challenge of investigating something difficult to define, difficult to measure, and therefore difficult to connect from other dimensions of development. I took up this challenge with the hope of being able to help unify a currently highly fragmented field. Executive Function (EF)…This umbrella term, this nebulous concept we never truly agree on… The uncertainty we have when we administer children an EF task we are not sure will be able to tell us something "valid" about their EF, something that will help us to understand why, when or where these children struggle…

Several studies (e.g., Toplak et al., 2011) confirm that performance tests suffer from ecological validity and do not reflect children's difficulties in various life contexts. Within this issue, the question I do to myself is if we can use storytelling to observe children's EF in an ecological way. This is a big question. There is still a long way to go to answer it, but I hope our results will help my colleagues look at storytelling for all the potential it offers in assessing child functioning: a window to understanding language and cognitive development.

We started way back because, as I mentioned, few studies have analysed the relationship between NC and EF in development. Although we could theoretically hypothesize that cognitive control processes are involved in storytelling, the evidence supporting their empirical association is unclear. The meta-analysis on EF and NC has provided insight that the link between these skills is variable throughout development. When considering the general association across childhood and adolescence, the medium effect size revealed that these skills are not related that much. But this is not always true. There is massive heterogeneity across studies (e.g., sample age and measures used to investigate the relationship between these dimensions). By analysing the sources of heterogeneity that moderates the effect size of the relationship between EF and NC, we could observe a period in child development when these skills are highly associated. We estimated that this period coincides with the preschool and early elementary school years. This period is also a critical time for the development of both skills. The association between these skills reveals that EF and NC could positively influence each other, but they could also hinder each other.

This tell us that children with difficulties in one of the two dimensions might suffer from a deleterious waterfall effect. This finding is significant. One reason behind the number of outputs I obtained when searching papers on EF and NC on Psycinfo is that these competencies predict a lot of important outcomes: social skills and academic success are some (Griffin et al., 2004; Johnston, 2008; Moffit et al., 2011). Both EF and NC proved to have a substantial impact on later development.

In the review of Diamond & Lee (2011), the authors claimed that atypically developing populations are the ones who benefit most from preschool interventions aimed at improving EF. We found the same results in a meta-analysis aimed at determining the effectiveness of cognitive training to enhance EF in preschoolers (Scionti et al., 2020). The evidence that the development of EF and NC is particularly interrelated in these populations suggests to me that an integrated intervention to foster both NC and EF may be helpful for these populations. Research has not yet established the possibility of generalising training effects to skills far associated with those directly trained. Implementing such kinds of studies requires enormous time and energy: setting up at least one experimental and control group to be tested pre- and post-intervention and possibly following children in both groups over time to observe long-term effects. Certainly, these studies would be beneficial for the clinical and scientific community. However, before investing time and resources in this, we need to keep our feet on the ground, we need to understand better the relationship between NC and EF. And taking this small step was the scope of my thesis.

The meta-analysis presented in the first chapter would be a starting point to clarify some aspects of this relationship. In summary, it served to establish four central points aimed at guiding further research:

1. the relationship between FE and NC decreases over time but is strongest in the early stage of development (indicatively, 4-7 years)
2. at this stage, the relationship is stronger in developmentally atypical children (with a diagnosis or developmental risk)
3. at this stage, the relationship of EF is stronger with the macrostructural aspects of NC.
4. at this stage, the various EF domains seem indistinctly related to NC. The magnitude of the association is moderate.

Also, the limitations of the meta-analysis were food for thought for studies presented in Chapters 2 and 3. As interesting as the meta-analysis results, the quantitative synthesis of such diverse studies does not allow us to understand whether the correlation observed is partly determined by other confounding factors or abilities, also related to both dimensions. Moreover, since studies on the topic are few, the meta-analysis allowed us to point out *general* considerations. It cannot answer *specific* questions about the topic, such as "which indices of oral narrative performance associate with which executive domains most strongly?"

Thus, we approached the second study with the purpose of understanding whether there were specific and stable patterns of association with EF among the macrostructural indices of narrative performance. Previous studies on this topic were heterogeneous, and the evidence provided has not been replicated. The scholars who investigated the association between EF and NC have done with different experimental paradigms, mainly administering retelling tasks. Furthermore, the EF tasks they administered did not attempt to isolate the linguistic component involved in task execution. Neither multiple tasks to index one component were administered, thus leaving the reader with much doubt about what variance was shared by the NC and EF measures. In 2013, Cannizzaro and Coelho published a fascinating study showing that *in adults*, NC and EF had some principal components in common. Through factor analyses of EF and NC measures, the authors found two components relevant to both measures. These components were: (a) the fluidity of cognitive processes that have been indexed by efficiency scores at shifting tasks and the episodic structure of the narration; (b) the organization-efficiency of cognitive processes that have been indexed by the inability to suppress interference at the Stroop test and the capacity to inhibit derailments within the narrative discourse. The authors' idea of analysing narrative discourse as an ecological measure of EF is exciting and falls right into place in the age when the scientific community is all aware of the limitations of the EF assessment.

In line with Cannizzaro and Coelho's findings, in Chapter 2, we found some interesting connections between macrostructural NC and EF. It turned out that children with a good working memory span tend to anticipate the events in the plot when telling the story. Although we did not think that the absence of anticipations in children's narrative production is an index of low working memory span, the results open some relevant considerations from an educational point of view. Narratives are used as educational strategies in school. They create a pleasant and creative learning environment and a constructive and enjoyable atmosphere for children (Nanson, 2021). Especially in preschools, there are frequent moments devoted to reading, in which the teacher tells a story to the children. These moments could easily become a shared reading time, where children participate actively. For instance, we would suggest that, in these moments, adults might guide children to reflect on the story's structure and ask the child to anticipate the problematic event in the story or the outcome of the protagonists' attempts. According to our results, this might be an opportunity to prompt working memory and foster the comprehension of the story's episodic structure.

It could also be interesting to foster adherence to the story's content in children's narration. Indeed, the results of our study show that the children who make the most errors in coherence are those who also show fragility in inhibitory skills. Telling a coherent story is fundamental for the narration's quality. The association might suggest that inhibitory processes play a key role in producing coherent stories. Although the direction of the association between FE and NC is unclear to date, and this thesis cannot answer this question, we would suggest that as the children are asked to tell a story, the adult might help the children to focus on the content of the illustrations, thereby prompting their interference control processes.

Similarly, helping the child to pay attention to the story's events when telling stories, preventing them from omitting important story episodes, would also be a way to prompt the inhibitory processes involved in this aspect of narration. The analysis showed that the children with the most difficulty managing the flanker effect (i.e., keeping attention to the little fish in the centre) omitted essential events during the narration and thus provided less information to the listener.

In addition, it might be interesting to foster the child to chunk into a single utterance the contents of two or more scenes observed in the picture book. This ability would seem to call into play planning skills closely related to language.

A limitation of this study is the ratio between the large number of variables investigated and the low sample size involved. Increasing the sample size is certainly desirable, and replications of these results are needed. However, the strength of this study is having considered multiple aspects that may confound the relationship between EF and NC (i.e., basic visual and verbal processes on which the investigated EF processes operate; age; socio-economic status; theory of mind and children's language competence). This makes it possible to establish more precise and defined patterns between EF and NC domains and, thus, explain why the indices of both performances were statistically correlated.

Finally, in Chapter 3, we deepen the relationship between EF and NC, focusing on a particular aspect of narrative competence: referential cohesion. The international literature on this topic is extensive. Cross-linguistically, there is consistent evidence that preschoolers struggle with choosing pragmatically adequate correct referential forms to keep track of characters within discourse. Analysing differences between Italian-speaking preschoolers and adults, we found that children showed adult-like competencies in referent introduction and maintenance. In line with previous findings, children show to distinguish between situations where the reference must be explicit (introduction and re-introduction) and where it could be pronominalised (maintenance), but they struggle to re-introduce characters appropriately after topic shifts. Specifically, they use null pronouns frequently, even when a null pronoun may result in a non-intended interpretation. We found that children who chose null pronouns to re-introduce characters within the discourse were those less proficient in verbal productivity, syntactic complexity and lexical diversity.

But interestingly, this pragmatically infelicitous choice was more related to children's TOM and interference control skills. The finding is consistent with the previous findings on the matter provided by Kuijper et al. (2015).

Furthermore, we thought it provides interesting clinical implications for professionals who work with Italian-speaking children that struggle to refer correctly to entities within their discourses. These implications could be summarised with the recommendation of working on the acquisition of first-order TOM and fostering interference control skills development before working to reduce the overproduction of null pronouns in re-introductions.

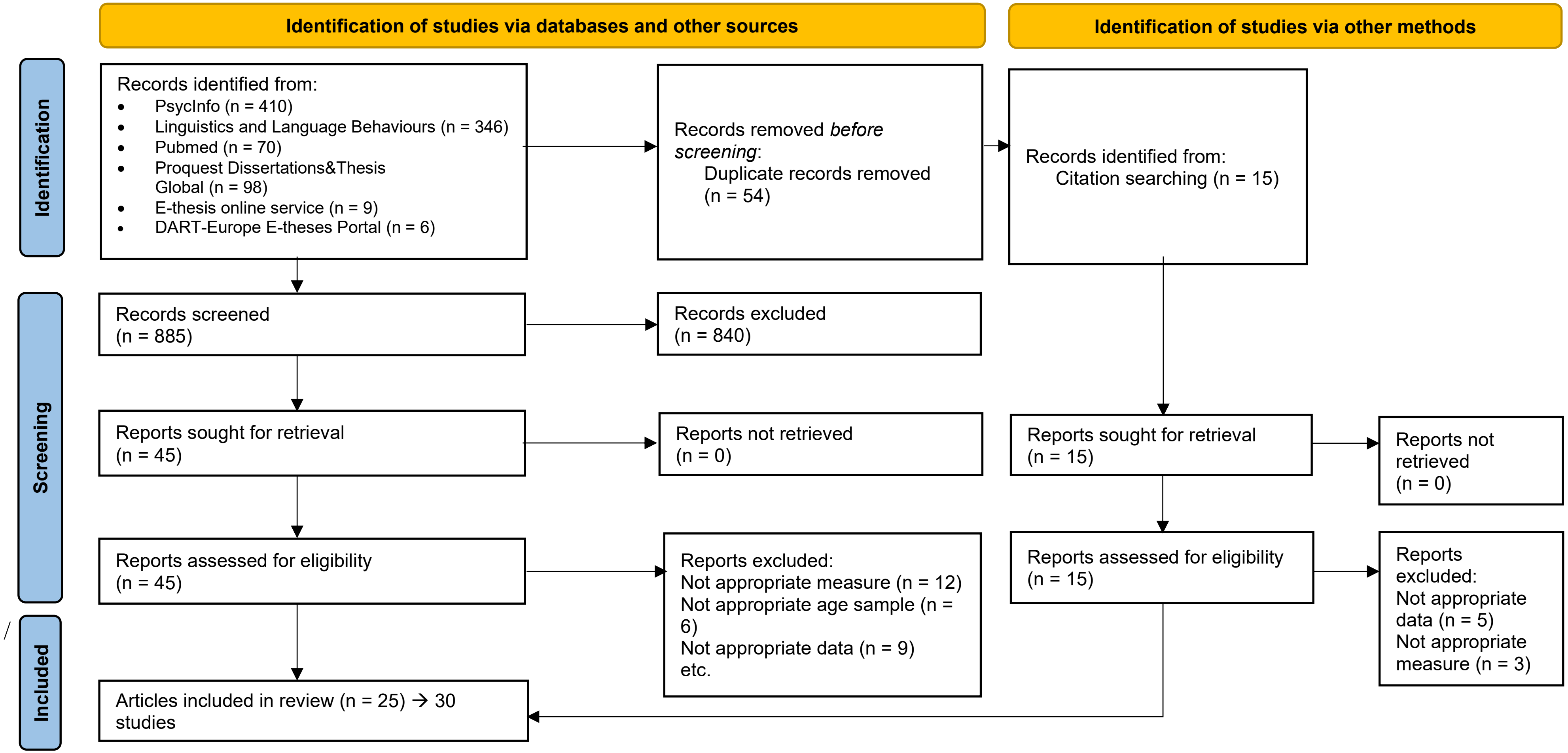
In summary, this thesis wants to offer some insights into the relationship between NC and EF that might be useful for educational and clinical purposes. Nonetheless, further investigations are needed to go deeper into the matter and establish the direction of this relationship across development. However, we hope that this work will help clinicians and educators who use narratives in their practice to know better what they can look at for observing or fostering the EF involved in oral narrative production.

APPENDIX A

1. Narrat* AND Executive Function [OR working memory OR Inhibit* OR flexibility OR shifting OR planning OR problem solving] (filtered by age: > 18 years excluded; by type of document: NOT review)

Storytelling AND Executive Function [OR working memory OR Inhibit* OR flexibility OR shifting OR planning OR problem solving] (filtered by age: > 18 years excluded; by type of document: NOT review)

APPENDIX B - Figure B1. Prisma Diagram



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

APPENDIX C - **Table C1.** Studies including participants aged 4-7 years old

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF domain	EF Task	Narrative Form	Narrative Competence	NC indicator	Fisher_Z, [95% CI]	SE
Balaban et al., 2020	Turkia	Typically developing (n = 18)	4,42	4-5	Behavioral Inhibition	Emotional Stroop Task	Oral	Macro-structural	Story Content - plot complexity	0,2554 [-0,2506, 0,7615]	0,2583
						Emotional Stroop Task		Micro-structural	Morphosintactic Complexity	0,4847 [-0,0214, 0,9908]	0,2583
Dodwell & Bavin, 2008	Australia	Specific Language Impairment (n = 16)	6,70	6-7	Working Memory capacity	Digit Span	Oral	Macro-structural	Information	0,182 [0,3616, 0,7256]	0,2773
					Working Memory capacity	Word Span			Information	0,3205 [0,2231, 0,8641]	0,2773
					Working Memory capacity	Recalling Sentences			Information	0,4059 [0,1377, 0,9495]	0,2773
Duinmeijer et al., 2012	Netherlands	Specific Language Impairment (n = 34)	7,35	6-9	Working Memory capacity	Digit Span	Oral	Micro-structural	Mean Length of Utterance	0,6416 [0,2896, 0,9936]	0,1797
Friend & Phoenix-Bates, 2014	USA	Typically developing (n = 38)	5,00	4-5	Shifting	ANT - executive attention subtest	Oral	-	Story content, lexicon and syntax	0,2693 [-0,062, 0,6006]	0,1691
					Shifting	ANT - executive attention subtest (latency)		-	Story content, lexicon and syntax	0,3062 [-0,0251, 0,6375]	0,1691
					Behavioral Inhibition	Tapping		-	Story content, lexicon and syntax	0,1861 [-0,1452, 0,5174]	0,1691
					Behavioral Inhibition	Tapping (latency)		-	Story content, lexicon and syntax	0,2059 [-0,1254, 0,5372]	0,1691
	USA	Typically developing (n = 42)	4,42	4-5	Behavioral Inhibition	Tapping		-	Story content, lexicon and syntax	0,1748 [-0,1391, 0,4886]	0,1600

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF domain	EF Task	Narrative Form	Narrative Competence	NC indicator	Fisher_Z, [95% CI]	SE
					Behavioral Inhibition	Tapping (latency)		-	Story content, lexicon and syntax	0,3172 [0,0034, 0,6311]	0,1600
					Shifting	ANT - executive attention subtest		-	Story content, lexicon and syntax	0,3406 [0,0267, 0,6544]	0,1600
					Shifting	ANT - executive attention subtest (latency)		-	Story content, lexicon and syntax	0,009 [-0,3048, 0,3228]	0,1600
Ketelaars et al., 2011	Netherlands	Specific Language Impairment (n = 77)	5,60	4-6	-	Nepsy subtests	Oral	Micro-structural	Total Lexical Production	0,3884 [0,1606, 0,6163]	0,1162
	Netherlands	Typically developing (n = 77)	5,60	4-6	-	Nepsy subtests		Micro-structural	Total Lexical Production	0,3095 [0,0817, 0,5374]	0,1162
Khan, 2013 (dissertation)	USA	Typically developing (n = 84)	4,50	3,5-5	Shifting	Verbal Fluency	Oral	Macro-structural	Story Content	0,2132 [-0,0046, 0,4309]	0,1109
					Planning	Tower of Hanoi			Story Content	0,2769 [0,0591, 0,4946]	0,1109
					Shifting	Card Sorting			Story Content	0,3316 [0,1139, 0,5494]	0,1109
Marini et al., 2020	Italy	Developmental Language Disorder (n = 16)	5,17	5	Working Memory capacity	Digit Span	Oral	Macro-structural	Information	0,3294 [-0,2142, 0,873]	0,2773
					Interference Control	Square/Circle		Micro-structural	Number Of Utterance	0,5101 [-0,335, 1,0537]	0,2773
						Square/Circle		Macro-structural	Information	0,6169 [0,0734, 1,1605]	0,2773
McNiven, 2010	Canada	Typically developing (n = 37)	6,95	5-8	Updating of Working Memory	Keep Track	Oral	Macro-structural	Cohesiveness - Referencial accuracy	0,3462 [0,0101, 0,6823]	0,1715
					Updating of Working Memory	N-back			Cohesiveness - Referencial accuracy	0,362 [0,0259, 0,6982]	0,1715
					Updating of Working Memory	Sound monitoring task			Cohesiveness - Referencial accuracy	0,4784 [0,1423, 0,8146]	0,1715

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF domain	EF Task	Narrative Form	Narrative Competence	NC indicator	Fisher_Z, [95% CI]	SE
Sacchetti, 2018 (dissertation)	Italy	Typically developing (n = 38-40)	4,92	3-5	Planning	Non-Narrative Sequences	Oral	Micro-structural	Total Lexical Production	0,4392 [0,1079, 0,7705]	0,1691
						Non-Narrative Sequences		Micro-structural	Lexical Variety	0,1186 [-0,2127, 0,4498]	0,1691
						Non-Narrative Sequences		Micro-structural	Morphosintactic Complexity	0,3417 [0,0104, 0,673]	0,1691
						Non-Narrative Sequences		Micro-structural	Mean Length of Utterance	0,2247 [-0,1066, 0,556]	0,1691
						Non-Narrative Sequences		Macro-structural	Story Content	0,5037 [0,1724, 0,835]	0,1691
						Non-Narrative Sequences		Macro-structural	Coherence of structure	0,5191 [0,1878, 0,8504]	0,1691
					Behavioral Inhibition	Go/NoGo		Micro-structural	Total Lexical Production	0,008 [-,3142, 0,3302]	0,1643
						Go/NoGo		Micro-structural	Lexical Variety	0,006 [-0,3162, 0,3282]	0,1643
						Go/NoGo		Micro-structural	Morphosintactic Complexity	0,1034 [-0,2188, 0,4256]	0,1643
						Go/NoGo		Micro-structural	Mean Length of Utterance	0,1409 [-0,1813, 0,4631]	0,1643
						Go/NoGo		Macro-structural	Story Content	0,1419 [-0,1803, 0,4642]	0,1643
						Go/NoGo		Macro-structural	Coherence of structure	0,044 [-0,2782, 0,3662]	0,1643
					Working Memory capacity	Vocal Span		Micro-structural	Total Lexical Production	0,1522 [-0,1701, 0,4744]	0,1643
						Vocal Span		Micro-structural	Lexical Variety	0,1624 [-0,1598, 0,4846]	0,1643
						Vocal Span		Micro-structural	Morphosintactic Complexity	0,051 [-0,2712, 0,3733]	0,1643
						Vocal Span		Micro-structural	Mean Length of Utterance	0,043 [-0,2792, 0,3652]	0,1643
						Vocal Span		Macro-structural	Information and Story Content	0,0832 [-0,239, 0,4054]	0,1643
						Vocal Span		Macro-structural	Coherence of structure	0,0852 [-0,237, 0,4074]	0,1643

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF domain	EF Task	Narrative Form	Narrative Competence	NC indicator	Fisher_Z, [95% CI]	SE
Tonér & Nilsson Gerholm, 2021	Sweden	Typically developing (n = 47)	5,30	4-6	Interference Control	Flanker	Oral	Micro-structural	Total Lexical Production	0,1409 [-0,1546, 0,4364]	0,1507
					Behavioral Inhibition	Head-Toes-Knees-Shoulders			Total Lexical Production	0,0701 [-0,2254, 0,3656]	0,1507
					Working Memory capacity	Digit Span			Total Lexical Production	0,01 [-0,2855, 0,3055]	0,1507
					Shifting	Dimensional Change Card Sorting			Total Lexical Production	0,01 [-0,2855, 0,3055]	0,1507
					Interference Control	Flanker		Micro-structural	Lexical Variety	0,3654 [0,0700, 0,6609]	0,1507
					Behavioral Inhibition	Head-Toes-Knees-Shoulders			Lexical Variety	0,2132 [-0,0823, 0,5086]	0,1507
					Working Memory capacity	Digit Span			Lexical Variety	0,2554 [-0,041, 0,5509]	0,1507
					Shifting	Dimensional Change Card Sorting			Lexical Variety	0,4847 [0,1892, 0,7802]	0,1507
					Interference Control	Flanker		Micro-structural	Morphosintactic Accuracy	0,4356 [0,1401, 0,7311]	0,1507
					Behavioral Inhibition	Head-Toes-Knees-Shoulders			Morphosintactic Accuracy	0,1206 [-0,1749, 0,4161]	0,1507
					Working Memory capacity	Digit Span			Morphosintactic Accuracy	0,2877 [0,0078, 0,5832]	0,1507
					Shifting	Dimensional Change Card Sorting			Morphosintactic Accuracy	0,2554 [-0,0401, 0,5509]	0,1507
					Interference Control	Flanker			Morphosintactic Complexity	0,1614 [-0,1341, 0,4569]	0,1507
					Behavioral Inhibition	Head-Toes-Knees-Shoulders			Morphosintactic Complexity	0,05 [-0,2454, 0,3455]	0,1507

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF domain	EF Task	Narrative Form	Narrative Competence	NC indicator	Fisher_Z, [95% CI]	SE
					Working Memory capacity	Digit Span			Morphosintactic Complexity	0,2448 [-0,0507, 0,5402]	0,1507
					Shifting	Dimensional Change Card Sorting			Morphosintactic Complexity	0,3428 [0,0474, 0,6383]	0,1507
					Interference Control	Flanker			Morphosintactic Complexity - Unified predicates	0,1717 [-0,1238, 0,4671]	0,1507
					Behavioral Inhibition	Head-Toes-Knees-Shoulders			Morphosintactic Complexity - Unified predicates	0,03 [-0,2655, 0,3255]	0,1507
					Working Memory capacity	Digit Span			Morphosintactic Complexity - Unified predicates	0,1206 [-0,1749, 0,4161]	0,1507
					Shifting	Dimensional Change Card Sorting			Morphosintactic Complexity - Unified predicates	0,3316 [0,0362, 0,6271]	0,1507
					Interference Control	Flanker		Macro-structural	Information	0,2877 [-0,0078, 0,5832]	0,1507
					Behavioral Inhibition	Head-Toes-Knees-Shoulders			Information	0,1104 [-0,185, 0,4059]	0,1507
					Working Memory capacity	Digit Span			Information	0,3095 [0,0140, 0,6050]	0,1507
					Shifting	Dimensional Change Card Sorting			Information	0,4722 [0,1768, 0,7677]	0,1507
Veraksa et al., 2020	Russia	Typically developing (n=269)	5,58	5-6	Working Memory capacity	Memory Design	Oral	Micro-structural	Morphosintactic Accuracy	0,1206 [0,0004, 0,2408]	0,0616
						Memory Design		Micro-structural	Number Of Syntagmas	0,1511 [0,0310, 0,2713]	0,0616
						Memory Design		Micro-structural	Number Of Simple Utterance	0,1511 [0,0310, 0,2713]	0,0616

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF domain	EF Task	Narrative Form	Narrative Competence	NC indicator	Fisher_Z, [95% CI]	SE
						Memory Design		Macro-structural	Coherence - Semantic adequacy	0,1614 [0,0412, 0,2816]	0,0616
						Memory Design		Micro-structural	Lexical Production	0,1614 [0,412, 0,2816]	0,0616
						Memory Design		Macro-structural	Coherence - programming	0,182 [0,0618, 0,3022]	0,0616
					Working Memory capacity	Sentence Repetition		Micro-structural	Number Of Simple Utterance	0,2027 [0,0826, 0,3229]	0,0616
						Sentence Repetition			Number Of Syntagmas	0,2237 [0,1035, 0,3438]	0,0616
					Working Memory capacity	Memory Design		Macro-structural	Coherence - Semantic completeness	0,2342 [0,114, 0,3544]	0,0616
						Memory Design			Coherence of structure	0,2554 [0,1352, 0,3756]	0,0616
					Working Memory capacity	Sentence Repetition		Micro-structural	Total Lexical Production	0,2554 [0,1352, 0,3756]	0,0616
					Working Memory capacity	Memory Design		Macro-structural	Coherence - narrative structure	0,2661 [0,1459, 0,3863]	0,0616
					Working Memory capacity	Sentence Repetition		Micro-structural	Morphosintactic Accuracy	0,3205 [0,2004, 0,4407]	0,0616
						Sentence Repetition		Macro-structural	Coherence - Semantic adequacy	0,4356 [0,3154, 0,5558]	0,0616
						Sentence Repetition			Coherence - narrative structure	0,4599 [0,3397, 0,5801]	0,0616
						Sentence Repetition			Coherence - programming	0,4847 [0,3645, 0,6049]	0,0616
						Sentence Repetition			Coherence - narrative type (complete, simplified, distorted)	0,5361 [0,4159, 0,6562]	0,0616
						Sentence Repetition			Coherence - Semantic completeness	0,5493 [0,4291, 0,6695]	0,0616

APPENDIX C - Table C2. Studies including participants aged 8-18 year old

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
Artico & Penge, 2016	Italy	Dyslexia and Dysgraphia (n = 54)	9,87	8-12	Shifting	Verbal Fluency	Written	Micro-structural	Lexical Variety	0,1003 [0,1741, 0,3748]	0,1400
						Verbal Fluency		Macro-structural	Cohesiveness	0,1003 [-0,1741, 0,3748]	0,1400
					Planning	Tower of London		Micro-structural	Morphosintactic Complexity	0,1307 [-0,1437, 0,4052]	0,1400
					Shifting	Response set (NEPSY II)		Micro-structural	Total Lexical Production	0,1409 [-0,1335, 0,4154]	0,1400
					Planning	Tower of London			Total Lexical Production	0,1409 [-0,1335, 0,4154]	0,1400
					Shifting	Verbal Fluency			Total Lexical Production	0,1717 [-0,1028, 0,4461]	0,1400
					Planning	Tower of London		Micro-structural	Lexical Variety	0,1820 [-0,0925, 0,4564]	0,1400
					Shifting	Response set (NEPSY II)		Macro-structural	Coherence	0,1820 [-0,0925, 0,4564]	0,1400
					Shifting	Switching NEPSY II		Macro-structural	Cohesiveness	0,1923 [-0,0821, 0,4668]	0,1400
					Planning	Tower of London			Cohesiveness	0,1923 [-0,0821, 0,4668]	0,1400
					Shifting	Switching NEPSY II		Micro-structural	Total Lexical Production	0,2027 [-0,0717, 0,4772]	0,1400
					Shifting	Response set (NEPSY II)		Micro-structural	LexicalVariety	0,2132 [-0,0613, 0,4876]	0,1400
					Shifting	Verbal Fluency		Macro-structural	Coherence	0,2132 [-0,0613, 0,4876]	0,1400
					Planning	Tower of London			Coherence	0,2132 [-0,0613, 0,4876]	0,1400
					Planning	Clocks		Macro-structural	Cohesiveness	0,2342 [0,4030, 0,5086]	0,1400
					Shifting	Switching NEPSY II		Macro-structural	Coherence	0,2342 [0,4030, 0,5086]	0,1400
					Shifting	Response set (NEPSY II)		Macro-structural	Cohesiveness	0,2448 [-0,0297, 0,5192]	0,1400
					Behavioural Inhibition	Go/NoGo		Macro-structural	Coherence	0,2448 [-0,0297, 0,5192]	0,1400
					Planning	Clocks		Micro-structural	Total Lexical Production	0,2877 [0,0132, 0,5621]	0,1400

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
					Shifting	Switching NEPSY II		Micro-structural	Lexical Variety	0,2986 [0,0241, 0,5730]	0,1400
					Shifting	Response set (NEPSY II)		Micro-structural	Morphosintactic Complexity	0,2986 [0,0241, 0,5730]	0,1400
					Behavioural Inhibition	Go/NoGo		Micro-structural	Total Lexical Production	0,3428 [0,0684, 0,6173]	0,1400
					Shifting	Verbal Fluency		Micro-structural	Morphosintactic Complexity	0,3541 [0,0796, 0,6285]	0,1400
					Behavioural Inhibition	Go/NoGo		Macro-structural	Cohesiveness	0,3541 [0,0796, 0,6285]	0,1400
						Go/NoGo		Micro-structural	Lexical Variety	0,3654 [0,0910, 0,6399]	0,1400
					Planning	Clocks		Micro-structural	Morphosintactic Complexity	0,3884 [0,1140, 0,6629]	0,1400
						Clocks		Macro-structural	Coherence	0,4001 [0,1256, 0,6745]	0,1400
						Clocks		Micro-structural	Lexical Variety	0,4236 [0,1492, 0,6981]	0,1400
					Shifting	Switching NEPSY II		Micro-structural	Morphosintactic Complexity	0,4599 [0,1854, 0,7343]	0,1400
					Behavioural Inhibition	Go/NoGo			Morphosintactic Complexity	0,5230 [0,2485, 0,7974]	0,1400
Balaban et al., 2020	Turkia	Typically Developing (n = 87)	8,17	7-11	Behavioural Inhibition	Emotional Stroop Task	Oral	Micro-structural	Syntactic Complexity	0,1717 [-0,0422, 0,3855]	0,1091
						Emotional Stroop Task		Macro-structural	Plot Complexity	0,3316 [0,1178, 0,5455]	0,1091
Balioussis et al., 2012	Canada	Typically Developing (n = 70)	9,83	8-9	Working Memory capacity	Letter Memory Task	Written	Micro-structural	Morphosintactic Complexity	0,3541 [0,1146, 0,5935]	0,1221
					Shifting	Contingency Naming Task		Micro-structural	Total Lexical Production	0,4599 [0,2204, 0,6993]	0,1221
					Working Memory capacity	Letter Memory Task			Total Lexical Production	0,3316 [0,0922, 0,5711]	0,1221
					Shifting	Contingency Naming Task		Micro-structural	Syntactic Complexity	0,3428 [0,1034, 0,5823]	0,1221
Drijbooms et al., 2017	Netherlands	Typically Developing (n = 93)	11,08	-	-	Trail Making Test; Tower of London	Written	Micro-structural	Total Lexical Production	0,03 [-0,1766, 0,2366]	0,1054
					-						

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
					-	Trail Making Test; Tower of London		Macro-structural	Story content	0,03 [-0,1766, 0,2366]	0,1054
					-	Digit Span; Letter Fluency; Ricerca visiva			Story content	0,0601 [-0,1465, 0,2667]	0,1054
					-	Digit Span; Letter Fluency; Ricerca visiva		Micro-structural	Morphosintactic Complexity	0,0701 [-1365, 0,2767]	0,1054
					-	Digit Span; Letter Fluency; Ricerca visiva		Micro-structural	Total Lexical Production	0,0701 [-0,1365, 0,2767]	0,1054
					-	Walk Don't Walk; Opposite Worlds; Trail Making Test; Letter Digit Substitution			Total Lexical Production	0,1717 [-0,0349, 0,3783]	0,1054
					-	Walk Don't Walk; Opposite Worlds; Trail Making Test; Letter Digit Substitution		Macro-structural	Story content	0,2027 [-0,0039, 0,4093]	0,1054
					-	Trail Making Test; Tower of London		Micro-structural	Morphosintactic Complexity	0,2237 [0,0171, 0,4303]	0,1054
					-	Walk Don't Walk; Opposite Worlds; Trail Making Test; Letter Digit Substitution			Morphosintactic Complexity	0,2554 [0,0488, 0,462]	0,1054
Drijbooms et al., 2015	Netherlands	Typically Developing (n = 102)	9,58	8-11	Planning	Tower of London	Written	Micro-structural	Morphosintactic Complexity	0,05 [-0,1469, 0,247]	0,1005
					Shifting	Trail Making Test		Micro-structural	Total Lexical Production	0,0701 [-0,1269, 0,2671]	0,1005
					Planning	Tower of London			Total Lexical Production	0,0701 [-0,1269, 0,2671]	0,1005
					Behavioural Inhibition	Opposite words		Macro-structural	Story content	0,1003 [-0,0966, 0,2973]	0,1005
					Shifting	Trail Making Test		Micro-structural	Morphosintactic Complexity	0,1104 [-0,0865, 0,3074]	0,1005

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
					Working Memory capacity	Digit Span		Macro-structural	StoryContent	0,1409 [-0,0561, 0,3379]	0,1005
						Digit Span		Micro-structural	Total Lexical Production	0,1511 [-0,0458, 0,3481]	0,1005
					Planning	Tower of London		Macro-structural	Story content	0,1511 [-0,0458, 0,3481]	0,1005
					Behavioural Inhibition	Walk don't Walk			Story content	0,1717 [-0,0253, 0,3687]	0,1005
					Shifting	Trail Making Test			Story content	0,1717 [-0,0253, 0,3687]	0,1005
					Behavioural Inhibition	Walk don't Walk		Micro-structural	Morphosintactic Complexity	0,182 [-0,015, 0,379]	0,1005
					Behavioural Inhibition	Opposite words			Morphosintactic Complexity	0,2132 [0,0162, 0,4102]	0,1005
					Working Memory capacity	Digit Span			Morphosintactic Complexity	0,2237 [0,0267, 0,4206]	0,1005
					Behavioural Inhibition	Opposite words		Micro-structural	Total Lexical Production	0,2448 [0,0478, 0,4418]	0,1005
					Behavioural Inhibition	Walk don't Walk			Total Lexical Production	0,2554 [0,0584, 0,4524]	0,1005
Fisher et al., 2019	USA	Dyslexia (n = 92)	9,25	-	Shifting	Card Sorting	Oral	Macro-structural	Coherence	0,1206 [-0,0872, 0,3283]	0,1058
					Interference Control	Stroop			Coherence	0,1614 [-0,0464, 0,3691]	0,1058
					Shifting	Trail Making Test			Coherence	0,1923 [-0,0154, 0,4001]	0,1058
					Working Memory capacity	Corsi			Coherence	0,2877 [0,0799, 0,4954]	0,1058
Park, 2014 (dissertation)	USA	Typically Developing (n = 10)	10,00	9-11	Shifting	Trail Making Test	Oral	Macro-structural	GAO units	0,4611 [-0,2797, 1,2019]	0,3780
						Trail Making Test		Macro-structural	Complete GAO units (Integrity)	0,1318 [-0,609, 0,8726]	0,3780
					Planning	Tower of London			Complete GAO units (Integrity)	0,0993 [-0,6415, 0,8401]	0,3780
						Tower of London		Macro-structural	GAO units - episodic structure	0,038 [-0,7028, 0,7788]	0,3780
					Shifting	Card Sorting		Macro-structural	Complete GAO units (Integrity)	0,2079 [-0,5329, 0,9487]	0,3780

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
					Working Memory capacity	Digit Span Backword			Complete GAO units (Integrity)	0,2586 [-0,4822, 0,9994]	0,3780
						Digit Span Backword		Macro-structural	GAO units	0,5682 [-0,1726, 1,3089]	0,3780
					Shifting	Card Sorting			GAO units	0,8053 [0,0645, 1,5461]	0,3780
		Deaf or hard to hearing (n = 11)	10,00	9-11	Planning	Tower of London	Oral	Macro-structural	GAO units	0,5874 [-0,1056, 1,2803]	0,3536
					Working Memory capacity	Digit Span Backword			GAO units	0,3451 [-0,3479, 1,038]	0,3536
					Shifting	Card Sorting			GAO units	0,2384 [-0,4545, 0,9314]	0,3536
					Working Memory capacity	Digit Span Backword		Macro-structural	Complete GAO units (Integrity)	0,1145 [-0,5785, 0,8074]	0,3536
					Planning	Tower of London			Complete GAO units (Integrity)	0,1155 [-0,5774, 0,8085]	0,3536
					Shifting	Trail Making Test			Complete GAO units (Integrity)	0,1348 [-0,5581, 0,8278]	0,3536
						Trail Making Test		Macro-structural	GAO units	0,231 [-0,4619, 0,924]	0,3536
					Shifting	Card Sorting		Macro-structural	Complete GAO units (Integrity)	0,4047 [-0,2882, 1,0977]	0,3536
Peristeri et al., 2020	Greece	Autism Spectrum Disorder (n = 20)	9,80	7-12	Updating of Working Memory	2-back	Oral	Micro-structural	Lexical Variety	0,1246 [-0,3507, 0,6]	0,2425
						2-back		Micro-structural	Morphosintactic Complexity	0,1522 [-0,3232, 0,6275]	0,2425
						2-back		Micro-structural	Number of subordinated clauses	0,2501 [-0,2253, 0,7254]	0,2425
						2-back		Micro-structural	Number of relative clauses	0,046 [-0,4293, 0,5214]	0,2425
						2-back		Macro-structural	Story Structure	0,146 [-0,3293, 0,6214]	0,2425
						2-back		Macro-structural	Referencial Accuracy	0,4153 [-0,06, 0,8907]	0,2425
					Interference Control	Local-to-Global (Accuracy)		Micro-structural	Lexical Variety	0,0993 [-0,376, 0,5747]	0,2425

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
						Local-to-Global (Accuracy)		Micro-structural	Morphosintactic Complexity	0,3272 [0,1482, 0,8026]	0,2425
						Local-to-Global (Accuracy)		Micro-structural	Number of subordinated clauses	0,031 [-0,4444, 0,5064]	0,2425
						Local-to-Global (Accuracy)		Micro-structural	Number of relative clauses	0,047 [-0,4283, 0,5224]	0,2425
						Local-to-Global (Accuracy)		Macro-structural	Story Structure	0,0591 [-0,4163, 0,5344]	0,2425
						Local-to-Global (Accuracy)		Macro-structural	Referencial Accuracy	0,353 [-0,1224, 0,8283]	0,2425
					Interference Control	Global-to-Local (Accuracy)		Micro-structural	Lexical Variety	0,1206 [-0,3548, 0,5959]	0,2425
						Global-to-Local (Accuracy)		Micro-structural	Morphosintactic Complexity	0,0621 [-0,4133, 0,5374]	0,2425
						Global-to-Local (Accuracy)		Micro-structural	Number of subordinated	0,0902 [-0,3851, 0,5656]	0,2425
						Global-to-Local (Accuracy)		Micro-structural	Number of relatives	0,019 [-0,4564, 0,4944]	0,2425
						Global-to-Local (Accuracy)		Macro-structural	Story Structure	0,4822 [0,0068, 0,9576]	0,2425
						Global-to-Local (Accuracy)		Macro-structural	Referencial Accuracy	0,0661 [-0,4093, 0,5415]	0,2425
					Interference Control	Local-to-Global (Reaction Time)		Micro-structural	Lexical Variety	0,4562 [-0,0191, 0,9316]	0,2425
						Local-to-Global (Reaction Time)		Micro-structural	Morphosintactic Complexity	0,3598 [-0,1156, 0,8351]	0,2425
						Local-to-Global (Reaction Time)		Micro-structural	Number of subordinated clauses	0,2942 [-0,1812, 0,7696]	0,2425
						Local-to-Global (Reaction Time)		Micro-structural	Number of relative clauses	0,3372 [-0,1381, 0,8126]	0,2425
						Local-to-Global (Reaction Time)		Macro-structural	Story Structure	0,037 [-0,4383, 0,5124]	0,2425
						Local-to-Global (Reaction Time)		Macro-structural	Referencial Accuracy	0,049 [-0,4263, 0,5244]	0,2425
					Interference Control	Global-to-Local (Reaction Time)		Micro-structural	Lexical Variety	0,4648 [-0,0105, 0,9402]	0,2425
						Global-to-Local (Reaction Time)		Micro-structural	Morphosintactic Complexity	0,2715 [-0,2039, 0,7468]	0,2425

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
						Global-to-Local (Reaction Time)		Micro-structural	Number of subordinated clauses	0,1013 [-0,374, 0,5767]	0,2425
						Global-to-Local (Reaction Time)		Micro-structural	Number of relative clauses	0,045 [-0,4303, 0,5204]	0,2425
						Global-to-Local (Reaction Time)		Macro-structural	Story Structure	0,482 [0,0068, 0,9576]	0,2425
						Global-to-Local (Reaction Time)		Macro-structural	Referencial Accuracy	0,0661 [-0,4093, 0,5415]	0,2425
Peristeri et al., 2020	Greece	Typically Developing (n = 20)	9,80	7-12	Updating of Working Memory	2-back	Oral	Micro-structural	Lexical Variety	0,1257 [-0,3497, 0,601]	0,2425
						2-back		Micro-structural	Morphosintactic Complexity	0,0862 [-0,3891, 0,5616]	0,2425
						2-back		Micro-structural	Number of subordinated clauses	0,2048 [-0,2705, 0,6802]	0,2425
						2-back		Micro-structural	Number of relative clauses	0,146 [-0,3293, 0,6214]	0,2425
						2-back		Macro-structural	Story Structure	0,1064 [-0,369, 0,5818]	0,2425
						2-back		Macro-structural	Referencial Accuracy	0,231 [-0,2443, 0,7064]	0,2425
					Interference Control	Local-to-Global (Accuracy)		Micro-structural	Lexical Variety	0,0621 [-0,4133, 0,5374]	0,2425
						Local-to-Global (Accuracy)		Micro-structural	Morphosintactic Complexity	0,2779 [-0,1974, 0,7533]	0,2425
						Local-to-Global (Accuracy)		Micro-structural	Number of subordinated clauses	0,045 [-0,4303, 0,5204]	0,2425
						Local-to-Global (Accuracy)		Micro-structural	Number of relative clauses	0,9417 [0,4663, 1,4171]	0,2425
						Local-to-Global (Accuracy)		Macro-structural	Story Structure	0,2342 [-0,2412, 0,7096]	0,2425
						Local-to-Global (Accuracy)		Macro-structural	Referencial Accuracy	0,1389 [-0,3365, 0,6142]	0,2425
					Interference Control	Global-to-Local (Accuracy)		Micro-structural	Lexical Variety	0,041 [-0,4343, 0,5164]	0,2425
						Global-to-Local (Accuracy)		Micro-structural	Morphosintactic Complexity	0,5139 [0,0386, 0,9893]	0,2425

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
						Global-to-Local (Accuracy)		Micro-structural	Number of subordinated clauses	0,0923 [-0,3831, 0,5676]	0,2425
						Global-to-Local (Accuracy)		Micro-structural	Number of relative clauses	0,7137 [0,2384, 1,1891]	0,2425
						Global-to-Local (Accuracy)		Macro-structural	Story Structure	0,3496 [-0,1258, 0,8249]	0,2425
						Global-to-Local (Accuracy)		Macro-structural	Referencial Accuracy	0,0701 [-0,4052, 0,5455]	0,2425
					Interference Control	Local-to-Global (Reaction Time)		Micro-structural	Lexical Variety	1,211 [0,7357, 1,6864]	0,2425
						Local-to-Global (Reaction Time)		Micro-structural	Morphosintactic Complexity	0,5308 [0,0554, 1,0062]	0,2425
						Local-to-Global (Reaction Time)		Micro-structural	Number of subordinated clauses	0,2877 [-0,1877, 0,763]	0,2425
						Local-to-Global (Reaction Time)		Micro-structural	Number of relative clauses	0,3507 [-0,1247, 0,8261]	0,2425
						Local-to-Global (Reaction Time)		Macro-structural	Story Structure	0,7582 [0,2828, 1,2335]	0,2425
						Local-to-Global (Reaction Time)		Macro-structural	Referencial Accuracy	0,2533 [-0,2221, 0,7286]	0,2425
					Interference Control	Global-to-Local (Reaction Time)		Micro-structural	Lexical Variety	0,1206 [0,3548, 0,5959]	0,2425
						Global-to-Local (Reaction Time)		Micro-structural	Morphosintactic Complexity	0,0741 [-0,4012, 0,5495]	0,2425
						Global-to-Local (Reaction Time)		Micro-structural	Number of subordinated clauses	0,1186 [-0,3568, 0,5939]	0,2425
						Global-to-Local (Reaction Time)		Micro-structural	Number of relative clauses	0,3586 [-0,1167, 0,834]	0,2425
						Global-to-Local (Reaction Time)		Macro-structural	Story Structure	0,6155 [0,1402, 1,0909]	0,2425
						Global-to-Local (Reaction Time)		Macro-structural	Referencial Accuracy	0,002 [-0,4734, 0,4774]	0,2425
Puranik, 2006 (dissertation)	USA	Typically Developing (n = 90)	10,22	8-12	Working Memory capacity	Competing Language Processing Task	Written	Micro-structural	Total Lexical Production	0,4001 [0,1899, 0,6102]	0,1072
					Working Memory capacity	Digit Ordering			Total Lexical Production	0,3316 [0,1215, 0,5418]	0,1072

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
					Working Memory capacity	Competing Language Processing Task		Macro-structural	Information	0,4118 [0,2017, 0,6219]	0,1072
					Working Memory capacity	Digit Ordering			Information	0,3884 [0,1783, 0,5986]	0,1072
					Working Memory capacity	Competing Language Processing Task		Micro-structural	Number of Utterance	0,2986 [0,0884, 0,5087]	0,1072
					Working Memory capacity	Digit Ordering			Number of Utterance	0,2661 [0,056, 0,4762]	0,1072
Salas & Silvente, 2020	Spain	Typically Developing (n = 1337)	10,17	7-14	Interference Control	Stroop	Written	Micro-structural	Mean Length of Utterance	0,0802 [0,0265, 0,1338]	0,0265
					Working Memory capacity	Digit Span		Micro-structural	Total Lexical Production	0,2237 [0,17, 0,2773]	0,0265
					Working Memory capacity	Digit Span		Micro-structural	Mean Length of Utterance	0,0802 [0,0265, 0,1338]	0,0265
					Interference Control	Stroop		Micro-structural	Total Lexical Production	0,2342 [0,1805, 0,2879]	0,0265
Swanson & Berninger, 1996a	USA	Typically Developing (n = 300)	11,09	9-12	Working Memory capacity	Listening Recall, Listening Generate Recall	Written	Micro-structural	Number of Utterance	0,2769 [0,1631, 0,3906]	0,0583
						Listening Recall, Listening Generate Recall		Macro-structural	Content and organization	0,2554 [0,1417, 0,3691]	0,0583
					Working Memory capacity	Matrix		Micro-structural	Number of Utterance	0,0601 [-0,0537, 0,1738]	0,0583
						Matrix		Macro-structural	Content and organization	0,1206 [0,0069, 0,2343]	0,0583
Swanson & Berninger, 1996b	USA	Typically Developing (n = 50)	10,50	9-12	Working Memory capacity	Sentence Span Test	Written	Macro-structural	Content	0,3095 [0,0236, 0,5945]	0,1459
						Sentence Span Test		Micro-structural	Mean Length of Utterance	0,2769 [-0,009, 0,5628]	0,1459
						Sentence Span Test		Micro-structural	Total Lexical Production	0,3654 [0,0796, 0,6513]	0,1459

References	Location	Clinical Risk Status of the sample	Mean Age (years)	Age Range	EF Domain	EF task	Narrative Form	Narrative Competence	NC indicator	Fisher's Z [95% CI]	SE
Vanderberg & Swanson, 2006	USA	Typically Developing (n = 160)	15,21	14-15	Working Memory capacity	Rhyming words	Written	Macro-structural	Structure	0,182 [0,0256, 0,3384]	0,0800
						Rhyming words		Micro-structural	Total Lexical Production	0,1511 [-0,0053, 0,3076]	0,0800
						Rhyming words		Micro-structural	Morphosintactic Complexity	0,0902 [-0,0662, 0,2467]	0,0800
					Working Memory capacity	Sentence Span		Macro-structural	Structure	0,1104 [-0,046, 0,2669]	0,0800
						Sentence Span		Micro-structural	Total Lexical Production	0,0701 [-0,0863, 0,2265]	0,0800
						Sentence Span		Micro-structural	Morphosintactic Complexity	0,1409 [-0,0155, 0,2973]	0,0800
					Working Memory capacity	Visual Matrix		Macro-structural	Structure	0,0902 [-0,0662, 0,2467]	0,0800
						Visual Matrix		Micro-structural	Total Lexical Production	0,1409 [-0,0155, 0,2973]	0,0800
						Visual Matrix		Micro-structural	Morphosintactic Complexity	0,0601 [-0,0964, 0,2165]	0,0800
					Working Memory capacity	Mapping		Macro-structural	Structure	0,01 [-0,1464, 0,1664]	0,0800
						Mapping		Micro-structural	Total Lexical Production	-0,0601 [-0,2165, 0,0964]	0,0800
						Mapping		Micro-structural	Morphosintactic Complexity	0,02 [-0,1364, 0,1764]	0,0800
Ygual Fernandez et al., 2010	Spain	Attention Deficit Hyperactivity Disorder (n = 26)	8,50	6-11	Behavioural Inhibition	Matching Familiar Figure Test	Oral	Macro-structural	Coherence	0,4236 [0,015, 0,8323]	0,2086
					Working Memory capacity	Digit Span	Oral		Coherence	0,1104 [-0,2982, 05191]	0,2086
					Interference Control	Stroop	Oral		Coherence	0,2661 [-0,1426, 0,6748]	0,2086
					Working Memory capacity	Rey Figure	Oral		Coherence	0,4973 [0,0886, 0,906]	0,2086

REFERENCE LIST

Abel, C., Nerren, J., & Wilson, H. (2015). Leaping the language gap: Strategies for preschool and head start teachers. *International Journal of Child Care and Education Policy, 9*, 7. https://doi.org/10.1186/s40723-015-0009-8

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-and intermediate-grade writers. *Journal of Educational Psychology, 85*(3), 478-508. https://doi.org/10.1037/0022-0663.85.3.478

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), 1698-1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental neuropsychology, 20*(1), 385-406. https://doi.org/10.1207/S15326942DN2001_5

Anderson, P. J., & Reidy, N. (2012). Assessing executive function in preschoolers. *Neuropsychology review, 22*(4), 345-360. https://doi.org/10.1007/s11065-012-9220-3

Antelmi, D. (1997). *La prima grammatica dell'Italiano*. Bologna: Il Mulino

Arán Filippetti, V., & Richaud, M. C. (2015). Do Executive Functions Predict Written Composition? Effects beyond Age, Verbal Intelligence and Reading Comprehension. *Acta Neuropsychologica 13*(4), 331-349.

Artico, M., & Penge, R. (2016). Uno studio trasversale sul contributo delle Funzioni Esecutive nella narrazione scritta in bambini con Disturbo Specifico di Apprendimento [A transversal study on the contribution of Executive Functions on written narratives in children with learning disabilities]. *Giornale di Neuropsichiatria dell'Età Evolutiva, 36*, 14-23.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child neuropsychology*, *5*(2), 115-129.

Ariel, M. (1990). Accessing noun-phrase antecedents. London: Routledge

Ariel, M. (2001). "Accessibility theory: an overview," in *Text Representation: Linguistic and Psycholinguistic Aspects*, eds T. Sanders, J. Schilperoord, and W. Spooren (Amsterdam; Philadelphia, PA: John Benjamins Publishing Company), 29–87

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187-203. doi: 10.1111/j.1749-818X.2010.00193.x

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. Journal of Memory and Language*,* 56(4), 521-536. doi:10.1016/j.jml.2006.09.007

Arnold, J. E., Kaiser, E., Kahn, J. M., & Kim, L. K. (2013). Information structure: linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 403-413. doi: 10.1002/wcs.1234

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*(3), 154-174. https://hdl.handle.net/11245/1.540264

Azar, Z., Backus, A., & Özyürek, A. (2019). General-and language-specific factors influence reference tracking in speech and gesture in discourse. *Discourse Processes*, *56*(7), 553-574.

Balaban, H. O., & Hohenberger, A. (2020). The development of narrative skills in Turkish-speaking children: A complexity approach*. PloS one, 15*(5), e0232579. https://doi.org/10.1371/journal.pone.0232579

Baldo, S., Sgariboldi, V., Venturelli, L., & Marzocchi, G. M. (2017). Standardizzazione della versione italiana del questionario SDQ per bambini dai 3 ai 5 anni. Uno strumento di screening in età prescolare. *Disturbi di attenzione e Iperattività, 13*(1), 11-33. doi: 10.14605/DdAI1311701

Balioussis, C., Johnson, J., & Pascual-Leone, J. (2012). Fluency and complexity in children's writing: The role of mental attention and executive function. *Rivista Di Psicolinguistica Applicata, 12*(3), 33–45.

Bamberg, M. (1987). *The Acquisition of Narratives*. Berlin: Mouton de Gruyter

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

Barkley, R. A. (2006). *Attention defcit hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed.). Guilford Press.

Belletti, A. (1999). Italian/romance clitics: Structure and derivation. In Clitics in the Language of Europe. Edited by H. van Riemsdijk. Berlin: Mounton de Gruyter, pp. 543–79.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Berman, R. A., & Slobin, D. I. (Eds.). (2013). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Berninger, V., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading connections between language by hand and language by eye. *Journal of Learning Disabilities, 35*(1), 39–56. https://doi.org/10.1177/002221940203500104

Berretta, M. (1986). Per uno studio dell'apprendimento dell'italiano in contesto naturale: il caso dei pronomi personali atoni. In Giacolone Ramat (ed.). L'apprendimento spontaneo di una seconda lingua. Bologna, il Mulino: 329-352.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development, 81*(6), 1641-1660. doi: 10.1111/j.1467-8624.2010.01499.x

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental review, 29*(3), 180-200. doi: 10.1016/j.dr.2009.05.002

Bigozzi, L., & Vettori, G. (2016). To tell a story, to write it: developmental patterns of narrative skills from preschool to first grade. *European Journal of Psychology of Education, 31*(4), 461-477. doi: 10.1007/s10212-015-0273-6

Bishop, D.M.V. (1979). Comprehension in developmental disorders. *Developmental Medicine and Child Neurology, 21*, 225-238

Bolander, T. (2014, January). *Seeing is Believing: Formalising False-Belief Task in dynamic epistemic logic* [paper presentation]. European Conference on Social Intelligence, Barcelona, Spain.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221–235). Russell Sage Foundation.

Botting, N., Morgan, G., Jones, A., Marshall, C., Denmark, T., & Atkinson, J. (2017). Nonverbal executive function is mediated by language: A study of deaf and hearing children. *Child Development, 88*, 1689–1700. doi: 10.1111/cdev.12659

Boudreau, D. M., & Chapman, R. S. (2000). The relationship between event representation and linguistic skill in narratives of children and adolescents with Down syndrome. *Journal of speech, language, and hearing research : JSLHR*, *43*(5), 1146–1159. https://doi.org/10.1044/jslhr.4305.1146

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge, UK: Cambridge University Press

Cannizzaro, M., & Coelho, C. A. (2002). Treatment of story grammar following traumatic brain injury: A pilot study. *Brain Injury*, *16*(12), 1065–1073

Cannizzaro, M., & Coelho, C.A. (2013). Analysis of Narrative Discourse Structure as an Ecologically Relevant Measure of Executive Function in Adults. *Journal of Psycholinguistic Research,* 42, 527-549.

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., & Rucián, M. (2016). Development of the updating executive function: From 7-year-olds to young adults. *Developmental Psychology, 52*(4), 666-678. doi: 10.1037/dev0000091

Castilla-Earls A., Petersen, D., Spencer T., & Hammer K. (2015). Narrative Development in Monolingual Spanish-Speaking Preschool Children, *Early Education and Development, 26*(8), 1166-1186. doi: 10.1080/10409289.2015.1027623

Centers for Disease Control and Prevention. (2022, April 11). *Why Act Early if You're Concerned about Development*?. https://www.cdc.gov/ncbddd/actearly/whyActEarly.html

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 2556). New York, NY: Academic Press.

Chafe, W. L. (1994). *Discourse, Consciousness, and Time*. Chicago, IL: Chicago University Press

Chen, L., & Pan, N. (2009). Development of English referring expressions in the narratives of Chinese–English bilinguals. *Bilingualism: Language and Cognition*, *12*, 429–445. doi:10.1017/S1366728909990216

Cochran, W. G. (1954). Some methods for strengthening the common χ 2 tests. *Biometrics*, *10*(4), 417-451. https://doi.org/10.2307/3001616

Coelho, C., Liles, B., and Duffy, R. (1995). Impairments of discourse abilities and executive functions in traumatically brain-injured adults. *Brain Injuries 9,* 471–477. doi: 10.3109/02699059509008206

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Cornish, F. (1999). *Anaphora, Discourse, and Understanding: Evidence from English and French*. Oxford: Clarendon Press

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., De Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric disease and treatment*, *12*, 1191-1202. doi: 10.2147/NDT.S104620

D'Amico, S., Devescovi, A., Marano, A., & Albano, S. (2008). La valutazione della competenza narrativa in bambini prescolari e scolari attraverso un libro illustrato. *Rivista di psicolinguistica applicata*, *8*(1/2), 1000-1018

Davidson, K., Norrie, J., Tyrer, P., Gumley, A., Tata, P., Murray, H., & Palmer, S. (2006). The effectiveness of cognitive behavior therapy for borderline personality disorder: results from the borderline personality disorder study of cognitive therapy (BOSCOT) trial. *Journal of personality disorders, 20*(5), 450-465. doi: 10.1521/pedi.2006.20.5.450

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system*. Psychological Corporation.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, *32*(4), 759-786.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168. doi: 10.1146/annurev-psych-113011-143750

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959-964. doi: 10.1126/science.1204529

Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research & Practice, 16*(4), 186–202. https://doi.org/10.1111/0938-8982.00019

Di Domenico, E., Ioli, B., & Andrea, C. (2020). Null and overt subject pronouns in topic continuity and topic shift: An investigation of the narrative productions of Italian Natives, Greek Natives and near-native second language speakers of Italian with Greek as a first language. *Glossa: a journal of general linguistics 5*(1): 117. 1–35. https://doi.org/10.5334/gjgl.1009

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: an investigation of their narratives and memory. *International journal of language & communication disorders, 43*(2), 201–218. https://doi.org/10.1080/13682820701366147

Drijbooms, E., Groen, M. A., & Verhoeven, L. (2015). The contribution of executive functions to narrative writing in fourth grade children. *Reading and Writing: An Interdisciplinary Journal, 28*(7), 989-1011. https://doi.org/10.1007/s11145-015-9558-z

Drijbooms, E., Groen, M. A., & Verhoeven, L. (2017). How executive functions predict development in syntactic complexity of narrative writing in the upper elementary grades. *Reading and Writing: An Interdisciplinary Journal, 30*(1), 209-231. https://doi.org/10.1007/s11145-016-9670-8

Duinmeijer, I., de Jong, J., & Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders, 47*(5), 542-555. doi: 10.1111/j.1460-6984.2012.00164.x

Fazio, B. B., Naremore, R. C. & Connell, P. J. (1996). Tracking children from poverty at risk for specific language impairment: A 3-year longitudinal study. *Journal of Speech and Hearing Research, 39,* 611–624. doi: 10.1044/jshr.3903.611.

Fernández, A. Y., Miranda, B. R., & Casas, A. M. (2010). Funciones ejecutivas, comprensión de historias y coherencia narrativa en niños con trastorno por déficit de atención con hiperactividad. *Revista de Logopedia, Foniatría y Audiología, 30*(3), 151-161. http://hdl.handle.net/11162/28945

Fiorentino, L., & Howe, N. (2004). Language Competence, Narrative Ability, and School Readiness in Low-Income Preschool Children. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *36*(4), 280-300.

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35*(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5

Friend, M., & Phoenix-Bates, R. (2014). The union of narrative and executive function: different but complementary. *Frontiers in psychology, 5,* 469. doi: 10.3389/fpsyg.2014.00469

Gamannossi, B. A., & Pinto, G. (2014). Theory of mind and language of mind in narratives: Developmental trends from kindergarten to primary school. *First Language, 34*(3), 262-272.

Gandolfi, E., Viterbori, P., Traverso, L., & Usai, M. C. (2014). Inhibitory processes in toddlers: a latent-variable approach. *Frontiers in psychology, 5,* 381. doi: 10.3389/fpsyg.2014.00381

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gioia, G. A., Isquith, P. K., & Kenealy, L. E. (2008). Assessment of behavioral aspects of executive function. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.), *Executive functions andthe frontal lobes* (pp. 179–202). Psychology Press

Givòn, T. (1983). Topic and continuity in discourse: An introduction. In T. Givo´n (Ed.), *Topic continuity in discourse: A quantitative crosslanguage study* (pp. 142). Amsterdam: John Benjamins Publishing.

Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., & Hulme, C. (2016). The development of executive function and language skills in the early school years. *Journal of child psychology and psychiatry, and allied disciplines, 57*(2), 180–187. https://doi.org/10.1111/jcpp.12458

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2003). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry*, *15*(1-2), 166-172.

Gordon, P. C., Grosz, B. J., and Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science 17*(3), 311–347. doi: 0.1207/s15516709cog1703_1

Grafman, J. (1995). Similarities and distinctions among current models of prefrontal cortical functions. In K. J. Holyoak & F. B. J. Grafman (Eds.), *Structure and functions of the human prefrontal cortex* (pp. 337–368). New York: New York Academy of Sciences.

Grafman, J., & Litvan, I. (1999). Importance of deficits in executive functions. *The Lancet*, *354*(9194), 1921–1923

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral Discourse in the Preschool Years and Later Literacy Skills. *First Language, 24*(2), 123–147. https://doi.org/10.1177/0142723704042369

Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering: a framework for modelling the local discourse. *Computational Linguistics 21*, 203–225. doi: 10.21236/ADA324949

Gundel, J. (1998). "Centering theory and givenness hierarchy:towards a synthesis," in *Centering Theory in Discourse*, eds M. Walker, A. Joshi, and E. Prince (Oxford: Clarendon Press), 183–198.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274-307.

Halliday, M.A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman

Heilmann, J., Nockerts, A., & Miller, J. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*(4), 393–404.

Hendriks, P., Englert, C., Wubs, E., & Hoeks, J. (2008). Age differences in adults' use of referring expressions. Journal of Logic, Language and Information, 17(4), 443-466. doi:10.1007/s10849-008-9065-6

Hendriks, P., Koster, C., & Hoeks, J. C. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, cognition and neuroscience*, *29*(4), 391-407. https://doi.org /10.1080/01690965.2013.766356

Henry, L. A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and Adolescent Mental Health, 15*(2), 110-119. http://dx.doi.org/10.1111/j.14753588.2010.00557.x

Hickmann, M. & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: a comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language 26*(2), 419–452 doi:10.1017/S0305000999003785

Hickmann, M. (2003). *Children's Discourse. Person, Space and Time across Languages.* Cambridge: Cambridge University Press.

Howard, S. J., & Melhuish, E. (2017). An early years toolbox for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. *Journal of Psychoeducational Assessment*, *35*(3), 255-275.

Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking

Johnston, J. R. (2008). Narratives: Twenty-five years later. *Topics in language disorders, 28*(2), 93-98.

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing schoolage children's narrative performances. *American Journal of Speech-Language Pathology, 15*(2), 177-191.

Justice, L. M., Bowles, R., Pence, K., & Gosse, C. (2010). A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly, 25*(2), 218-234. https://doi.org/10.1016/j.ecresq.2009.11.002

Kail, M. (1998). Referent introduction in narratives as a function of mutual knowledge: a developmental comparison between French and Spanish. *Psychology of Language and Communication 2*, 23–46.

Kail, Miche`le, Sanchez y Lopez, Ines, 1997. Referent introduction in Spanish children's narratives as a function of contextual constraints: a crosslinguistic perspective. *First Language 17*(51), 103–130. https://doi.org/10.1177/014272379701705010

Karmiloff-Smith, A. (1981). The grammatical marking of thematic structure in the development of language production. In: Deutsch, W. (Ed.), The Child's Construction of Language. Academic Press, New York.

Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes 1*(1), 61–85 https://doi.org/10.1080/01690968508402071

Khan, K. S. (2013). *The relation between cognitive skills and language skills in typically developing 3 ½ to 6 year olds*. [Doctoral dissertation, The Pennsylvania State University]. PennState University Libraries.

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://hdl.handle.net/11245/1.355042

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in medicine, 22*(17), 2693-2710. doi: 10.1002/sim.1482

Kuijper, S. J. M., Hartman, C. A., and Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of Ambiguous pronouns. *PLoS ONE* 10:e0132408. doi: 10.1371/journal.pone.0132408

Kochhar, R. (2017). Middle Class Fortunes in Western Europe. LIS Working papers 702, LIS Cross-National Data Center in Luxembourg.

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

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology, 108*(1), 1-24.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech, Language, and Hearing Research, 38*(2), 415-425.

Luo, F., & Timler, G.R. (2008). Narrative organization skills in children with attention deficit hyperactivity disorder and language impairment: Application of the causal network model. *Clinical Linguistic & Phonetics, 22,* 25-46

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* Mahwah, NJ: Erlbaum.

Malvern, D., Richards, B., Chipere, N., & Purán, P. (2004). *Lexical diversity and language development*. New York, NY: Palgrave Macmillan.

Mäkinen, L., Loukusa, S., Nieminen, L., Leinonen, E., & Kunnari, S. (2014). The development of narrative productivity, syntactic complexity, referential cohesion and event content in four-to eight-year-old Finnish children. First Language*,* 34(1), 24-42. http://dx.doi.org/10.1177/0142723713511000

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia, 42*(10), 1414-1434. doi: 10.1016/j.neuropsychologia.2003.12.016

Marini, A., Sperindè, P., Ruta, I., Savegnago, C., & Avanzini, F. (2019). Linguistic skills in bilingual children with developmental language disorders: A pilot study. *Frontiers in Psychology*, *10*, 493.

Marini, A., Piccolo, B., Taverna, L., Berginc, M., & Ozbič, M. (2020). The complex relation between executive functions and language in preschoolers with Developmental Language Disorders. *International journal of environmental research and public health*, *17*(5), 1772-1792.

McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing, 24*, 459-488. https://doi.org/10.1177/0265532207080767

McCormack, T., & Atance, C. M. (2011). Planning in young children: A review and synthesis. *Developmental Review, 31*(1), 1-31. https://doi.org/10.1016/j.dr.2011.02.002

McCormack, T., & Hanley, M. (2011). Children's reasoning about the temporal order of past and future events. *Cognitive Development, 26*(4), 299-314. https://doi.org/10.1016/j.cogdev.2011.10.001

McNiven, C. L. (2007). *The Relation Between Cognitive Skills and Language Skills in Typically Developing 3 1/2 to 6 Year Olds*. [Master dissertation, The University of British Columbia]. PennState University Libraries.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental psychology, 49*(2), 270-291. doi: 10.1037/a0028228

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. doi: 10.1006/cogp.1999.0734.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the national Academy of Sciences, 108*(7), 2693-2698. https://doi.org/10.1073/pnas.1010076108

Monette, S., Bigras, M., & Lafrenière, M. A. (2015). Structure of executive functions in typically developing kindergarteners. *Journal of experimental child psychology, 140*, 120-139. doi: 10.1016/j.jecp.2015.07.005

Morra, S., Panesi, S., Traverso, L., & Usai, M. C. (2018). Which tasks measure what? Reflections on executive function development and a commentary on Podjarny, Kamawar, and Andrews (2017). *Journal of Experimental Child Psychology, 167*, 246-258. doi:10.1016/j.jecp.2017.11.004

Mozeiko, J., Le, K., Coelho, C., Krueger, F., and Grafman, J. (2011). The relationship of story grammar and executive function following TBI. *Aphasiology 25*, 826–835. doi: 10.1080/02687038.2010.543983

Müller, U., Lieberman, D., Frye, D., & Zelazo, P. D. (2008). Executive function, school readiness, and school achievement. In S. K. Thurman & C. A. Fiorello (Eds*.), Applied cognitive research in K–3 classrooms* (pp. 41–83). Routledge/Taylor & Francis Group.

Nanson, A. (2021). *Storytelling and ecology: Empathy, enchantment and emergence in the use of oral narratives.* Bloomsbury Publishing.

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of child psychology and psychiatry*, *58*(4), 361-383. https://doi.org/10.1111/jcpp.12675

Norbury, C. F., & Bishop, D. V. (2003). Narrative skills of children with communication impairments. *International journal of language & communication disorders*, *38*(3), 287-313. Doi:10.1080/136820310000108133

Odato, C. V. (2013). The development of children's use of discourse like in peer interaction. *Speech, 88,* 117–143. doi: 10.1215/00031283-2346825

Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin, 118*(1), 155–164. https://doi.org/10.1037/0033-2909.118.1.155

Olinghouse, N. G. (2008). Student- and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing, 21*(1–2), 3–26. https://doi.org/10.1007/s11145-007-9062-1

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool Children's Narratives and Performance on the Peabody Individualized Achievement Test – Revised: Evidence of a Relation between Early Narrative and Later Mathematical Ability. *First Language, 24*(2), 149–183. https://doi.org/10.1177/0142723704043529

Orsolini, M. (1990). Episodic structure in children's fantasy narratives: Breakthroughs to decontextualized discourse. *Language and Cognitive Processes, 5*(1), 53-79

Orsolini, M. (1995). *L'acquisizione delle competenze pragmatiche* [The acquisition of pragmatic competencies], in G. Sabbadini (Eds.), *Manuale di neuropsicologia dell'età evolutiva*. Bologna: Il Mulino.

Orsolini, M., Rossi, F., & Pontecorvo, C. (1996). Re-introduction of referents in Italian children's narratives. *Journal of Child Language*, *23*(2), 465-486.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews, 10*(1), 1-11. doi: /10.1136/bmj.n71

Park, H. (2014). Language Skills, Oral Narrative Production, and Executive Functions of Children Who Are Deaf or Hard of Hearing. [Doctoral Dissertation, The University of Tennessee Health Science Center]. doi:10.21007/etd.cghs.2014.0238

Peristeri, E., Baldimtsi, E., Andreou, M., & Tsimpli, I. M. (2020). The impact of bilingualism on the narrative ability and the executive functions of children with autism spectrum disorders. *Journal of communication disorders, 85*, 105999. https://doi.org/10.1016/j.jcomdis.2020.105999

Perner, J. and Lang, B. (1999) Development of Theory of Mind and Executive Control. Trends in Cognitive *Sciences, 9*, 337-344. http://dx.doi.org/10.1016/S1364-6613(99)01362-5

Petersen, D. B. (2011). A systematic review of narrative-based language intervention with children who have language impairment. *Communication Disorders Quarterly, 32*(4), 207-220. https://doi.org/10.1177/1525740109353937

Peterson, R. A., & Brown, S. P. (2005). On the Use of Beta Coefficients in Meta-Analysis. *Journal of Applied Psychology, 90*(1), 175–181. https://doi.org/10.1037/0021-9010.90.1.175

Piantadosi, S., Byar, D. P., and Green, S. B. (1988). The ecological fallacy. *American Journal of Epidemiology, 127*(5), 893–904. ttps://doi.org/10.1093/oxfordjournals.aje.a114892

Pinto, G. (2003). *Il suono, il segno, il significato. Psicologia dei processi di alfabetizzazione* [Sound, sign, meaning. Psychology of literacy processes]. Carocci Editore.

Pinto, G., Tarchi, C., & Bigozzi, L. (2016). Development in narrative competences from oral to written stories in five-to seven-year-old children. *Early Childhood Research Quarterly*, *36*, 1-10.

Pizzuto E., & Caselli M.C. (1992). The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language, 19*, 491-557. https://doi.org/10.1017/S0305000900011557

Puranik, C. S. (2006). *Expository writing skills in elementary school children from third through sixth grades and contributions of short -term and working memory* [Doctoral Dissertation, University of Florida]. ProQuest Dissertations Publishing.

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research 49*(4), 793-808. https://doi.org/10.1044/1092-4388(2006/056)

Ripich, D. N., & Griffith, P. L. (1988). Narrative abilities of children with learning disabilities and nondisabled children: Story structure, cohesion, and propositions. *Journal of learning disabilities*, *21*(3), 165-173.

RStudio Team (2020). *RStudio: Integrated Development for R. RStudio*, PBC, Boston, MA URL http://www.rstudio.com/.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. Applied Psycholinguistics, 37(01), 49-67. doi:10.1017/S0142716415000417

Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., et al. (2002). Executive control function: A review of its promise and challenges for clinical research. A report from the committee on research of the American neuropsychiatric association. *Journal of Neuropsychiatry Clinical Neuroscience*, *14*(4), 377–405

Roth, F. P., & Spekman, N. J. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally achieving students. *Journal of Speech and Hearing Disorder, 51*(1), 8-23.

Sacchetti, S. A. (2018). *Lo sviluppo delle competenze narrative e funzioni esecutive in età prescolare* [The development of narrative competence and executive functions in preschool age].
[Master dissertation, University of Milan Bicocca].

Salas, N., & Silvente, S. (2020). The role of executive functions and transcription skills in writing: A cross-sectional study across 7 years of schooling. *Reading and Writing, 33*(4), 877-905. doi: 10.1007/s11145-019-09979-y

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools, 41*, 459–473. doi:10.1044/0161-1461(2010/09-0040)

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton narrative norms instrument. *Journal of speech language pathology and audiology, 30*(4), 224-235.

Schraeder, T., Quinn, M., Stockman, I. J., & Miller, J. (1999). Authentic assessment as an approach to preschool speech-language screening. *American Journal of SpeechLanguage Pathology, 8*(3), 195-200. https://doi.org/10.1044/1058-0360.0803.195

Scionti, N., Cavallero, M., Zogmaister, C., & Marzocchi, G. M. (2020). Is cognitive training effective for improving executive functions in preschoolers? A systematic review and meta-analysis. *Frontiers in psychology, 10*, 2812. https://doi.org/10.3389/fpsyg.2019.02812

Scionti, N., & Marzocchi, G. M. (2021). The dimensionality of early executive functions in young preschoolers: Comparing unidimensional versus bidimensional models and their ecological validity. *Child Neuropsychology, 27*(4), 491-515. https://doi.org/10.1080/09297049.2020.1868419

Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. Applied Psycholinguistics*,* 26(3), 437-462.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British journal of psychology* (London, England: 1953), *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental psychology, 27*(6), 960-974. doi: 10.1037/0012-1649.27.6.960

Shonkoff, J. & Phillips, D. 2000. *From neurons to neighbourhoods: The science of early childhood development.* National Academy Press.

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

Spencer, T. D., Weddle, S. A., Petersen, D. B., & Adams, J. A. (2018). Multi-tiered narrative intervention for preschoolers: A Head Start implementation study. *NHSA Dialog, 20*(1), 1-28.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Frredle (Ed.) *New directions in discourse processing.* Erlbaum.

Stein, N.L. (1988). *The development of children's storytelling skill*. In M. Franklin & S. Barten (Eds.), *Child language* (282-287). Oxford University Press.

Strömqvist, S., & Verhoeven, L. T. (2004). *Relating events in narrative: Typological and contextual perspectives* (Vol. 2). Psychology Press

Swanson, H. L., & Berninger, V. W. (1996a). Individual differences in children's working memory and writing skill. *Journal of experimental child psychology, 63*(2), 358-385. doi: 10.1006/jecp.1996.0054.

Swanson, H. L. & Berninger, V. W. (1996b). Individual differences in children's writing: A function of working memory or reading or both processes? *Reading and Writing: An Interdisciplinary Journal 8*, 357-383.

Thompson, S. G. and Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*(11), 1559–1573. https://doi.org/10.1002/sim.1187

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental science, 12*(1), 106-113. doi: 10.1111/j.1467-7687.2008.00745.x

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209–253. doi: 10.1016/S0010-0277(99)00069-4

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

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct?. *Journal of child psychology and psychiatry, 54*(2), doi: 131-143. 10.1111/jcpp.12001

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

Usai, M. C., Viterbori, P., Traverso, L., & De Franchis, V. (2014). Latent structure of executive function in five-and six-year-old children: A longitudinal study. *European Journal of Developmental Psychology, 11*(4), 447-462. https://doi.org/10.1080/17405629.2013.840578

Vallotton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly, 26*(2), 169-181. doi: 10.1016/j.ecresq.2010.09.002

Van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adult's memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. *Child Development, 67*, 3010–3028.

Van der Linden, M., Meulemans, T., Marczewski, P., & Collette, F. (2000). The relationships between episodic memory, working memory, and executive functions: The contribution of the prefrontal cortex. *Psychologica Belgica*, *40*(4) 275-297.

Van der Sluis, S., De Jong, P. F., & Van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*(5), 427-449. https://doi.org/10.1016/j.intell.2006.09.001

Vanderberg, R., & Lee Swanson, H. (2007). Which components of working memory are important in the writing process? *Reading and Writing, 20*(7), 721-752. https://doi.org/10.1007/s11145-006-9046-6

Veraksa, A., Bukhalenkova, D., Kartushina, N., & Oshchepkova, E. (2020). The relationship between executive functions and language production in 5–6-year-old children: Insights from working memory and storytelling. *Behavioral Sciences, 10*(2):52. https://doi.org/10.3390/bs10020052

Viechtbauer, W., & Viechtbauer, M. W. (2015). *Package 'metafor'. The Comprehensive R Archive Network.* Package 'metafor'. http://cran. r-project. org/web/packages/metafor/metafor. pdf.

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory–Is younger, better?. *Developmental Review, 32*(4), 360-387. https://doi.org/10.1016/j.dr.2012.07.001

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental neuropsychology, 4*(3), 199-230. https://doi.org/10.1080/87565648809540405

Westerveld, M. F., Gillon, G. T., & Miller, J. F. (2004). Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech Language Pathology, 6*(4), 195–208.

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. doi: 10.1016/j.jecp.2010.08.008

Willoughby, M. T., Blair, C. B., Kuhn, L. J., & Magnus, B. E. (2018). The benefits of adding a brief measure of simple reaction time to the assessment of executive function skills in early childhood. *Journal of experimental child psychology*, *170*, 30-44.

Ylvisaker, M., Szekeres, S., & Feeney, T. (2008). Communication disorders associated with traumatic brain injury. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 879–962). Philadelphia: Lippincott, Williams & Wilkins

Zanchi, P., & Zampini, L. (2020). The narrative competence task. European Journal of Psychological Assessment, 37, 15-22 https://doi.org/10.1027/1015-5759/a000569.

Zelazo, P. D., & Jacques, S. (1996). *Children's rule use: Representation, reflection, and cognitive control.* In R. Vasta (Ed.), Annals of child development: A research annual, 12 (pp. 119–176). Jessica Kingsley Publishers.