Journal of Experimental Child Psychology 233 (2023) 105693

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

How executive functioning, sentence processing, and vocabulary are related at 3 years of age

Crystal Lee [a], Andrew Jessop [b], Amy Bidgood [c], Michelle S. Peter [d], Julian M. Pine [b], Caroline F. Rowland [b,e,f,*], Samantha Durrant [g]

[a] *Department of Psychology, Princeton University, Princeton, NJ 08540, USA*
[b] *Department of Psychology, University of Liverpool, Liverpool L69 7ZA, UK*
[c] *School of Psychology, Liverpool John Moores University, Liverpool L3 3AF, UK*
[d] *North Thames Genomic Laboratory Hub, Great Ormond Street Hospital for Children, London WC1N 3BH, UK*
[e] *Language Development Department, Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands*
[f] *Donders Centre for Cognition, Radboud University, 6525 GD Nijmegen, The Netherlands*
[g] *School of Health Sciences, University of Manchester, Manchester M13 9PL, UK*

ARTICLE INFO

Article history:
Received 22 June 2022
Revised 30 March 2023

Keywords:
Executive functions
Language development
Sentence processing
Eye tracking
Individual differences
Child development

ABSTRACT

There is a wealth of evidence demonstrating that executive function (EF) abilities are positively associated with language development during the preschool years, such that children with good executive functions also have larger vocabularies. However, why this is the case remains to be discovered. In this study, we focused on the hypothesis that sentence processing abilities mediate the association between EF skills and receptive vocabulary knowledge, in that the speed of language acquisition is at least partially dependent on a child's processing ability, which is itself dependent on executive control. We tested this hypothesis in longitudinal data from a cohort of 3- and 4-year-old children at three age points (37, 43, and 49 months). We found evidence, consistent with previous research, for a significant association between three EF skills (cognitive flexibility, working memory [as measured by the Backward Digit Span], and inhibition) and receptive vocabulary knowledge across this age range. However, only one of the tested sentence processing abilities (the ability to maintain multiple possible referents in mind) significantly mediated this relationship and only for one of the tested EFs (inhibition). The results suggest that children who are better able to inhibit incorrect responses are also better able to maintain multiple possible referents in mind

* Corresponding author.
E-mail address: caroline.rowland@mpi.nl (C.F. Rowland).

https://doi.org/10.1016/j.jecp.2023.105693

while a sentence unfolds, a sophisticated sentence processing ability that may facilitate vocabulary learning from complex input.

Introduction

Executive function (EF; sometimes called cognitive or executive control) is an umbrella term referring to the abilities implicated in tasks that require individuals to regulate and control their behavior (e.g., paying attention, organizing, planning, self-monitoring; Wiebe et al., 2011). EFs are positively associated with language development during childhood in that children with good EFs tend to have larger vocabularies across the preschool and early school years. This applies to both children with language disorders (Henry et al., 2012; Wittke et al., 2013) and those without them (Carlson et al., 2005; Fuhs & Day, 2011; Gathercole & Pickering, 2000; Kuhn et al., 2016; O'Neill & Miller, 2013; Weiland et al., 2014).

However, the underlying cause of the association between EF and language is still not clear. Bishop et al. (2014) proposed three possible models. One is that some children develop faster than others across the board. In this view, there is no direct or indirect causal relationship between the development of EF and language; they simply grow in synchrony (e.g., as a result of development of the frontal lobes). A second model is that there is a direct causal relationship in which having a larger vocabulary promotes EF (e.g., Kuhn et al., 2014). For example, some have suggested that vocabulary supports the development of EF by facilitating the creation of rules and symbols that allow children to create internal representations, hence enhancing their ability to monitor their own behavior (Jones et al., 2020; Zelazo & Frye, 1998).

There is, however, a third possibility, which was the focus of the current study. This is the idea that as EF skills develop, these developments have a direct effect on children's ability to learn language by enabling them to process the incoming input faster and more efficiently (e.g., Diamond, 2013; Weiland et al., 2014). On this view, the speed of language acquisition is at least partially dependent on a child's processing ability, which is itself associated with EF ability. Language acquisition by children often seems effortless to observers, but in fact it is underpinned by a number of online processing tasks, all of which impose a certain level of cognitive load. For example, in order to learn from the input, children need to be able to sustain attention to the stimuli, rapidly switch attention between multiple streams of information (from visual to auditory cues and back again in spoken language; e.g., Yu et al., 2019), retain information in phonological working memory (Gathercole & Baddeley, 1993), and monitor and rapidly update incorrect inferences (Diamond, 2013; Weiland et al., 2014). All these processing tasks recruit, and thus are dependent on, the same EF mechanisms (including working memory, inhibition, and cognitive flexibility) that govern the regulation and control of behavior in both adults (Huettig & Janse, 2016; Nozari et al., 2016; Vuong & Martin, 2014) and children (Khanna & Boland, 2010; Woodard et al., 2016). Thus, developing language is inherently reliant on successful online sentence processing strategies, many of which themselves are reliant on executive functioning. And executive functioning is itself developing during this period.

It has proven challenging to develop a comprehensive theory of EF development, partly because it is not always clear what functions different tasks tap into and partly because it is not yet clear whether EFs during the preschool years reflect a single unitary factor (e.g., Wiebe et al., 2011) or diverse but correlated latent factors (e.g., Camerota et al., 2020). However, it is clear that children's EF ability develops rapidly during the preschool and early school years, from the ability to solve simple tasks (e.g., inhibiting a motor response, keeping strings of increasing length in working memory) to mastery of more complex control tasks (e.g., the development of complex rules that control two or more responses) (see Best & Miller, 2010, for a developmental review).

At the same time, children's processing abilities are developing in tandem with executive functioning, such that faster processors tend to have better performance on EF tasks (Willoughby et al., 2020).

In fact, processing speed has been argued to underpin performance on many, if not all, EF constructs (e.g., working memory; Fry & Hale, 1996) as well as how EF relates to academic achievement (e.g., Gordon et al., 2018). Similarly, online processing abilities are also associated with vocabulary development (Borovsky et al., 2012; Fernald & Marchman, 2012; Fernald et al., 2006). For example, in Fernald and Marchman's (2012) longitudinal study of typically developing children and late talkers, children who were faster to recognize words in sentences in a looking-while-listening task at 18 months of age (e.g., faster to look at the ball when asked to "look at the ball") were more likely to show accelerated vocabulary growth compared with their counterparts with slower processing speed. Borovsky et al. (2012) reported similar results in a sentence processing task; they found that children's vocabulary size, not age, was related to their ability to predict referents in an upcoming speech stream.

This sentence processing task of Borovsky et al. (2012) is of particular interest because it measured children's ability to integrate information from two elements in a sentence in order to successfully predict a third element, an ability that may well rely on good executive functioning. In this task, children's eye movements were tracked using an eye tracker while the children saw four pictures on a screen and listened to sentences. For example, children might hear "The pirate hides the treasure" and see images of the *treasure* (target), a *ship* (agent-related foil), *bones* (action-related foil), and a *cat* (unrelated foil). Eye movements were analyzed after the verb was heard but before the onset of the final noun (*treasure*). The key ability tested was whether participants made anticipatory looks at the target image (*treasure*) before the onset of the target word itself. In other words, could children correctly anticipate the identity of the final noun in the sentence by integrating information from the agent and action words? Borovsky et al. reported that even 3-year-old children made successful predictions, but importantly those children who were better able to predict the identity of the target image had better age-normalized scores on a test of receptive vocabulary.

Crucially, relying on information from the action or agent word alone did not provide sufficient information for target identification in this task because two of the foils were associated with these elements (*pirate* is also associated with *ship,* and *hide* is also associated with *bones*). Instead, in order to anticipate the target, children needed to perform a "complex calculation of higher order contingencies among the agent, action, and patient" (Borovsky et al., 2012, p. 429), which required them to integrate information not only from the agent and action but also from their real-world knowledge of events and situations (that *pirates* are more likely to *hide treasure* than to *hide bones* or *cats*). And they needed to do this quickly. Thus, it is highly probable that such complex sentence processing tasks are, at least in part, reliant on children having good EF ability.

In sum, children with good EFs should be faster at processing sentences, and those who are faster at processing sentences should be able to learn more words more rapidly and thus develop a larger lexicon. We suggest that this might be a (partial) explanation of the association between EF ability and language during early childhood. In other words, we predicted that the association between EF and language would be (at least partially) mediated by children's online language processing ability.

We tested this hypothesis using longitudinal data from 3- and 4-year-old children (aged 37, 43, and 49 months). At each age point, we collected measures of receptive vocabulary, EF ability (cognitive flexibility, working memory, and inhibition), and sentence processing abilities. To measure EF, the children took part in three EF tasks (measuring cognitive flexibility, working memory, and inhibition, respectively) at each of the three age points (37, 43, and 49 months). Tasks were chosen based on Carlson's (2005) review of developmentally sensitive EF tasks. Within the literature, there is debate about whether the different components of EF reflect a single unitary EF ability during the preschool years or whether they are distinct but correlated abilities, with different components developing at different rates (see Best & Miller, 2010). As a result, we preregistered a confirmatory factor analysis (CFA) to determine whether performance on our tasks could be best described as an underlying unitary latent EF ability or as three separate but related components. To foreshadow the results, we did not find evidence for a latent variable, so we continued the analyses using the separate EF tasks.

To measure sentence processing, we used the same task as Borovsky et al. (2012), described above, and calculated two different types of anticipatory eye movements. First, we calculated the anticipatory looks that the children made to the target image (*treasure* in the example above) before hearing the target noun. This measures integration ability because it is only by integrating information from both the agent and verb (by paying attention to both *pirate* and *hides*) that it is possible to identify the likely target (*treasure*) rather than the other items (*ship* and *bone,* which are associated with either *pirate* or *hide* but not both). We suggest that children who possess better EF abilities should be more likely to integrate information from the agent and verb to successfully predict the identity of the target. Second, we calculated the anticipatory looks that children made to the agent-related distractor (*ship)* before hearing the target noun, as a proportion of all the looks to the distractors (i.e., excluding looks to the target). This measures maintenance ability; the ability to simultaneously keep (or maintain) a plausible alternative referent in mind, which we argue is a useful sentence processing skill in cases where the listener has mis-parsed the speech stream. We suggest that children who possess better EF abilities should be more able to maintain the identity of the agent as a potential target and thus look more at the agent-related distractor than at the other distractors (see Woodard et al., 2016, who demonstrated that EF abilities are related to comprehension of garden path sentences in children, which require a similar kind of ability).

Method

Participants

The data presented here come from the Language 0–5 Project, a large longitudinal project that tracked the spoken English language development of a cohort of 95 children living in North West England (United Kingdom) from 6 to 54 months of age. At the time of recruitment, all infants were typically developing and born full-term, and none was low birthweight. Over the course of the project, participants took part in regular testing at specified age points approximately 3 to 6 months apart.

The tasks discussed in the current article were administered at three age points: 37 months (M_{age} = 37;13 [months;days], range = 36;29–38;03, *n* contributing data to at least one task = 73 [38 girls], 63 of whom took part in the sentence processing task), 43 months (M_{age} = 43;14, range = 42;12–44;05, *n* = 75 [39 girls], 65 in the sentence processing task), and 49 months (M_{age} = 49;17, range = 48;17–50;09, *n* = 71 [37 girls], 59 in the processing task). These ages were chosen because the predictive sentence processing task of Borovsky et al. (2012) demonstrated anticipatory looking from 3 years of age, and the same task was used here to assess the children's sentence processing abilities. At each age point, we collected data on receptive vocabulary, EF, and sentence processing abilities.

Stimuli and procedure

British Picture Vocabulary Scale

The British Picture Vocabulary Scale–Third Edition (BPVS) is a standardized measure of receptive vocabulary size, normed with 3- to 16-year-old children learning British English, with good reliability and validity (Dunn et al., 2009). Children are asked to point to the picture that best matches a word's meaning from an array of four images. A stopping rule is applied when children respond incorrectly to 8 or more target items in a set (there are 14 sets, each containing 12 arrays/target items). Our analyses used total raw score, which is calculated by summing the total number of correct picture choices made by children before the stopping rule is applied.

EF battery

To choose EF tasks that both were suitable for children at this young age (3–4 years) and measured core abilities of EF, we reviewed the literature on EF in preschoolers. We used, in particular, Carlson (2005), which provides an extensive analysis of a number of EF tasks, detailing which tasks are most reliable and discriminatory in preschool children and distinguishing between different difficulty levels (e.g., demonstrating that preschoolers find the Monkey–Crocodile inhibition task easier than Simon Says). We chose three different EF tasks that primarily measure three core abilities of EF (e.g., Best & Miller, 2010; Hughes, 1998): the Dimensional Change Card Sort (DCCS) task that primarily measures cognitive flexibility, the Forward Digit Span (FDS) task that primarily measures working memory, and the Monkey–Crocodile/Simon Says (Monkey–Croc) task that primarily measures inhibition. (Note that no single task can ever be a pure test of a particular EF.)

In older children and adults, it has been argued that the FDS is not a working memory task because it can be passed using basic recall unless a distractor is introduced. This is not the case for young preschoolers, for whom other tasks are also too difficult (e.g., the Backward Digit Span [BDS] includes inhibition as well as working memory demands; Carlson, 2005). However, we added two additional working memory tasks to the later age points; the BDS from the British Ability Scales-3 was administered at 43 and 49 months, and the Corsi blocks task described in Farrell Pagulayan et al. (2006) was administered at 49 months. Neither was administered at 37 months because these tasks are too difficult for children that young (and in fact the results for the BDS at 43 months were at floor; see later).

The tasks were always presented in the same order for all children at all age points: FDS → BDS (at 43 and 49 months) → Monkey–Croc → DCCS → Corsi blocks (at 49 months). All items were scored online and were checked offline (using video-recordings of the session) where necessary.

Cognitive flexibility (shifting): DCCS. In this task, children are required to place cards into one of two boxes based on a specific feature of the image on the card. Over the course of the task, the sorting rule changes, requiring children to shift their response accordingly. We followed the procedure outlined by Zelazo (2006) in which test trials are administered at levels of increasing difficulty, with each level introducing a new more difficult change to the sorting rule. Children were introduced to two boxes; one had a card with a blue rabbit attached, and one had a card with a red boat attached. The experimenter first explained that they were going to play a color game in which they were going to sort the cards into the two boxes by their color (red or blue). Children first completed two practice trials, one for each color. At Level 1 (6 trials), the experimenter presented a card to children, described it using the color feature, and asked children to place it in one of the boxes (e.g., "Here's a blue one. Where does it go?"). After all 6 Level 1 trials, the experimenter introduced the first rule change (Level 2), explaining that they now needed to sort the images by shape (boat or rabbit). This was followed by 6 Level 2 test trials where the experimenter presented a card to children, described it using the shape feature, and asked children to place it in one of the boxes (e.g., "This is a rabbit. Where does it go?"). At the 43- and 49-month data points, the experimenter introduced the third, more complex, rule change in which the cards were to be sorted according to the presence or absence of a star on the card; cards with a star were sorted by the color feature, and cards without a star were sorted by the shape feature. Again, 6 Level 3 test trials were administered. An item was considered correct if children placed the card in the correct box.

Working memory: FDS. We administered the FDS task from the British Ability Scales-3, a standardized test with good validity and reliability (Elliot & Smith, 2012). The task consisted of eight blocks of digits that the experimenter read out loud and then asked children to repeat (36 items in total). The items increased in difficulty in eight levels from Level 1 (in which items were two digits long) to Level 8 (nine digits long). The experimenter first used a puppet to demonstrate how to play the game and then asked children to play the game like the puppet. The experimenter said "ready" prior to each item and read each sequence at a pace of two digits per second with a drop in intonation for the last digit. The test was administered with a basal rule (administer first item in each block until this item is failed, then go back and administer items in previous block) and a stopping rule (when basal is established, continue forward on all administered items until children complete a block with no more than one pass). Items were marked correct if children repeated all digits in the correct order.

Working memory: BDS. The BDS task was identical to the FDS task, but children were asked to repeat the digits in reverse order. The task consisted of 25 items in five blocks that increased in difficulty from Level 1 (items were two digits long) to Level 5 (eight or nine digits long). Items were marked correct if children repeated all digits in the correct reverse order. The task proved to be too difficult for the 43-month-olds (all children scored 0 except for 1 child who accurately repeated only 1 item), so we report the results from the BDS at 49 months only.

Working memory: Corsi blocks. The Corsi blocks task (Farrell Pagulayan et al., 2006) is a test of nonverbal (visuospatial) working memory. Nine identical wooden blocks are placed in front of children in a grid. The experimenter taps/points to certain blocks in order, and children are asked to point to the blocks in the same order. The task consisted of 45 items that increased in difficulty in nine levels from Level 1 (5 items in which one block was tapped) to Level 9 (5 items in which nine blocks were tapped). The test was administered with a basal rule (administer first item in each block until this item is failed, then go back and administer items in previous block) and a stopping rule (when basal is established, continue forward on all administered items until all items in a block are failed and then stop). Items were marked correct if children tapped all blocks in the correct order.

Inhibition: Monkey–Croc. This task was the modified version of the graded bear–dragon task described by Meuwissen and Carlson (2015). It requires children to inhibit a response in some cases but not in others. The test contained 80 items in eight levels of increasing difficulty, with difficulty determined by the performance data from preschoolers reported in Meuwissen and Carlson (2015). There were 10 trials in each level, 5 action trials and 5 inhibition trials, although only inhibition trials were scored. First, in a warm-up phase, the experimenter checked that children were comfortable with all the actions required (e.g., "Touch your nose"). Then the experimenter went through the eight levels. Each level began with practice trials, where incorrect responses were corrected up to four times, followed by test trials. In test trials, children were given two attempts to follow each instruction, with the experimenter waiting 4 s for a response.

The initial four levels of the task were administered with the support of two puppets, a "nice" monkey and a "not very nice" crocodile. Children were encouraged to do what the monkey said and not to do what the crocodile said (inhibit the response). The task became harder throughout the levels (e.g., at Level 1 all the action trials were presented first, followed by the inhibition trials, whereas at Level 4 the inhibition and action trials were presented in turn). Levels 5 to 8 followed the traditional Simon Says game format in which children needed to follow only a verbal prompt, which is harder for preschool children than the Monkey–Croc task (Carlson, 2005). They were told to follow instructions that started with "Simon says" but to not follow (inhibit the response) when the experimenter did not say "Simon says." Again, the task became more difficult as the children advanced through the levels (e.g., at Level 5 the experimenter provided both verbal and visual cues, whereas at Level 8 only the verbal cue was provided). Only inhibition responses were scored; items were marked correct if children correctly inhibited their response.

Scoring. The literature reports a range of different practices for scoring preschool children's performance on EF tasks (e.g., highest level achieved, raw score, scaled score). Because our tasks were designed in levels of increasing difficulty, we used a proportional scaled score that credited children with higher scores for passing items at more difficult levels. For each correct item, the raw score was multiplied by the level at which it was administered. For example, items at Level 1 were multiplied by 1, and items at Level 3 were multiplied by 3. Then we calculated proportional scaled scores (scaled score / total possible scaled score) for each child for each task. Using proportional scores facilitates direct comparison across tests across ages (the tests contained very different total numbers of items, and for the DCCS task we administered only two of the three possible levels to the 37-month-olds).

Sentence processing task

This task was a replication of Borovsky et al. (2012) in which children's eye movements were tracked using an eye tracker while the children saw four pictures on a screen and listened to sentences. For example, children might hear "The pirate hides the treasure" and see images of the *treasure* (target), a *ship* (agent-related foil), *bones* (action-related foil), and a *cat* (unrelated foil). Eye movements were analyzed in the period following the verb and before the onset of the final noun (*treasure*).

Speech stimuli. At all three age points, we presented the same 32 sentences used by Borovsky et al. (2012; see Table S1 in the online supplementary material for a full list of sentences). The sentences were recorded by a female native speaker of British English whose local accent would be familiar to all participants and who spoke with a child-directed intonation. Recordings were made using Audacity and were processed using Praat (Boersma, 2001). Sentences were modified so that the onsets were matched for the first article (2000 ms), agent word (2220 ms), action word (3180 ms), second article (4387 ms), and target word (4480 ms). Pitch and volume were normalized. Two counterbalance orders were created, each containing 16 of the 32 possible sentences, to ensure that no agents or verbs were repeated within participants. Participants were randomly assigned to one counterbalance order at each age point.

Visual stimuli. The images were prototypical exemplars of each object, one shown in each quarter of the screen. Images depicted the target, an agent-related distractor, an action-related distractor, and an unrelated distractor. Images were located in the corner of each quarter on a white background measuring 400 × 400 pixels.

Procedure. During the experiment, children typically sat in a car seat with their caregiver standing behind them or on their caregiver's lap. They were positioned approximately 60 cm away from a 17-inch LCD monitor attached to an adjustable arm mount. An EyeLink 1000+ eye tracker (SR Research: Ottawa, Ontario, Canada) was used to present the stimuli and record eye movements. The speech stimuli were presented in stereo from speakers situated on either side of the monitor.

During setup, children watched a short animated cartoon on the screen with engaging music but no speech, which was followed by a manual 5-point calibration in which children saw a looming black and white circular shape accompanied by a twinkly sound. Following successful calibration, a centrally positioned gaze-contingent attention getter appeared on the screen. This helped to minimize trial loss due to inattention (Delle Luche et al., 2015). After a continuous fixation of 400 ms, a trial began. Caregivers were instructed to engage minimally with their children and to not name the images or talk during the task unless necessary (e.g., if their children were distressed).

Each trial consisted of a display of four images located in the four corners of the screen presented for 8000 ms and an audio-recorded sentence time-locked to begin 2000 ms after the images appeared. Participants viewed 16 trials, with each quartet of images presented twice, each time with a different agent and verb included in the sentence. During presentation, the experimenter recorded, for each trial, whether a trial was usable or unusable. Trials were coded as unusable if there was parental interference or audible disruption during any part of the sentence (e.g., talking, crying, distortion to the sound). During the lab visit, parents completed a checklist containing all targets and foils presented during the task to indicate which of these words their children understood.

Scoring. Eye movements were recorded using the EyeLink 1000+ eye tracker in remote mode, where head and eye movements are monitored using a target sticker positioned on the forehead of participants prior to calibration. This provides stable eye tracking for child participants in that it allows for some movement of the head in relation to the camera. Eye movements were recorded for the full duration of each trial, and the default settings of the EyeLink system were used to identify saccades, fixations, and blinks. Interest areas were added using DataViewer, the custom software designed for working with the EyeLink eye trackers, and captured the full area covered by each image. Data were extracted at the sample level, with one sample every 10 ms, and included information about the interest area in which the gaze was located for each sample. Samples were included only if they formed part of a continuous fixation lasting 100 ms or longer (Casillas & Frank, 2017). Shorter fixations were not included in analyses of looking time. Following fixation identification, individual trials were excluded according to the following criteria applied in this order: (a) at least one of the key words (*agent, action,* or *patient*) included in the sentence was not understood by children based on the checklist completed by parents; (b) the trial was marked as unusable by the experimenter during the testing session; and (c) there were no fixations in any of the interest areas for the duration of the trial. Participants were excluded if they failed to provide looking time data for 50% of total trials (i.e., for 8 trials). After application of the inclusion criteria, we retained data from 63 participants at 37 months, 65 participants at 43 months, and 59 participants at 49 months.

We included looks during a time window that began 300 ms after the onset of the verb and ended at the onset of the target word. The 300-ms delay ensured that only shifts in response to the speech stimuli were included in the analyses (Swingley et al., 1999). During this time window, children have had access to sufficient information to accurately identify the target but had not yet heard it labeled.

Integration ability was calculated in two ways: (a) the proportion (target / target + distractors) of time spent looking at the target image after the onset of the verb (+300 ms) but before the onset of the target noun (logit transformed) and (b) the latency of initial saccades (i.e., speed to orientate) to the target image after the onset of the verb (+300 ms) but before the onset of the target noun (log transformed). The latency score can be calculated only on trials where children are not looking at the target image at the onset of the time window of interest. Following the speed of processing literature (e.g., Peter et al., 2019), we only included data from participants with at least 2 trials where a latency score could be calculated.

Maintenance ability was calculated as the proportion of time (logit transformed) spent looking at the agent-related image after the onset of the verb (+300 ms) but before the onset of the target noun, expressed as a proportion of time looking at all distractors (excluding target; i.e., agent-related distractor / all distractors).

Analysis

All analyses were conducted in R Version 4.0.4 (R Core Team, 2021). All outliers were included unless it was determined that the data point reflected experimenter or participant error (i.e., an impossible value such as a score of 80 on a 70-point scale). All participants who provided at least one data point for the EF battery, BPVS, and sentence processing task were included.

A preregistration of our analysis and analysis code, data, and output are available on our Open Science Framework (OSF) project page (https://osf.io/zqmf2/?view_only = 8fb7ce33f7fb4 f2aa9d485b8db433d53). For full disclosure, the current article includes some preliminary growth curve analyses that were not preregistered but were added in response to reviewer comments as well as some exploratory analyses. Note too that we preregistered some ancillary analyses (a conceptual replication of Borovsky et al.'s (2012) main finding of vocabulary and age predicting integration scores and a correlation between scores on our EF battery and on the Behavior Rating Inventory of Executive Functions, Preschool Version [BRIEF-P]) that we do not present here.

Results

We present results in the following order: (a) preliminary analyses, (b) effect of EF on vocabulary size and growth over time, and (c) results testing our primary hypothesis that sentence processing ability (integration and maintenance) mediates the relationship between receptive vocabulary and EF. We finish with (d) exploratory analyses. The code for all analyses can be found on the OSF project page https://osf.io/zqmf2/).

Preliminary analysis

Table 1 shows the descriptive statistics for the EF and vocabulary measures, illustrating the mean (and standard deviation) proportional scaled scores for the EF tasks and the raw mean (and standard deviation) scores for the vocabulary task. Fig. 1 shows the scores for vocabulary and the three EF tasks for which we have data at all three age points in graphical form, demonstrating age-related changes. Fig. 2 shows the scores for the three measures taken from the sentence processing task across age, illustrating the mean proportion of anticipatory looks to the target image (Integration measure 1) and agent-related distractor (Maintenance measure) and the mean latency to the target after the onset of the verb (Integration measure 2).

Growth over time

We used growth curve modeling to explore age-related changes in vocabulary, in sentence processing, and in the EF tasks for which we had data at three age points in more detail. In particular, we were concerned with determining whether task performance was best fitted by a linear or quadratic age

Table 1
Mean (and standard deviation) proportional scaled scores on the EF battery and raw scores on the vocabulary (BPVS) task at each age point

Age point	Cognitive flexibility	Working memory			Inhibition	Vocabulary
	DCCS	FDS	BDS	Corsi blocks	Monkey–Croc	BPVS
37 months	0.53 (0.27)	0.13 (0.07)	–	–	0.18 (0.18)	48.3 (14.6)
43 months	0.61 (0.16)	0.17 (0.08)	0.0002 (0.002)		0.36 (0.22)	58.6 (13.3)
49 months	0.66 (0.16)	0.20 (0.09)	0.01 (0.03)	0.08 (0.04)	0.51 (0.17)	70.4 (13.5)

Note. EF, executive function; BPVS, British Picture Vocabulary Scale–Third Edition; DCCS, Dimensional Change Card Sort task; FDS, Forward Digit Span task; BDS, Backward Digit Span task; Monkey–Croc, Monkey–Crocodile/Simon Says task.

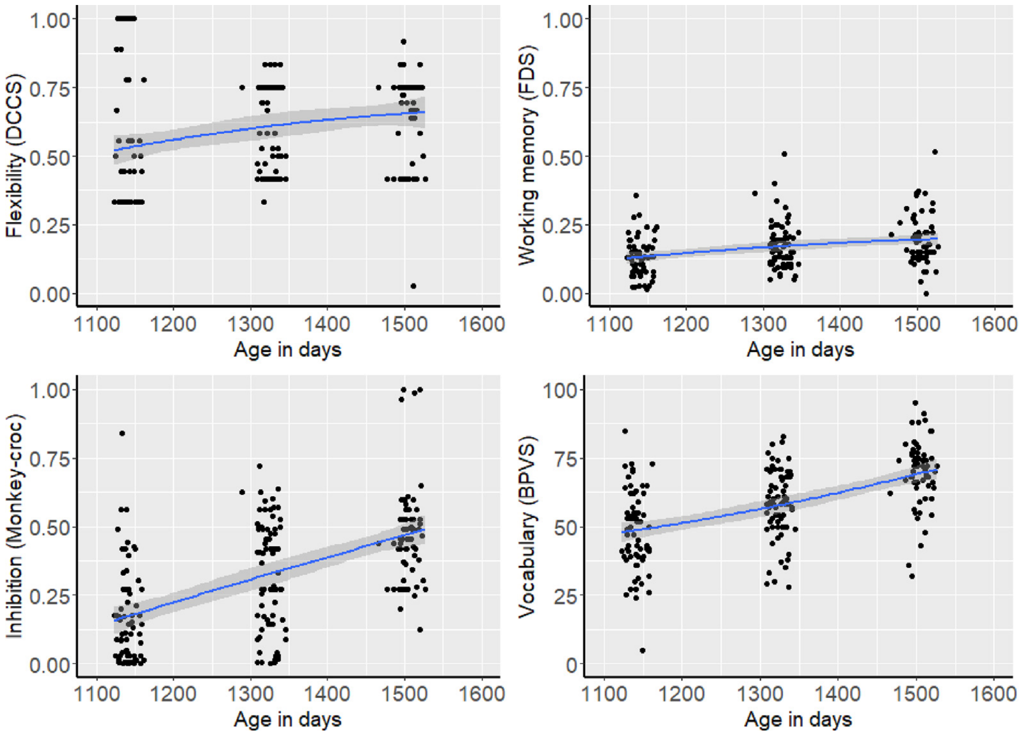

Fig. 1. Longitudinal results from three executive function (EF) tasks and the vocabulary task by age. Age is show in days (the first set of points represents data taken at the 37-month age point, the second at 43 months, and the third at 49 months). Each point indicates performance by 1 child. The line is a polynomial regression line with the 95% confidence interval band indicated in gray. The figure shows proportional scaled scores for the EF tasks and raw scores for vocabulary. DCCS, Dimensional Change Card Sort task; FDS, Forward Digit Span task; Monkey–Croc, Monkey–Crocodile/Simon Says task; BPVS, British Picture Vocabulary Scale–Third Edition.

curve because this affects the types of analyses we can perform. We ran mixed effects models predicting task performance from linear age and, using analysis of variance (ANOVA), compared them with models that predict task performance from both linear and quadratic age. For all analyses, we used age in days as the predictor to add precision. In each case, we used the nloptwrap optimization algorithm and maximum likelihood (ML) estimation (Zuur et al., 2009). We determined that the quadratic model was a better fit to the data if the Akaike information criterion (AIC) scores for the models including quadratic and linear age were more than 2 units lower than those for the models including only linear age. For the EF tasks and vocabulary, the models that converged included random effects of

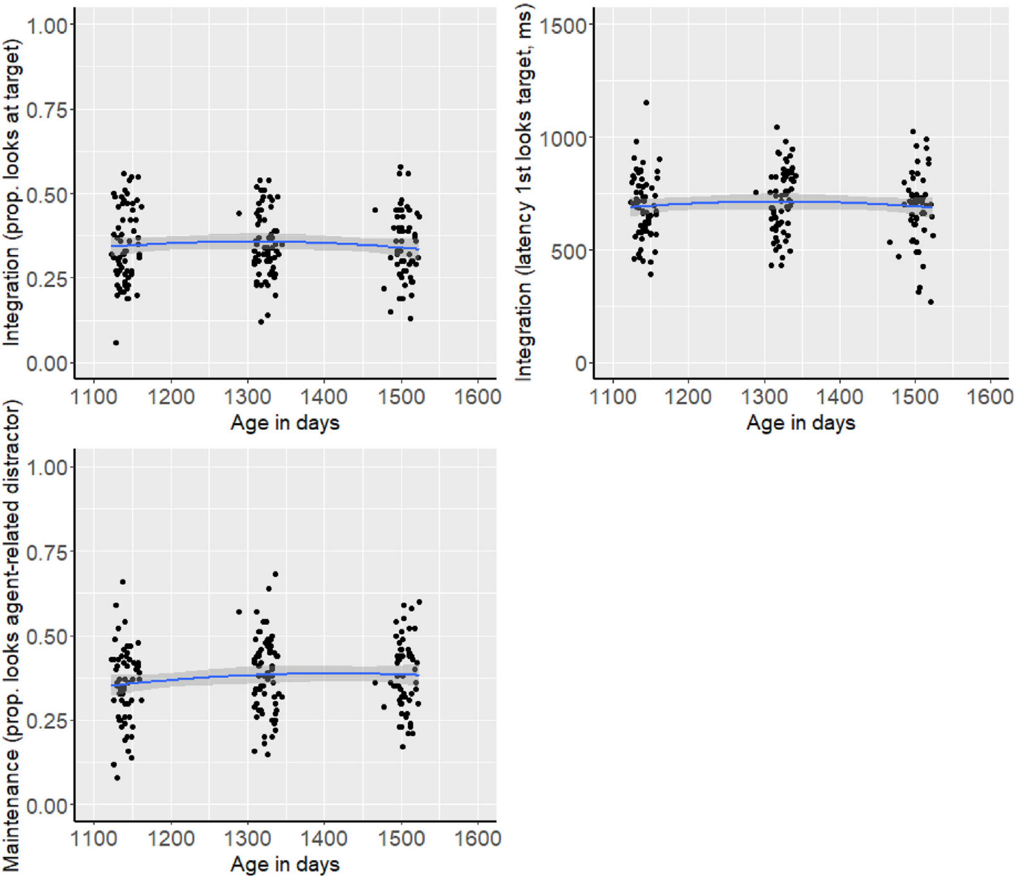

Fig. 2. Results from the sentence processing task: Proportion of looks to the target [Integration (prop.)], latency (speed) of first look at the target [Integration (latency)], and proportion of looks to the agent-related distractor (Maintenance) by age in days. Each point indicates performance by 1 child. The line is a polynomial regression line with the 95% confidence interval band indicated in gray. The figure shows true scores; transformed scores were used in the analysis. (*ns*: 37 months = 63, 43 months = 65, 49 months = 59.)

participant and random slopes for linear age by participant. For the sentence processing tasks, the models that converged included only random effects of participant, but note that the integration models also yielded singular fits (see below).

Full results of the best fitting models (linear or quadratic) for EFs and vocabulary are presented in Table 2 (see also Fig. 1). For all EF tasks and for the BPVS vocabulary task, there was a main effect of linear age, suggesting that performance improved between 37 and 49 months. For the FDS, Monkey–Croc, and BPVS, the best fitting model included only linear age, suggesting steady growth over time. For the DCCS task, although adding quadratic age did not explain significantly more variance, the model including both linear and quadratic age was a better fit to the data (AIC linear age = −81.52, AIC linear + quadratic age = −92.52). From Fig. 1, we can see that this is because the rate of growth slows down marginally over time.

For the three sentence processing measures, none of the models yielded significant effects of either linear or quadratic age (see Table 3 and Fig. 2), indicating that there was no improvement in performance with age. However, for the two integration measures, even the simplest models yielded singular fits, suggesting overfitting. Thus, to confirm the results, we also ran fixed effects models

Table 2
Results of the models predicting flexibility (DCCS), working memory (FDS), inhibition (Monkey–Croc), and vocabulary (BPVS) from linear age

	Flexibility (DCCS)	Working memory (FDS)	Inhibition (Monkey–Croc)	Vocabulary (BPVS)
(Intercept)	0.60***	0.16***	0.32***	58.64***
	[0.56, 0.63]	[0.15, 0.18]	[0.28, 0.36]	[56.33, 61.92]
Linear age	0.76***	0.39***	1.78***	129.79***
	[0.38, 1.12]	[0.28, 0.51]	[1.48, 2.07]	[116.87, 144.52]
Quadratic age	−0.12			
	[−0.47, 0.19]	NA	NA	NA
Number of observations	211	208	206	208
R^2 conditional	.42	.69	.58	.80
R^2 marginal	.06	.10	.31	.29
Akaike information criterion	−82.7	−512	−118	1588
Bayesian information criterion	−59.2	−492	−97.6	1608

Note. Quadratic age is included only for Dimensional Change Card Sort task (DCCS) because it did not add additional variance in the other models. Bootstrapped 95% confidence intervals are in brackets. FDS, Forward Digit Span task; Monkey–Croc, Monkey–Crocodile/Simon Says task; BPVS, British Picture Vocabulary Scale–Third Edition; NA, not available.
***$p < .001$.

Table 3
Results of the models predicting integration (proportion), integration (latency), and maintenance from linear age

	Integration (proportion looks to target)	Integration (latency)	Maintenance (proportion looks to agent-related distractor)
(Intercept)	-0.67***	6.53***	−0.56
	[−0.73, −0.60]	[6.49, 6.55]	[−0.64, −0.47]
Linear age	−0.07	−0.05	0.95
	[−1.07, 1.00]	[−0.51, 0.40]	[−0.13, 1.74]
Number of observations	192	186	192
R^2 conditional	NA!	NA!	.31
R^2 marginal	.00009	.0003	.02
Akaike information criterion	282	−12	290
Bayesian information criterion	295	0.89	310

Note. Quadratic age is not included because it did not explain additional variance. Bootstrapped 95% confidence intervals are in brackets. !, random effect variances not available due to singular fit.
***$p < .001$.

(repeated-measures ANOVAs; as suggested by Oberpriller et al., 2022, in the case of singular fits). These too yielded no significant effect of age (integration [proportion looks]: $F = 0.94$, $df = 2{,}92$, $p = .94$; integration [latency]: $F = 2.38$, $df = 2{,}86$, $p = .10$). Given that all tasks show strong linear effects of age, with a small effect of quadratic age only for the DCCS task, all subsequent analyses used linear models.

Confirmatory factor analysis

Because there is doubt about whether EF represents a unitary or multicomponent ability in preschoolers, we preregistered CFAs to determine whether performance on the EF tasks can be best described as representing an underlying unitary latent EF ability or three separate but related components. We ran CFAs for the three tasks for which we had longitudinal data both at separate age points and longitudinally across age (analysis output can be found on the project OSF page: https://osf.io/zqmf2/). Unfortunately, the age-specific analyses were saturated and the longitudinal analysis yielded error messages that indicated the model had not properly converged, both of which meant that we could not extract overall fit statistics from any of the models. However, the results suggested that there was no good evidence that the data were best represented by a unitary factor (e.g., very high correlations within tasks across age and low correlations across tasks within age, latent factors explained very little of the variance of performance on each task; see Willoughby et al., 2014, for a similar conclusion). As a result, we ran separate analyses for each component task of the EF battery.[1]

Effect of EF on vocabulary

The hypothesis that processing speed will mediate the relation between EF and vocabulary assumes that there is a relationship between EF and vocabulary to be explained. To check whether this was the case, for those EF tasks for which we had data from three age points (cognitive flexibility, working memory [as measured by the FDS], and inhibition), we ran linear mixed effects models. Centered proportional scaled EF score and centered age (in days) were predictors (log age in days for the FDS due to a scaling warning), and raw vocabulary (BPVS) score was the outcome variable. The models that converged included random effects of participant. For the BDS and Corsi blocks, for which we had reliable scores only at 49 months, we ran linear regressions, with centered proportional scaled scores as predictors and raw vocabulary score as the outcome variable.

Results are presented in Tables 4 and 5. There were significant positive effects of cognitive flexibility, working memory as measured by the BDS at 49 months, and inhibition (Monkey–Croc) on vocabulary. There was no effect of working memory as measured by the FDS or Corsi blocks task. For those tasks for which we had longitudinal data, there were main effects of age but no interactions between EF and age.

Mediation analyses

We tested whether the relationship between EF abilities and receptive vocabulary was mediated by sentence processing abilities. We predicted that the size of the direct effect of a particular EF ability (cognitive flexibility, working memory, or inhibition) on vocabulary would be significantly reduced when the mediator processing variable (integration or maintenance ability) was included in the model. Mediation analysis is appropriate only if the predictor has a significant effect on both the outcome variable (Step 1) and the mediator (Step 2) and if the mediator has a significant effect on the outcome variable (Step 3).

Step 1

We have already established that cognitive flexibility (DCCS), verbal working memory (as measured by the BDS), and inhibition are significant predictors of vocabulary (Step 1). Thus, for these variables, we moved on to Step 2. Working memory (as measured by the FDS and Corsi blocks tasks) did not have an effect on vocabulary (see analyses above), so we do not continue with these variables.

Step 2

Step 2 was to determine whether there was an effect of EF performance on the mediator (sentence processing) variables. For all analyses except BDS, we ran mixed effects models. For the BDS, we ran linear regressions because we had only 49-month data. We ran all models using proportional scaled EF score and centered age in days (for all but BDS) as predictors and using logit/log transformed integration/maintenance as outcome variables. Interactions with age were not included in order to simplify the models.

Table 6 illustrates the effect of linear mixed effects models analyzing the effect of cognitive flexibility and inhibition on the sentence processing scores, and Table 7 indicates the effect of a linear regression model analyzing the effect of BDS on sentence processing scores. There were no significant effects of DCCS or BDS on any of the three sentence processing measures. There was also no effect of

[1] We originally included the scores of a parent report EF instrument (the BRIEF-P) in the CFA analyses but removed them at the request of a reviewer. Those analyses converged, and the overall fit statistics also did not point to a unitary latent variable.

Table 4
Results of the models predicting vocabulary from age in days and flexibility (DCCS; Model 1), working memory (FDS; Model 2), and inhibition (Monkey–Croc; Model 3)

	1. Cognitive flexibility (DCCS)	2. Working memory (FDS)	3. Inhibition (Monkey–Croc)
(Intercept)	58.69*** [55.91, 61231]	58.63*** [55.89, 61.51]	58.52*** [55.82, 61.54]
EF task	9.79* [2.14, 17.15)	15.14 [−9.85, 38.81]	8.51* [0.05, 16.76]
Age in days	0.06*** [0.05, 0.07]	79.95*** [69.89, 89.41]	0.06* [0.05, 0.07]
EF*Age in days	0.01 [−0.03, 0.05]	−55.04 [−188.84, 91.73]	−0.008 [−0.05, 0.04]
Number of observations	208	205	203
R^2 conditional	.78	.77	.77
R^2 marginal	.31	.30	.30
Akaike information criterion	1600	1551	1567
Bayesian information criterion	1620	1571	1586

Note. Bootstrapped 95% confidence intervals are in brackets. DCCS, Dimensional Change Card Sort task; FDS, Forward Digit Span task; Monkey–Croc, Monkey–Crocodile/Simon Says task; EF, executive function.
*p < .05.
***p < .001.

Table 5
Results of the linear regression models predicting vocabulary from Backward Digit Span (verbal working memory) and Corsi blocks (nonverbal working memory) tasks

	Backward Digit Span	Corsi blocks
(Intercept)	70.77*** [67.81, 73.60]	70.61*** [67.32, 73.89]
EF task	224.64*** [62.21, 371.95]	34.09 [−44.92, 113.04]
Multiple R^2	.20	.011
Adjusted R^2	.18	−.005
F	15.10	0.70
df	1,62	1,62
p	.0003	.41

Note. Bootstrapped 95% confidence intervals are in brackets. EF, executive function.
***p < .001.

inhibition (Monkey–Croc) on either integration measure. Therefore, mediation analysis is not appropriate and we do not continue with these variables to Step 3. (Note that for the two integration measures, a singular fit error indicated that the models may be overfitted, so we also ran fixed effects models at each age [correlations in this case given that our EF measures are continuous] to confirm the lack of association between DCCS/inhibition and integration. All correlations were nonsignificant (all *R*s < .26, all *p*s < .05].) There was, however, a significant effect of inhibition (Monkey–Croc) on maintenance, suggesting that mediation analysis may be appropriate in this case.

Step 3

Step 3 of a mediation analysis is to check whether the mediator has a significant effect on the outcome variable. We ran a mixed effects model with BPVS raw score as the outcome measure and proportional scaled inhibition (Monkey–Croc) score and logit transformed and centered maintenance scores as predictors. We do not include interaction terms for simplicity. The model that converged included a random effect of participant. Table 8 illustrates the results. There was a significant effect of maintenance on vocabulary, meaning that mediation analysis is appropriate.

Table 6
Results of the models predicting integration (proportion), integration (latency), and maintenance from cognitive flexibility and inhibition

	Integration (proportion looks to target)	Integration (latency)	Maintenance (proportion looks to agent-related distractor)
Cognitive flexibility (DCCS)			
(Intercept)	−0.66***	6.53***	−0.52***
	[−0.75, −0.59]	[6.49, 6.56]	[−0.61, −0.46]
DCCS	0.18	0.007	0.08
	[−0.12, 0.58]	[−0.16, 0.18]	[−0.21, 0.45]
Age in days	−0.00004	−0.00002	0.0003
	[−0.0006, 0.0005]	[−0.0003, 0.0001]	[−0.00004, 0.0009]
Number of observations	185	185	185
R^2 conditional	NA!	NA!	.16
R^2 marginal	.006	.0002	.01
Akaike information criterion	287	9.5	281
Bayesian information criterion	288	25.6	297
Inhibition (Monkey–Croc)			
(Intercept)	−0.65***	6.53***	−0.52***
	[−0.73, −0.55]	[6.50, 6.56]	[−0.60, −0.44]
Monkey–Croc	0.25	−0.09	0.41*
	[−0.17, 0.69]	[−0.35, 0.013]	[0.03, 0.82]
Age in days	−0.00004	−0.00001	0.0003
	[−0.0005, 0.0004]	[−0.0002, 0.0003]	[−0.0001, 0.0007]
Number of observations	181	181	181
R^2 conditional	NA!	NA!	.17
R^2 marginal	.009	.006	.04
Akaike information criterion	282	7.69	270
Bayesian information criterion	298	23.7	286

Note. Bootstrapped 95% confidence intervals are in brackets. DCCS, Dimensional Change Card Sort task; Monkey–Croc, Monkey–Crocodile/Simon Says task; !, random effect variances not available due to a singular fit error.
$^{*}p < .05$.

Table 7
Results of the linear regression models predicting integration and maintenance from Backward Digit Span (verbal working memory) at 49 months

	Integration (proportion looks to target)	Integration (latency)	Maintenance (proportion looks to agent-related distractor)
(Intercept)	−0.69***	6.50***	−0.48***
	[−0.82, −0.56]	[6.43, 6.57]	[−0.60, −0.36]
Backward Digit Span	0.96	1.73	−2.21
	[−3.03, 5.19]	[−0.58, 4.36]	[−5.44, 0.81]
Multiple R^2	.003	.03	.02
Adjusted R^2	−.02	.02	.00005
F	0.16	1.89	1.003
df	1, 54	1,54	1, 54
p	.69	.18	.32

Note. Bootstrapped 95% confidence intervals are in brackets.
$^{***}p < .001$.

Table 8
Results of the mixed effects model predicting |vocabulary from inhibition (Monkey–Croc) and maintenance scores

	Vocabulary
(Intercept)	58.84***
	[55.99, 61.71]
Inhibition (Monkey–Croc)	5.16
	[−4.31, 14.52]
Maintenance	3.64*
	[0.28, 6.61]
Age in days	0.06***
	[0.05, 0.07]
Number of observations	181
R^2 conditional	.77
R^2 marginal	.30
Akaike information criterion	1393
Bayesian information criterion	1412

Note. Monkey–Croc, Monkey–Crocodile/Simon Says task.
*$p < .05$.
***$p < .001$.

Mediation

We used the R package "mediation" Version 4.5.0 (Tingley et al., 2014) to run mediation models to calculate whether the mediator (maintenance in this case) had a significant mediating effect on the relationship between inhibition (Monkey–Croc) and vocabulary (BPVS score). We preregistered a decision to use one-tailed *p* values for our mediation analysis because our hypothesis predicted unidirectional effects (Cho & Abe, 2013). As Fig. 3 illustrates, the regression coefficients between inhibition and vocabulary (0.41) and between maintenance and vocabulary (3.64) were significant, but the direct effect of inhibition on vocabulary was not once maintenance was taken into account (5.16). The indirect effect was (0.41)*(3.64) = 1.49. Unstandardized indirect effects were computed for each of 1000 bootstrapped samples, and the 95% confidence interval was computed. The bootstrapped unstandardized indirect effect was 1.49, and the 95% confidence interval ranged from −0.03 to 3.84. The indirect effect was significant and one-tailed ($p = .031$). In sum, the effect of inhibition on BPVS score was at least partially mediated by the maintenance score.

Exploratory analyses

In sum, we found only one significant preregistered mediating relationship: Maintenance ability partially mediated the relation between inhibition and vocabulary. For all other analyses, mediation was not appropriate because there was no relationship either between EFs and vocabulary or between EFs and sentence processing ability. To explore this in more detail, we ran correlations to determine (a) whether the results of our EF battery were stable over time (i.e. across development) for those tasks for which we had data at all three age points (37, 43, and 49 months; see Fig. 4) and (b) whether the results of our sentence processing tasks were stable over the time period (37, 43, and 49 months; see Fig. 5). For the EF battery, the key results are the correlations of each EF task with itself over time (highlighted by shaded boxes). There were small or medium correlations across age for all three tasks, suggesting that the tasks were assessing somewhat stable constructs across age. However, for the sentence processing scores (see Fig. 5), there were very few significant correlations over time (highlighted by blue boxes); only the correlation between maintenance scores at 37 and 43 months was significant ($r = .22$, $df = 57$, $p = .02$). Note too that the correlations between the two integration scores (proportion looking at target and latency of first look at target; highlighted by yellow box in Fig. 5), which should be measuring the same underlying ability, yielded only medium negative correlations that reached conventional levels of significance at 37 and 43 months (37 months: $r = -.39$, $df = 61$, $p = .001$; 43

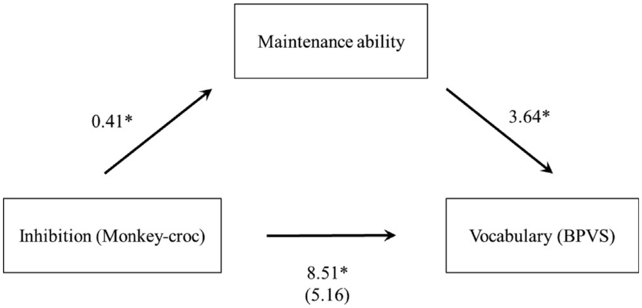

Fig. 3. The mediation model predicting vocabulary from inhibition (Monkey–Croc) with mediation from maintenance ability. Numbers represent beta estimates taken from the tables above. The beta estimates between inhibition and vocabulary reflect whether the mediating factor (maintenance ability) was (in parentheses) or was not included in the model. $^{*}p < .05$. Monkey–Croc, Monkey–Crocodile/Simon Says task; BPVS, British Picture Vocabulary Scale–Third Edition.

months: $r = -.41$, $df = 63$, $p = .001$) but not at 49 months. In sum, the sentence processing results do not seem to be stable over time.

Discussion

A wealth of research has found a consistent relationship between EF and language during the preschool years, but the causal mechanism underlying this link has remained elusive. We tested a novel question in the current experiment: Is sentence processing ability one mediating variable that explains the connection between EF and receptive vocabulary? We found that this was the case in only one of our analyses.

We employed a battery of behavioral EF tasks, a standardized test of receptive vocabulary (BPVS), and an eye-tracking-based sentence processing task (Borovsky et al., 2012) in which we measured two sentence processing abilities (integration and maintenance). We replicated previous findings of a relationship between EF and vocabulary for cognitive flexibility, working memory (as measured by the BDS), and inhibition.

Using mediation models, we then sought to determine whether there was a relationship between EF and vocabulary, with sentence processing abilities as a mediating variable. We hypothesized that sentence processing abilities might recruit EF mechanisms, such as working memory, inhibition, and flexibility, and that vocabulary growth in turn relies on such sentence processing abilities to comprehend the incoming and ephemeral speech input. However, for there to be grounds for a mediation analysis, there must be a significant relationship both between the predictor and the outcome and between the predictor and the mediator variable. This was not the case for either integration measure—integration as measured by proportion of looks to the target or integration as measured by latency (speed) of the first look to the target. Given that there were no grounds for a mediation analysis, we can conclude that our prediction that integration mediates the relation between EF and vocabulary was not upheld.

Turning to maintenance ability, there were no grounds for a mediation analysis for the analyses that tested the effects of flexibility and working memory; neither of these EF abilities predicted the mediator (maintenance). However, for inhibition (as measured by the Monkey–Croc task) there were grounds for mediation; there was both an effect of inhibition on maintenance (in that children with higher inhibition scores also looked more at the agent-related distractor in the sentence processing task) and an effect of maintenance on vocabulary (in that children who looked more at the agent-related distractor also had better vocabulary scores). Mediation analysis showed that the relationship between inhibition and vocabulary was partially, but not fully, mediated by maintenance ability. In other words, the children who scored high in a task that required them to inhibit automatic/impulsive responses were also more likely to maintain multiple possible referents in mind when processing sen-

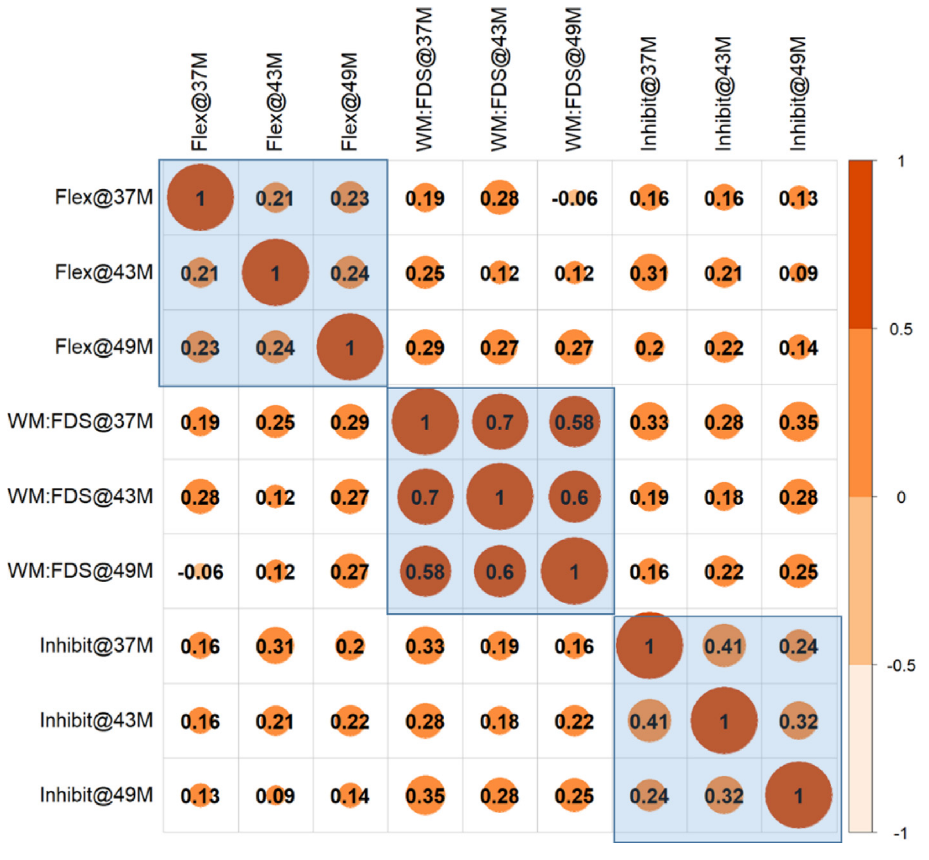

Fig. 4. Correlation matrix illustrating Pearson's *r* values of correlations between executive function tasks across age groups. The shading and size of the circles indicate the strength of the correlation. Flex, cognitive flexibility; WM:FDS, working memory as measured by Forward Digit Span; Inhibit, inhibition; @, age group (37 months, 43 months, 49 months). Scores are proportional scaled scores.

tences, and this relationship partially explained why such children also had bigger vocabularies in our study. However, the direct relationship between inhibition and vocabulary remained substantial in the mediation model, so there remains an effect of inhibition on vocabulary that cannot be explained by maintenance ability.

In sum, only one of our nine sets of analyses yielded the predicted result. Three possible reasons for this are (1) that our EF measures were not sensitive enough to pick up on meaningful individual differences, (2) that our sentence processing measures were not sensitive enough to pick up on individual differences, and (3) that the relationship is more nuanced than stated in the Introduction, in other words, that the relationship between some EFs and vocabulary is mediated by *some* sentence processing abilities. Option 1 is, we think, the least likely explanation. There were small or medium correlations in EF abilities across age, and we replicated the finding from the literature of a relationship between EF and receptive vocabulary in 3- and 4-year-old children, which would have been unlikely with very noisy EF measures.

Option 2, the idea that our sentence processing measures were not sensitive enough, is a more likely explanation. None of our three sentence processing scores correlated highly with themselves across time or with each other, nor did our integration measures predict vocabulary. These results contradict those of Borovsky et al. (2012), who did find a relationship between sentence processing

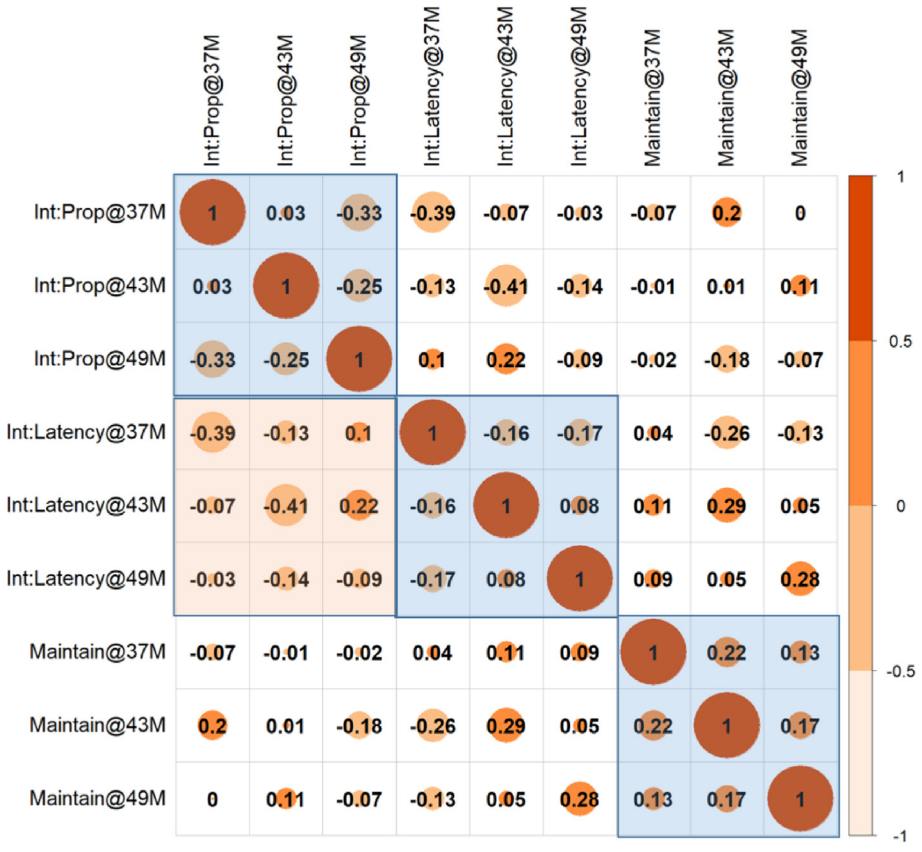

Fig. 5. Correlation matrix illustrating Pearson's *r* values of correlations between sentence processing tasks across age groups. Int:Prop = integration as measured by proportion looking at target, The shading and size of the circles indicate the strength of the correlation. Blue highlighted boxes indicate correlations across the same task over time. Yellow highlighted box indicates correlations between two integration scores (negative correlations predicted). Int:Latency, integration as measured by latency of first look to target; Maintain, maintenance as measured by proportion looking at agent-related distractor; @, age group (37 months, 43 months, 49 months). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and vocabulary in the same task, albeit over a much larger age range (3 years–adult). This discrepancy across studies using the same task speaks to the issue of the reliability of looking time measures when used to assess individual differences, an issue that has been raised in previous work (Byers-Heinlein et al., 2022; Durrant et al., 2021). These previous articles highlighted the need for a better understanding of looking time tasks when used with young children because very little is known about the cognitive mechanisms driving looking behavior in infants and young children (Byers-Heinlein et al., 2022), because looking time measures from children may be very noisy because they are often calculated on very few trials (see Egger et al., 2020, for a possible solution) and because task difficulty needs to be carefully calibrated in order to extract meaningful individual differences (tasks that are too difficult or too easy yield floor and ceiling effects, respectively; Hedge et al., 2018). More information about the strategies children use when responding to different types of stimuli in different kinds of looking time tasks is needed.

That said, Option 3—that the relationship between some EFs and vocabulary is mediated by some sentence processing abilities—remains a plausible explanation. In particular, the relationship between inhibition and maintenance in our study partially mediated the relationship between inhibition and vocabulary. Maintenance—the ability to maintain multiple referents in mind while a sentence unfolds over time—is an important sentence processing skill, especially in cases of complex sentences where initial responses often need to be suppressed (e.g., passive sentences in which the first noun is, unusually for English, the patient—not the agent—of the sentence; e.g., "The dog was bitten by the cat"). Thus, the ease with which 3- and 4-year-old children learn new vocabulary may well depend on their ability to keep a number of possible referents in mind in order to rapidly update incorrect inferences (Diamond, 2013; Weiland et al., 2014), which itself may depend on their ability to inhibit automatic responses.

However, the fact that we did not find mediating relationships in eight of our nine analyses suggests that sentence processing abilities (at least those we have studied here) cannot fully explain the relation between EF and language acquisition. In particular, the idea of a direct relation, rather than an indirect relation, between EF skills and language acquisition remains open. For example, it is possible that language supports the development of EF directly through the creation of rules and symbols that allow children to build internal representations. Once children label a referent, this may enhance their ability to attune and direct their attention to that referent and may improve their ability to monitor their own explicit goal-directed behavior toward that referent (Kuhn et al., 2014). Better language ability may provide a better foundation for earlier and/or faster developing EF abilities while, simultaneously, better EF abilities may underpin more accurate sentence processing and thus support faster language acquisition.

Limitations

The current study has a number of limitations. First, although we were careful to choose EF tasks that both were suitable for 3- and 4-year-old children and measured core abilities of EF, one of these—the FDS—has been argued to reflect recall rather than working memory limitations (albeit in older children and adults). We added additional working memory measures at the older ages, but these were not wholly successful with children this young. Future studies should explore the effect of other working memory tasks. Second, all EF tasks were administered in a fixed order, which means that fatigue effects might have influenced performance on the later tasks. This design decision was made in order to minimize individual differences in performance due to testing order, but it does limit comparisons across tasks. Third, although we had multiple data points from a relatively large sample (full $N = 95$), the study was underpowered to detect interactions because we did not have complete data from all participants. Future studies should aim to test bigger samples. Fourth, and finally, the exploratory analyses revealed that the results of the sentence processing tasks were noisy in that they did not show big correlations across time. Given that looking time analyses are widely used across developmental science, it is crucial that we develop more reliable looking time methodologies (see Byers-Heinlein et al., 2022, and Egger et al., 2020, for suggestions).

Conclusion

In this study, we tested the hypothesis that sentence processing abilities might mediate the association between EF skills and receptive vocabulary knowledge in that the speed of language acquisition is at least partially dependent on children's processing ability, which is itself dependent on executive control. We found evidence, consistent with previous research, for a significant association between three EF skills (cognitive flexibility, working memory [as measured by the BDS task], and inhibition) and receptive vocabulary knowledge across this age range. However, only one of the tested sentence processing abilities (the ability to maintain multiple possible referents in mind) significantly mediated this relationship and only for one of the tested EFs (inhibition). We suggest that the relationship between some EFs and vocabulary is mediated by some sentence processing abilities. In particular, children who are better able to inhibit incorrect responses are also better able to maintain multiple possible referents in mind while a sentence unfolds, a sophisticated sentence processing ability that may facilitate vocabulary learning from complex input.

Author contributions

Crystal Lee: conceptualization, formal analysis, writing–original draft, data curation, visualization; Andrew Jessop: conceptualization, formal analysis, writing–review & editing, supervision; Amy Bidgood: conceptualization, methodology, investigation, resources, writing–review & editing; Michelle Peter: conceptualization, methodology, investigation, resources, writing–review & editing; Julian Pine: conceptualization, writing–review & editing, funding acquisition; Caroline Rowland: conceptualization, methodology, formal analysis, resources, data curation, writing–review & editing, visualization, project administration, funding acquisition; Samantha Durrant: conceptualization, methodology, investigation, resources, writing–review & editing, supervision.

Data availability

Code and data are available at https://osf.io/zqmf2/

Acknowledgments

This work was supported by the Economic and Social Research Council (ESRC) International Centre for Language and Communicative Development (LuCiD), funded by the U.K. Economic and Social Research Council (ES/L008955/1 and ES/S007113/1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2023.105693.

References

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

Bishop, D. V. M., Nation, K., & Patterson, K. (2014). When words fail us: Insights into language processing from developmental and acquired disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634). https://doi.org/10.1098/rstb.2012.0403 20120403.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9), 341–345.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005.

Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2022).). Six solutions for more reliable infant research. *Infant and Child Development, 31*(5). https://doi.org/10.1002/icd.2296 e2296.

Camerota, M., Willoughby, M. T., & Blair, C. B. (2020). Measurement models for studying child executive functioning: Questioning the status quo. *Developmental Psychology, 56*, 2236–2245. https://doi.org/10.1037/dev0001127.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616.

Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science, 16*(8), 609–616. https://doi.org/10.1111/j.1467-9280.2005.01583.x.

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language, 92*, 234–253. https://doi.org/10.1016/j.jml.2016.06.013.

Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research, 66*(9), 1261–1266. https://doi.org/10.1016/j.jbusres.2012.02.023.

Delle Luche, C., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the intermodal preferential looking paradigm: Methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development, 40*, 151–172. https://doi.org/10.1016/j.infbeh.2015.05.005.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750.

Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *The British Picture Vocabulary Scale–Third Edition (BPVS 3)*. GL Assessment.

Durrant, S., Jessop, A., Chang, F., Bidgood, A., Peter, M. S., Pine, J. M., & Rowland, C. F. (2021). Does the understanding of complex dynamic events at 10 months predict vocabulary development? *Language and Cognition, 13*(1), 66–98. https://doi.org/10.1017/langcog.2020.26.

Egger, J., Rowland, C. F., & Bergmann, C. (2020). Improving the robustness of infant lexical processing speed measures. *Behavior Research Methods, 52*(5), 2188–2201. https://doi.org/10.3758/s13428-020-01385-5.

Elliot, C., & Smith, P. (2012). *British Ability Scales 3*. GL Assessment.

Farrell Pagulayan, K., Busch, R. M., Medina, K. L., Bartok, J. A., & Krikorian, R. (2006). Developmental normative data for the Corsi block-tapping task. *Journal of Clinical and Experimental Neuropsychology, 28*(6), 1043–1052. https://doi.org/10.1080/13803390500350977.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237–241. https://doi.org/10.1111/j.1467-9280.1996.tb00366.x.

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at Head Start. *Developmental Psychology, 47*(2), 404–416. https://doi.org/10.1037/a0021065.

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education, 8*(3), 259–272. https://doi.org/10.1007/BF03174081.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*(2), 177–194. https://doi.org/10.1348/000709900158047.

Gordon, R., Smith-Spark, J. H., Newton, E. J., & Henry, L. A. (2018). Executive function and academic achievement in primary school children: The use of task-related processing speed. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00582 582.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.

Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry, 53*(1), 37–45. https://doi.org/10.1111/j.1469-7610.2011.02430.x.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459.

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34*(6), 1326–1339. https://doi.org/10.1037/0012-1649.34.6.1326.

Jones, A., Atkinson, J., Marshall, C., Botting, N., St Clair, M. C., & Morgan, G. (2020). Expressive vocabulary predicts nonverbal executive function: A 2-year longitudinal study of deaf and hearing children. *Child Development, 91*(2), e400–e414. https://doi.org/10.1111/cdev.13226.

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *Quarterly Journal of Experimental Psychology, 63*(1), 160–193. https://doi.org/10.1080/17470210902866664.

Kuhn, L. J., Willoughby, M. T., Vernon-Feagans, L., & Blair, C. B. (2016). The contribution of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function. *Journal of Experimental Child Psychology, 148*, 20–34. https://doi.org/10.1016/j.jecp.2016.03.008.

Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., & Blair, C. B. (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Development, 85*(5), 1898–1914. https://doi.org/10.1111/cdev.12249.

Meuwissen, A. S., & Carlson, S. M. (2015). Fathers matter: The role of father parenting in preschoolers' executive function development. *Journal of Experimental Child Psychology, 140*, 1–15. https://doi.org/10.1016/j.jecp.2015.06.010.

Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin & Review, 23*(6), 1942–1953. https://doi.org/10.3758/s13423-016-1068-8.

Oberpriller, J., de Souza Leite, M., & Pichler, M. (2022). Fixed or random? On the reliability of mixed-effects models for a small number of levels in grouping variables. *Ecology and Evolution, 12*(7). https://doi.org/10.1002/ece3.9062 e9062.

O'Neill, G., & Miller, P. H. (2013). A show of hands: Relations between young children's gesturing and executive function. *Developmental Psychology, 49*(8), 1517–1528. https://doi.org/10.1037/a0030241.

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology, 115*. https://doi.org/10.1016/j.cogpsych.2019.101238 101238.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*(2), 73–108. https://doi.org/10.1016/s0010-0277(99)00021-9.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05.

Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience, 29*(3), 312–325. https://doi.org/10.1080/01690965.2013.836231.

Weiland, C., Barata, M. C., & Yoshikawa, H. (2014). The co-occurring development of executive function skills and receptive vocabulary in preschool-aged children: A look at the direction of the developmental pathways. *Infant and Child Development, 23*(1), 4–21. https://doi.org/10.1002/icd.1829.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology, 108*(3), 436–452. https://doi.org/10.1016/j.jecp.2010.08.008.

Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research and Perspectives, 12*(3), 69–95. https://doi.org/10.1080/15366367.2014.929453.

Willoughby, M., Hong, Y., Hudson, K., & Wylie, A. (2020). Between- and within-person contributions of simple reaction time to executive function skills in early childhood. *Journal of Experimental Child Psychology, 192*. https://doi.org/10.1016/j.jecp.2019.104779 104779.

Wittke, K., Spaulding, T. J., & Schechtman, C. J. (2013). Specific language impairment and executive functioning: Parent and teacher ratings of behavior. *American Journal of Speech-Language Pathology, 22*(2), 161–172. https://doi.org/10.1044/1058-0360(2012/11-0052).

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology, 141*, 187–209. https://doi.org/10.1016/j.jecp.2015.08.005.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science, 22*(1). https://doi.org/10.1111/desc.12735 e12735.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297–301. https://doi.org/10.1038/nprot.2006.46.

Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science, 7*(4), 121–126. https://doi.org/10.1111/1467-8721.ep10774761.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R. *Springer*. https://doi.org/10.1007/978-0-387-87458-6.