

A cross-sectional study of executive and semantic flexibility in childhood using task-switching

Received: 22 January 2026

Accepted: 8 April 2026

Published online: 29 April 2026

Cite this article as: Viviani G., Ambrosini E., Porru A. *et al.* A cross-sectional study of executive and semantic flexibility in childhood using task-switching. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-48580-6>

Giada Viviani, Ettore Ambrosini, Annamaria Porru, Silvia Benavides-Varela, Erin Buchanan, Irene Di Pietro, Daniela Lucangeli & Maria Montefinese

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

A cross-sectional study of executive and semantic flexibility in childhood using task-switching

Giada Viviani¹, Ettore Ambrosini^{2, 3}, Annamaria Porru¹, Silvia Benavides-Varela^{1,2}, Erin Buchanan⁴,
Irene Di Pietro², Daniela Lucangeli¹, Maria Montefinese^{1,*}

¹ Department of Developmental Psychology and Socialisation, University of Padova, Via Venezia 8, Padova, 35131, Italy

² Padova Neuroscience Center, University of Padova, Via Giuseppe Orus 2/B, 35129, Padova, Italy

³ Department of Neuroscience, University of Padova, Via Belzoni 160, 35121, Padova, Italy

⁴ Harrisburg University of Science and Technology, 326 Market Street, 17101, Harrisburg, Pennsylvania, USA

* Corresponding author:

Maria Montefinese,

Department of Developmental Psychology and Socialisation, University of Padova, Italy.

Via Venezia, 8 - 35131, Padova, Italy.

Email: maria.montefinese@unipd.it

Abstract

Cognitive control supports adaptive behavior through stability and flexibility, with executive flexibility typically assessed through task-switching paradigms. However, while executive flexibility is a well-studied construct, it is unclear whether it relies on the same mechanisms as semantic flexibility – the ability to switch between meanings based on context. An ideal approach to arbitrate this debate is to compare their developmental patterns, a method hampered by the fact that semantic flexibility's development remains largely uncharted. Here, 4- to 10-year-old children performed parallel task-switching paradigms: a classic visuospatial paradigm assessing executive flexibility and a novel semantic task assessing semantic flexibility by requiring them to alternate between semantic judgments about whether meaningful concepts were *living vs. non-living* or *moving vs. non-moving*. Results revealed a strong correlation between executive and semantic switch costs, suggesting shared control mechanisms, alongside domain-specific differences and age-related modulations influenced by semantic distance, revealing a growing interplay between semantic knowledge and control as children's conceptual systems mature. These findings provide novel insights into the maturation of cognitive control components in childhood, highlighting the interplay between domain-general executive processes and semantic control mechanisms in flexible cognition.

Keywords: Cognitive Control, Semantic Flexibility, Task-Switching Paradigms, Executive Flexibility, Controlled Semantic Cognition

Introduction

To effectively navigate a constantly changing environment, we rely on cognitive control – a family of top-down mechanisms that organize thoughts and actions to align with our goals¹. According to the cognitive stability-flexibility framework², a first fundamental component of cognitive control is cognitive stability, which contributes to maintaining focus on a relevant "task set" (the rules for a task)² while shielding it from distraction, as typically assessed by conflict paradigms such as the Stroop task^{3,4}. While crucial, stability alone is insufficient. Our dynamic world often requires cognitive flexibility, a complementary component to rapidly switch between task sets as demands change⁵. Thus, the crucial role of cognitive control is to balance stability and flexibility to effectively achieve our goals. For instance, while driving, we must maintain the stable goal of following a route, yet be able to flexibly react to an unexpected road closure by quickly planning a new one. Given its central role in adaptive cognition, flexibility is extensively investigated using task-switching paradigms^{6,7}, whereby participants alternate between two (or more) tasks. The performance difference between "Switch" trials (where the task changes) and "Repeat" trials (where the task remains the same) is known as the (local) 'switch cost'. This cost serves as a key behavioral measure of flexibility, with smaller switch costs indicating greater efficiency in shifting between tasks.

While the task-switching paradigm provides a robust method for measuring cognitive flexibility, many versions typically rely on stimuli that lack real-world meaning, such as simple geometric shapes or digits⁷. While this approach offers experimental control, it reduces ecological validity, as it does not fully capture how we flexibly manipulate meaningful information in everyday life. In natural settings, we frequently need to adapt our use of semantic knowledge based on the current demands. For example, we may be required to shift from the meaning of the word 'jam' referring to a fruit preserve to its alternative meaning referring to a traffic obstruction. This ability requires a form of cognitive control known as *semantic control*^{8–10}, which flexibly prioritizes some of a concept's features or meanings while actively suppressing others that are irrelevant to the current goal.

The concept of semantic control being distinct from domain-general cognitive control (hereafter called *executive control*) is a central tenet of the Controlled Semantic Cognition (CSC) framework¹⁰. This framework posits that semantic control is "a set of executive control processes that regulate the activation and deployment of semantic knowledge. These allow flexible, context- and task-appropriate responses by ensuring that only relevant aspects of semantic representations are used to direct thought and behavior."¹¹ (p. 259). Context-appropriate responding refers to selecting the meaning that is relevant given the surrounding situation, whereas task-appropriate responding refers to selecting the feature required by the current task rule. For example, when viewing an image of a *car*, contextual cues or prior thoughts may make its ability to *move* particularly salient. However, if the

task at hand requires judging whether the object is *living or non-living*, task-appropriate responding requires selecting “non-living” while ignoring the contextually highlighted feature. In many situations these demands align, but when they diverge, cognitive control is required to prioritize task-relevant over contextually salient information.

A dilemma exists, however, regarding the independence of executive and semantic control systems. While semantic control is supported by specific brain regions (such as the left posterior middle temporal gyrus and the inferior frontal gyrus)^{12,13}, neuroimaging evidence also shows partial overlap with the multiple-demand network that underpins executive control, a set of frontoparietal regions consistently recruited across a wide range of cognitively demanding tasks and thought to support domain-general control processes¹⁴. The distinction primarily originates from research on patients with semantic aphasia, who show deficits in using conceptual knowledge despite intact semantic representations^{15–17}. However, these patients also exhibit deficits in executive control, challenging the assumption that the two mechanisms are entirely independent^{15–17}. Further evidence from neuropsychological and behavioral studies suggests that semantic control relies on mechanisms that partially overlap with, but are not identical to, domain-general executive control¹⁸.

Recent work has further emphasized the interplay between domain-general and domain-specific control processes in shaping flexible cognition^{19,20}. To better untangle this relationship, it is crucial to consider that much like its executive counterpart, semantic control comprises components of both stability (maintaining focus on relevant meanings) and flexibility (shifting between meanings as the context demands). To date, however, research exploring the link between the two mechanisms has primarily focused on the stability dimension²¹, whereas the interplay between executive flexibility and semantic flexibility remains largely under-explored. An exception is our recent work with healthy adults, which showed a moderate-to-strong correlation between switch costs in executive and semantic tasks, suggesting shared underlying mechanisms²².

Studying developmental patterns in executive and semantic control from childhood offers a unique window into the architecture of cognitive control, allowing researchers to determine whether the observed associations in adulthood reflect shared developmental origins or later functional convergence. This developmental perspective moves beyond correlational evidence in adults, providing insight into how these mechanisms emerge and interact over time. Given that the link between the two control systems is particularly under-explored for the flexibility component, in both adults²² and children, charting the developmental course of both executive and semantic flexibility is a critical step forward.

In children, executive flexibility is a core component of cognitive control that is crucial for school readiness and later academic success²³. Its developmental trajectory is not a simple, linear

improvement. Although the behavioral ability to switch between tasks matures relatively early, showing significant gains between ages 7 and 11 and reaching adult-like performance around age 12, the underlying neural systems supporting this skill undergo a far more prolonged and complex maturation into adolescence²⁴. Neuroimaging research has sought to pinpoint the specific brain regions that drive this maturation. One influential line of evidence proposes that the sub-processes of task switching are supported by distinct prefrontal regions that mature at different rates²⁵. This model suggests the (pre-)supplementary motor area, a medial frontal region critical for suppressing the previous task set, shows an adult-like activation pattern by early adolescence. In contrast, the ventrolateral prefrontal cortex, a lateral region essential for retrieving and maintaining the current task rule, continues to show immature activation patterns throughout the teenage years²⁵. However, this region-specific dissociation is debated. An alternative, network-based perspective suggests that even young children recruit a broad frontoparietal “multiple-demand network” that is spatially similar to the one used by adults²⁶. From this viewpoint, development is characterized not by different regions coming online at different times, but by a process of network-wide refinement, where activation becomes more efficient and neurally selective with age and improving ability²⁷. In summary, while there is a clear consensus that the neural substrates of executive flexibility undergo a prolonged, non-linear development, the literature presents an active debate. It remains unresolved whether this maturation is best described by the staggered development of specific prefrontal regions handling distinct sub-processes, or by the holistic refinement of a pre-existing, domain-general cognitive control network.

Research^{28,29} has also explored the development of semantic flexibility using stimuli with semantic content. Studies have shown that between the ages of 3 and 6, children improve in switching between verbal rules for sorting cards and in flexibly using changing semantic cues to infer the meanings of new words^{28,29}. For example, preschoolers can learn to switch between sorting objects by a perceptual feature like shape and a more semantic feature like function²⁸. Similarly, word-learning tasks have demonstrated that 4- to 6-year-olds can flexibly use different verbal cues to decide whether a new word refers to an object's shape, material, or another part, a skill that is less developed in 3-year-olds²⁸. Verbal fluency tasks, which measure how many words a child can generate from a specific semantic category, have also been used as an indicator of semantic flexibility, with performance improving with age^{30–33}. These paradigms, while providing valuable insights, often differ from the trial-by-trial cued task-switching methodology typically used to measure executive flexibility. This methodological divergence makes it difficult to determine whether observed improvements are driven by the same underlying control mechanisms isolated by classic switch-cost designs. The development of semantic flexibility is further complicated by its interaction with other cognitive

domains. For instance, a child's growing ability to flexibly use concepts likely reflects the combined contributions of both executive control and the acquisition of domain-specific semantic knowledge³⁴.

Despite this existing body of work, a significant gap remains. No experimental study has explicitly investigated the development of semantic flexibility in children using a cued, trial-by-trial task-switching paradigm, nor has one directly compared it to domain-general executive flexibility using a parallel design. This methodological gap leaves a fundamental developmental question unanswered: are semantic and executive flexibility distinct processes that mature independently, or do they rely on a shared cognitive process that follows a common developmental timeline? Addressing this question has broader theoretical implications, as it informs our understanding of how domain-general and domain-specific control mechanisms interact during development to produce the mature forms observed in adulthood.

To address this question, the present study investigates the developmental patterns of both executive and semantic flexibility. Importantly, because the present study is cross-sectional, the observed age-related differences cannot be interpreted as direct evidence of within-individual developmental change. Longitudinal designs will be necessary to confirm the developmental trajectories suggested by these findings. As previously validated with adult participants²², we employed a task-switching methodology with two parallel paradigms: a classical visuospatial task with minimal semantic content to measure executive flexibility, and a novel task requiring semantic judgments about pictures of meaningful concepts to measure semantic flexibility. Our study has two primary aims. First, we aim to elucidate the relationship between these two core cognitive abilities across development, assessing their developmental interplay. By directly comparing performance on these parallel tasks, we can provide crucial evidence as to whether they are supported by distinct or shared control mechanisms. Second, we aim to gain a deeper understanding of semantic flexibility itself, characterizing its unique developmental features. We will examine how its development is modulated by semantic distance between consecutive concepts, a factor intrinsic to the organization of semantic information.

Methods

In line with the journal's guidelines for transparency and openness, this section details the sample size determination, all criteria for data and participant inclusion/exclusion, and all experimental manipulations and measures. Exclusion criteria were established a priori, before the start of data analysis. While the study was not preregistered, we provide full access to the research materials. The anonymized raw data, materials and codes for the experimental task and analyses are publicly available in our project repository on the Open Science Framework (OSF) platform at <https://osf.io/zr4dn/overview> for reuse by other researchers.

Participants

A total of 655 children participated in the first session of a larger two-session study, which included the experimental paradigms reported here. From these, we then excluded 37 participants (see Data analysis). The final experimental sample included in the analyses consisted of 618 children (290 females, 328 males, see Table 1) aged between 3 years and 10 months and 10 years and 10 months. Within this sample, 531 children provided valid data for both the executive and semantic tasks and were thus included in the analyses comparing executive and semantic flexibility, while 76 children provided valid data for the executive task only, and 11 for the semantic task only.

Table 1. Distribution of the full recruited sample by grade level and gender.

Grade	F	M	Total
Preschool	47	69	116
1 st grade	35	42	77
2 nd grade	76	75	151
3 rd grade	62	65	127
4 th grade	51	55	106
5 th grade	19	22	41
Total	290	328	618

Notes: F, females; M, Males

Participants were recruited from several preschools and primary schools in the Northern part of Italy, within the province of Padova. All children were Italian speakers with typical development, as reported by their parents, no history of neurological, neurodevelopmental, or learning disorders, nor any uncorrected sensory impairments (vision or hearing). Written informed consent was obtained from the parents or legal guardians of all participants, and verbal assent was obtained from each child. The study was conducted in accordance with the ethical standards of the 2013 Declaration of Helsinki for human studies of the World Medical Association and received approval from the Ethical Committee for the Psychological Research of the University of Padova (protocol number: 5272).

General procedure

Participants performed two versions of a two-choice cued task-switching paradigm: an executive and a semantic version. While targeting different domains, both paradigms were designed with an identical structure. The task-switching paradigms were part of a broader experimental battery administered within a larger project. The present study focuses specifically on the executive and semantic flexibility measures, so additional data collected in the same session, not directly relevant to the current research questions, will be reported in separate publications. The task-switching paradigms were both administered during the first of two experimental sessions. To avoid participants by task order interactions, the order of presentation of the two task-switching paradigms was kept

fixed, with all participants performing the executive one before the semantic one. Furthermore, the sequence of trials within each paradigm was also fixed, ensuring all children completed the same trial list.

The experiment was administered using the online version of Labvanced³⁵, displayed in full-screen mode on tablets with a 1280 x 800 pixel resolution (landscape orientation). Data collection took place in a quiet room within the children's schools. In each session, the battery of tasks was administered collectively to the entire class, with each child working on an individual tablet under the supervision of a team of at least three experimenters. Children responded by tapping on one of two designated response areas on the left or right side of the tablet screen (see details below).

The procedure for each domain (executive and semantic) followed a structured sequence. Before each task-switching paradigm, participants first completed two single-task blocks (20 trials each). In these blocks, they performed two discrimination tasks in isolation. These blocks were designed to assess baseline visuospatial and semantic abilities of each child. While the results from these single-task blocks are not reported directly, performance metrics derived from them were used as predictors in our main analyses to control for baseline abilities and obtain more precise estimates of the experimental effects of interest. Once the two single-task blocks were completed, the task-switching was administered in a single, continuous block of 48 trials. Specifically, the two simple discrimination tasks were mixed, and participants had to perform the relevant one based on a trial-by-trial visual cue (see below).

To ensure sustained engagement from our young participants, a gamified approach was adopted throughout the procedure. Each experimental block (executive and semantic single-task and task-switching ones) was preceded by an instructional and training procedure. The lead experimenter (always the same, the author GV) provided standardized instructions to the entire group, adapting the language to the age of the class and using visual aids (slides). To capture the children's attention and enhance compliance and motivation, the instructions were framed within a short, engaging narrative. This was followed by several practical examples where the experimenter collectively guided the children, asking different children to indicate the correct response to ensure the rules were fully understood. After the group instructions, each child completed a brief, individual training phase (4 trials for the single-task blocks, 6 trials for the task-switching blocks) with immediate, visually intuitive on-screen feedback (a happy face for correct answers; a sad face for incorrect/slow responses). During the experimental blocks, motivational images (e.g., a superhero character) were displayed every eight trials to maintain focus and compliance. Finally, both paradigms used child-friendly stimuli, as detailed in the following sections.

Each task-switching paradigm took approximately 15–20 min to complete, including the group instructions, individual training phases, breaks, and the experimental blocks. Specifically, the experimental phase itself was brief (max. 10 min: 2 min for each single-task block and 6 min for the task-switching block). Importantly, the actual duration was participant-dependent because the stimuli were presented until a response was made.

Executive Task-Switching

To assess executive flexibility, we adapted a classic color/shape task-switching paradigm³⁶, implementing a gamified version with child-friendly stimuli. The task required children to sort stimuli representing cartoon aliens based on one of two visuo-spatial features: their color (red, RGB: 192, 0, 0; or blue, RGB: 0, 112, 192) or the shape of their head (squared or rounded) (see Figure 1). Across all stimuli, the body and facial elements were held constant. To promote the use of abstract rules and enhance comparability with the semantic paradigm, four unique variations of each stimulus were created. These variations were generated by combining two internal color-fill patterns (zigzag vs. striped) and two head contours (e.g., square vs. rectangle for the squared-head aliens; circular vs. oval for the rounded-head aliens).

The target stimuli (180 x 255 pixels) were presented at the center of the screen. The experiment was presented on a background depicting a lunar landscape with two extraterrestrial houses, one on the left and one on the right side of the screen, which served as the response areas. Crucially, signs outside each house indicated the correct feature-response mapping. The sign on the left house displayed a red paintbrush and a black square, while the sign on the right house showed a blue paintbrush and a black circle (see Figure 1). The color task therefore required children to sort the alien by color, tapping the left house for red and the right house for blue aliens. The shape task required them to sort it by the shape of its head, tapping the left house for squared-head aliens and the right house for rounded-head aliens. Therefore, the employed stimuli were bivalent, that is, they afforded both a color and a shape judgment, requiring children to flexibly select the relevant stimulus feature based on the currently relevant task, which was signaled by a visual task cue (255 x 106 pixels) presented at the top-center of the screen: an image of two paintbrushes for the color task or two geometric shapes for the shape task (see Figure 1). Moreover, we used only the two stimulus types that were task-incongruent: the red alien with the rounded head, requiring a left response for the color task but a right response for the shape task, and the blue alien with the squared head, requiring a right response for the color task but a left response for the shape task. This was critical to maximize flexibility demands through cross-task interference, while ensuring that correct responses on Switch trials required an actual task-set shift.

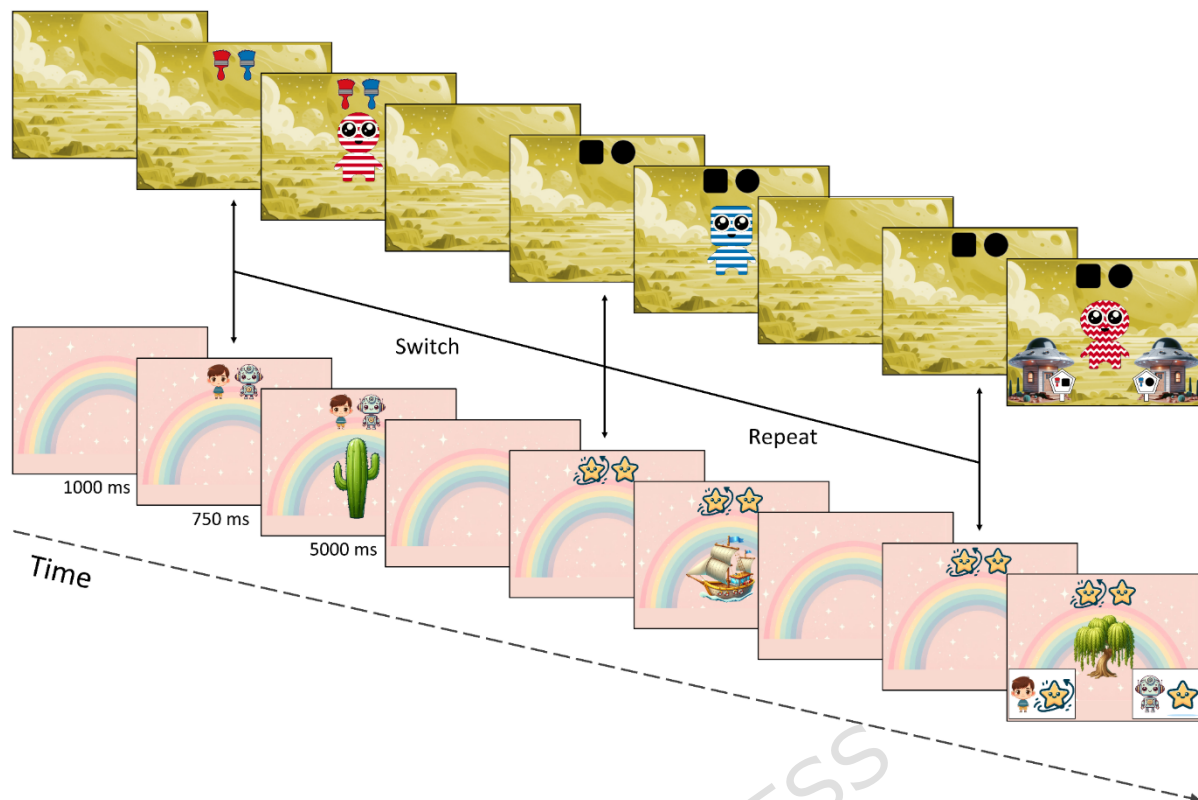


Figure 1. Task-switching experimental paradigms. Top panel: Executive task-switching. Children judged whether a cartoon alien was red/blue (color task) or had a squared/rounded head (shape task), based on a visual cue (paintbrushes for the color task; shapes for the shape task). Responses were made by tapping on one of two extraterrestrial houses, which had signs displaying the feature-response mapping (left house: red and squared; right house: blue and rounded). Bottom panel: Semantic task-switching. Children judged whether an image referred to a living/non-living or moving/non-moving concept, based on a visual cue (a child/robot for the living/non-living task; a stationary/spinning star for the moving/non-moving task). Responses were made by tapping on one of two areas displaying the concept-response mapping (left area: living and moving; right area: non-living and non-moving). For both paradigms, three example trials are shown. The second trial is a Switch trial, where children were required to switch from the task performed in the previous trial to the other task (e.g., from discriminating alien's color to discriminating alien's head shape, or from classifying the concept as living/non-living to classifying it as moving/non-moving). The third trial is a Repeat trial, where children kept performing the same task as in the previous trial (e.g., discriminating the alien's head shape or classifying the concept as moving/non-moving).

The entire procedure was framed within a narrative where children had to help the lost aliens return to the correct house. They were told that to select the correct house first they had to follow a clue (paintbrushes or shapes). Each trial began with the presentation of the lunar landscape for 1000

ms. The task cue then appeared for 750 ms, after which the target alien appeared in the center while the cue remained visible. Both the cue and stimulus stayed on screen until a response was made or for a maximum of 5000 ms. Trials could either be Repeat (same task as the previous trial) or Switch trials (different task) (see Figure 1).

The executive task-switching paradigm consisted of a single block of 48 experimental trials, preceded by one buffer trial. Repeat and Switch trials were interspersed and equally probable (50% switch probability). The trial sequence was pseudorandomized to prevent immediate repetitions of the exact same stimulus image and to avoid more than four consecutive repetitions of the same task or response.

Semantic Task-Switching

Semantic flexibility was assessed using a semantic task-switching paradigm adapted for children from a version previously validated with adults²². While maintaining the same underlying structure as the executive task, this paradigm used stimuli with semantic meaning (instead of visuo-spatial information) to tax flexible control within the semantic domain, requiring participants to flexibly switch between different meaning aspects of a concept³⁷. Children were required to categorize cartoon images based on one of two conceptual dimensions: whether the object depicted was living or non-living, or whether it was capable of moving or non-moving.

The target stimuli were colorful, child-friendly cartoon images (255 x 255 pixels) presented at the center of the screen (see Figure 1). These images depicted concepts belonging to two categories, Plants (living and non-moving concepts) or Vehicles (non-living and moving concepts). These categories were selected to include concepts that are highly familiar to young children and typically acquired early in development, and that are frequent in everyday language. In addition, during the training phase, we ensured that all participants correctly recognized the stimuli and understood the task requirements before proceeding to the experimental trials.

Four exemplars were selected for each category (Plants: tree, sunflower, cactus, potted plant; Vehicles: bicycle, tractor, boat, helicopter). Each image was created ex-novo to be engaging and colorful, and all shared a consistent visual style. As in the executive task-switching paradigm, the selected categories made target stimuli bivalent and task-incongruent (i.e., stimuli that both tasks can be performed on and that required different responses based on the task cue).

The experiment was presented on a rainbow-themed background with two response areas on the left and right. Each response area displayed images representing the possible conceptual features. The left area showed both a cartoon child (representing living concepts) and a spinning star (representing moving concepts). The right area showed both a cartoon robot (representing non-living concepts) and a stationary star (representing non-moving concepts). Therefore, for the living/non-living task,

children had to tap the left area for living images (i.e., Plants) and the right area for not-living images (i.e., Vehicles), while for the moving/non-moving task, they had to tap the left area for moving images (i.e., Vehicles) and the right area for non-moving images (i.e., Plants) (see Figure 1). The task to be performed was signaled by a cue stimulus (300 x 160 pixels) presented at the top-center of the screen. The cue for the living/non-living task was an image showing both the child and the robot side-by-side, while the cue for the moving/non-moving task showed the two stars. The narrative instructed the children to sort the pictures based on the specific cue they saw on each trial.

All other methodological details, including trial timing and the pseudo-randomization of the 48-trial sequence, were kept identical to the executive task-switching paradigm.

Data analysis

Statistical analyses were performed using R (<https://www.r-project.org/>, version 4.4.1³⁸) on inverse-transformed RTs (iRTs), computed as $1000/\text{RTs}$. This transformation was used to mitigate the heavy right skewness common in RT distributions and normalize their residuals. Crucially, iRTs serve as a measure of processing speed or performance, where higher values correspond to better performance (i.e., faster responses, interpreted as responses-per-second). We chose this transformation as our primary approach because it provided the most stable model convergence and the best alignment of residuals with the normality assumption, compared to LMM performed on log-transformed RTs and on raw RTs, as well as to GLMM performed on raw RTs using the gamma distribution (see Figure S1 in the online supplementary materials at <https://osf.io/zr4dn/files/m24pj> for residual diagnostic plots). It is important here to note that these alternative analyses replicated our primary findings, including the critical interactions, demonstrating the robustness of our results across different analytical strategies (see Tables S1-3 in the online supplementary materials at <https://osf.io/zr4dn/files/m24pj>).

Thirty-seven participants (5.65% of the recruited participants) were excluded from the analyses because they did not complete the two single-task and the two task-switching blocks or showed poor compliance/performance (i.e., failed to provide a response in more than 80% of trials, provided the same response in more than 90% of trials, and/or provided the incorrect response in more than 90% of trials).

Practice trials and the first trial of each block were excluded from analyses. We then excluded experimental trials with incorrect responses ($n = 10472$, 18.98% of trials), missed responses ($n = 3058$, 5.54% of trials) and anticipations (i.e., trials with RTs < 150 ms, $n = 79$, 0.14% of trials), which were both considered as errors, as well as trials with extreme iRTs in the preceding trial (i.e., those with an absolute z-scored iRT > 3; $n = 195$, 0.035 % of trials).

To address both our aims, we used linear mixed-effects model (LMM) analyses³⁹. This analytical approach accounts for the hierarchical structure of our data, where observations are nested within individual participants and experimental stimuli. Unlike traditional ANOVA, LMMs allow us to simultaneously model both subject-specific and stimulus-specific variance (random effects), thereby increasing the generalizability of our findings. Furthermore, LMMs are robust to the unbalanced nature of datasets often encountered in developmental studies, such as missing trials or varying response counts across participants, providing more accurate and unbiased estimates of the developmental trajectories under investigation.

Aim 1 - Executive vs Semantic Flexibility

First, to compare the effects of executive and semantic flexibility on behavioral performance and their interplay across development, we performed a within-subject analysis testing an a-priori defined model (M_EvsS_iRT) on the performance data (iRT) from both the executive and semantic paradigms. The model's Wilkinson-notation formula was: $iRT \sim iRT_{pre} + Trial*Age + postERR*Age + Dom*Cond*Age + Dom*Cond*PerfBase + (Dom*Cond|SS) + (1|Stim)$.

Control variables. The fixed part of the model included several confounding predictors: i) a continuous predictor for the iRT of the preceding trial (iRT_{pre}) to account for temporal dependency in iRTs⁴⁰; ii) a continuous predictor for trial number ($Trial$) to account for time-on-task effects (e.g., learning and/or fatigue); iii) a dichotomous predictor for an error in the preceding trial ($postERR$) to account for post-error slowing⁴¹. The latter two predictors were tested in interaction with Age to capture any age-related modulation of these effects.

Fixed effects. To test how executive-semantic flexibility interplay modulated the switch cost across development, we included predictors for Domain (Dom : Executive vs. Semantic), Condition ($Cond$: Repeat vs. Switch) and the continuous predictor for Age , modeling them in a three-way interaction ($Dom*Cond*Age$). We focused on this trial-by-trial 'local' switch cost – the performance difference between Switch and Repeat trials – because it provides a purer index of task-set reconfiguration, which is the core mechanism of cognitive flexibility. Conversely, we avoided the so-called 'global' switch cost (the performance difference between mixed and single-task blocks), as this measure is often confounded by sustained task-set maintenance and block-level monitoring demands, which do not reflect the capacity for flexible shifting. However, to control for how the $Dom*Cond$ interaction was modulated by children's baseline abilities, we also included the three-way interaction between Dom , $Cond$ and baseline performance ($PerfBase$), where $PerfBase$ is a continuous predictor representing each participant's mean iRT in executive and semantic domain in single-task blocks, that is, when no flexibility was required.

Random effects. To rigorously account for individual variability and avoid anti-conservative estimates, we employed a maximal random-effects structure justified by the data. The random effect structure included a by-participant (SS) random intercept for overall performance differences, as well as by-participant random slopes for Domain, Condition, and their interaction, allowing us to account for individual differences in domain sensitivity and switch costs. A by-stimulus (Stim) random intercept was also included to account for variability attributable to specific stimuli. Random slopes for trial-level covariates (e.g., preceding iRT, trial order) were omitted to ensure model convergence and prevent overparameterization.

Variable coding. To facilitate model convergence and interpretation, the continuous confounding predictors (Trial and iRTpre) were centered and scaled (from -.5 to .5) and z-scored, respectively, at the participant level. The categorical predictor postErr was factor-coded (0 = after correct, 1 = after error), with the 0 acting as the reference level. Age was z-scored across the entire sample and PerfBase by-domain. The predictors of interest, Dom and Cond, were treated as continuous numerical variables and centered around zero (-0.5 for Executive/Repeat, 0.5 for Semantic/Switch), so that the estimated main effects and interactions refer to the average across conditions.

Outliers. Finally, after fitting the initial model, we examined the residuals to check for the presence of extreme outliers and re-fitted a trimmed version, excluding data points with absolute standardized residuals greater than 3.

For the trimmed model, we reported the estimated coefficients (b), standard error (SE), t and p values for each fixed effect. Satterthwaite's approximation of degrees of freedom was used to calculate p -values and to derive effect size estimates (d_s) for the experimental effects of interest, which were also calculated using Westfall's approach (d_w)⁴². We additionally calculated the conditional R^2 values using the MuMIn package⁴³. An alpha level of .05 was set as the cut-off for statistical significance.

Switch cost correlation. Then, to directly compare executive and semantic flexibility effects on performance and further assess their commonalities independent of developmental maturation and general processing speed, we correlated the individual executive and semantic switch costs. These costs were calculated using the participants' Best Linear Unbiased Predictors (BLUPs) extracted from the LMM. Because the LMM explicitly modeled age and baseline performance as fixed effects interacting with task conditions, the extracted BLUPs inherently represent individual task-specific abilities adjusted for age and baseline speed. We computed both Pearson's r and Spearman's ρ correlation coefficients between these age-adjusted individual switch costs using the package pcor.test⁴⁴.

Aim 2 - Semantic Flexibility

Our second aim was to characterize the development of semantic flexibility by examining how it is modulated by the organization of semantic information. To this end, we conducted a focused analysis on iRT during the semantic paradigm alone. We tested an a priori defined model (M_Sem_iRT) based on our theoretical assumptions and previous findings²². The model's Wilkinson-notation formula was: $iRT \sim iRT_{pre} + Trial*Age + postERR*Age + Cond*Age + Cond*PerfBase + Age*SemDist + (1|SS) + (1|Stim)$.

Control variables. As in the previous model (M_EvsS_iRT), we included confounding predictors (iRT_{pre}, Trial*Age, postERR*Age) and two key interactions: Cond*Age, to test how age modulates the semantic switch cost, and Cond*PerfBase, to assess the role of baseline abilities.

Fixed Effects. Crucially, building on our previous evidence showing that the representational distance between concepts influences performance^{17,22,45,46}, we also included a continuous predictor for semantic distance (SemDist). SemDist reflects the cosine distance between the concepts of the current and previous stimulus (trial n vs. n-1), derived from fastText, a distributional semantic model trained on Italian corpora⁴⁷⁻⁴⁹. Our previous results revealed that when stimuli change on every trial (like in our paradigm), participants are required to flexibly navigate through the semantic space. When stimuli are semantically more distant from each other, they require greater flexibility – a mechanism we termed stimulus-level flexibility because it operates regardless of the task condition. Therefore, we tested SemDist in interaction with Age to explore whether this stimulus-level flexibility changes across development.

However, in addition to this stimulus-level effect, semantic distance might also modulate the degree of the type of flexibility required to switch between tasks, reflected in the switch cost. Therefore, to test whether SemDist modulates the semantic switch cost and whether this modulation changes with age, we specified a second, more complex model (M_SemFull_iRT), which added a three-way interaction (Cond*Age*SemDist). The formula was: $iRT \sim iRT_{pre} + Trial*Age + postERR*Age + Cond*Age + Cond*PerfBase + Cond*Age*SemDist + (1|SS) + (1|Stim)$. We then formally compared the fit of M_Sem_iRT and M_SemFull_iRT using a log-likelihood ratio test to determine if the inclusion of this critical interaction was statistically justified. For this analysis, the SemDist predictor was centered, and the rest of the model fitting procedure was identical to that described above.

Random effects. The random effect structure included a by-participant (SS) random intercept to account for individual differences in overall performance, and a by-stimulus (Stim) random intercept to account for variability attributable to specific stimuli. The inclusion of the Cond random slope at the participant level led to convergence issues.

Sensitivity Analysis on Accuracy

To control for possible speed-accuracy trade-offs, the analyses detailed above were also performed on accuracy as the dependent variable, but in this case using a generalized LMM (with the binomial family) and considering the mean accuracy in the two single-task blocks for the PerfBase predictor. Therefore, the M_EvsS_Acc model compared the effects of executive and semantic flexibility on children's accuracy and their interplay across development, while the M_SemFull_Acc model assessed SemDist modulations of the semantic flexibility. It is worth noting that we attempted to fit the same maximal random-effects structure to our binary accuracy data but, due to the inherent complexity of generalized LMMs (which involve non-linear link functions), these models failed to converge. Given these computational limitations, we maintained a more parsimonious random-effects structure for accuracy analyses, only including the by-participants and by-stimulus random intercepts. Moreover, in these analyses, Satterthwaite's approach could not be used to compute the effect sizes because this method is not applicable to generalized mixed models; therefore, effect sizes were only computed following Westfall's (2014) approach³⁶.

Power Analysis

The sample size was determined based on the aims of the larger study that included, among others, the two task-switching paradigms described here. We recruited as many participants as possible based on the availability of resources and access to local schools. Nonetheless, we performed a sensitivity power analyses using the method introduced by Westfall and colleagues³⁶ for a fully-crossed linear mixed-effects model, conservatively assuming participant and stimulus intercepts and residual variance partitioning coefficients of .2, .1, and .7, respectively⁴². The variance partitioning coefficients for the participant and stimulus slopes and the participant-by-stimulus intercept were set to 0 because those effects were not included in the models we tested. This power analysis revealed that a sample size of 618 participants and 16 stimuli was large enough to detect a very small effect size (Cohen's $d = .036$) with a power of .80. It should be noted, however, that this approach is not adequate for complex mixed effect models like the one used in this work^{42,50}, but it nonetheless provides a useful estimation of the so-called minimal statistically detectable effect for our study (i.e., the lower bound of the range of effect sizes that can be detected⁵¹). Indeed, to the best of our knowledge, to date there are no accepted analytical approaches to accurately compute statistical power for such models.

Results

Aim 1 - Executive vs. Semantic Flexibility

The first analysis compared the behavioral effects of executive and semantic flexibility and explored their interplay across development. The conditional R^2 of the trimmed LMM model

(M_EvsS_iRT) provided a good fit to the data (Conditional $R^2 = .38$), with 0.66% of observations removed as extreme outliers (>3 absolute standardized residuals).

All confounding predictors modulated participants' performance (iRT) significantly (see Table 2 for full model details).

Table 2 - Results of the LMM model assessing the domain-dependent switch effects on iRTs

	Estimate	SE	df	t	p	d_s	d_w
(Intercept)	0.5954	0.0057	84.7	104.67	< .0001	11.38	2.33
iRT preceding trial	0.0410	0.0011	40439.2	37.12	< .0001	0.18	0.18
Trial	0.0141	0.0040	11583.7	3.53	.0004	0.03	0.07
Age	0.0066	0.0057	857.0	1.16	.2448	0.04	0.03
Post-error	-0.0483	0.0028	38769.2	-17.17	< .0001	-0.09	-0.24
Domain	0.0194	0.0085	28.3	2.28	.0306	0.43	0.10
Condition	-0.0627	0.0025	773.4	-25.43	< .0001	-0.91	-0.83
Baseline Performance	0.0777	0.0051	1078.8	15.09	< .0001	0.46	0.94
Trial:Age	-0.0120	0.0036	39598.3	-3.35	.0008	-0.02	-0.12
Age:Post-error	-0.0386	0.0029	31135.0	-13.53	< .0001	-0.08	-0.30
Domain:Condition	0.0367	0.0044	9150.4	8.29	< .0001	0.09	0.26
Age:Domain	0.0048	0.0072	572.1	0.66	.5076	0.03	0.03
Age:Condition	-0.0151	0.0032	785.5	-4.75	< .0001	-0.17	-0.10
Domain:Baseline Performance	-0.0277	0.0074	583.9	-3.75	.0002	-0.16	-0.16
Condition:Baseline Performance	-0.0115	0.0030	889.9	-3.83	.0001	-0.13	-0.07
Age:Domain:Condition	-0.0098	0.0058	9492.5	-1.70	.0884	-0.02	-0.06
Domain:Condition:Baseline Performance	0.0112	0.0056	6911.7	1.99	.0468	0.02	0.07

Notes: SE, standard error; df, degrees of freedom computed with the Satterthwaite's approximation; d_s , effect size estimates calculated with Satterthwaite's approach; d_w , effect size estimates calculated with Westfall's approach.

The analyses revealed various significant effects. Specifically, performance (iRT) improved significantly with higher baseline performance. Crucially, there was a main effect of Condition, confirming the presence of a significant switch cost, with better performance on Repeat trials compared to Switch trials. Moreover, the Age*Condition interaction was significant, indicating that performance improved with age more steeply for Repeat trials than for Switch trials (Fig. 2).

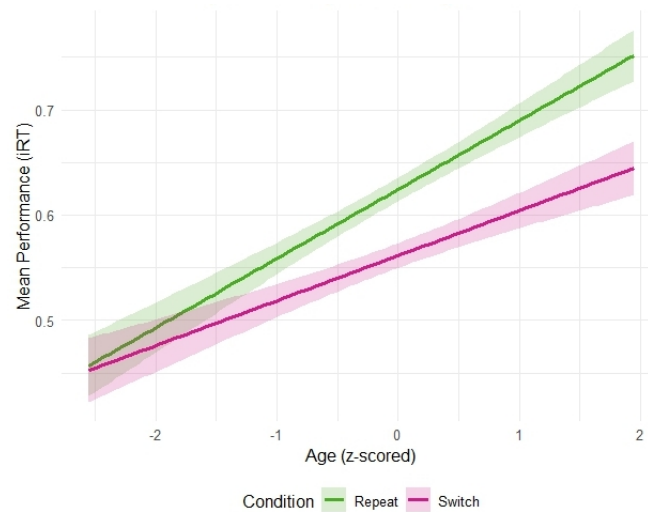


Figure 2. Results of the LMM on iRTs, Age by Condition interaction. The plot shows the behavioral performance (iRTs) corresponding to the marginal effects estimated by the linear mixed model as a function of Age (x-axis) and Condition (Repeat, green vs Switch, fuchsia lines). Shaded regions indicate the 95% confidence interval.

We also found a significant interaction between Domain and the baseline performance (PerfBase), showing that the greater the baseline performance the greater the performance in task-switching, especially in the executive domain. The Condition*PerfBase interaction was also significant, showing that the performance benefit of having higher baseline abilities was more pronounced for Repeat trials than for Switch trials.

Especially relevant to our research question was the Domain*Condition interaction, which was significant, revealing that the magnitude of the switch cost was significantly larger in the executive domain compared to the semantic domain (Fig. 3).

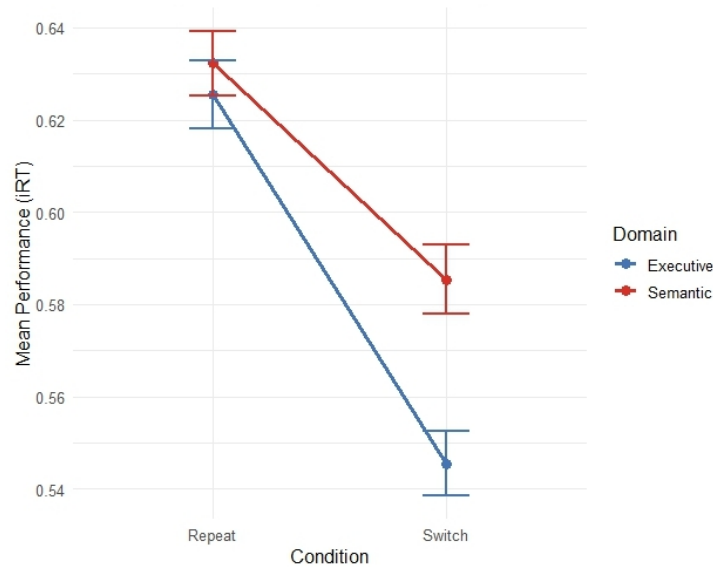


Figure 3. Results of the LMM on iRTs, Domain by Condition interaction. The plot shows the behavioral performance (iRTs) corresponding to the marginal effects estimated by the linear mixed model as a function of Condition (x-axis) and Domain (Executive, blue vs Semantic, red lines). Shaded regions indicate the 95% confidence interval.

Finally, the three-way Age*Domain*Condition interaction was not significant, whereas the p -values for the Domain*Condition*PerfBase interaction was just below .05 (and it is not significant in the analysis on accuracy, see below), thus it will not be discussed further.

To further probe the relationship between the two forms of flexibility, we correlated the individual switch cost effects estimated by the LMM model. Despite the difference in magnitude, the executive and semantic switch costs were strongly and positively correlated (Pearson's $r = .810$, $t(529) = 31.727$, $p < .00$; Spearman's $\rho = .812$, $t(529) = 31.973$, $p < .0001$; see Figure 4).

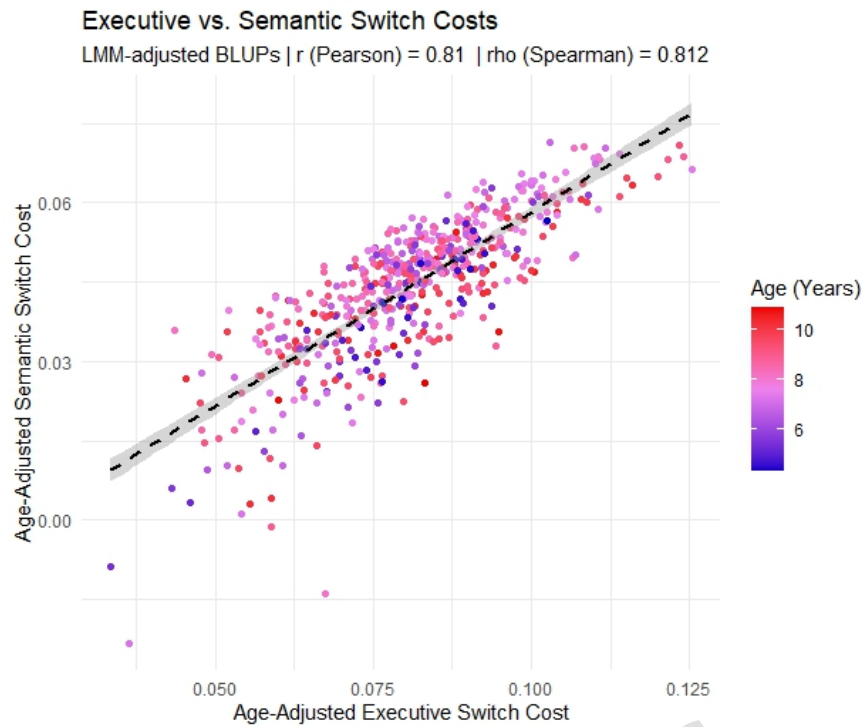


Figure 4. Correlation between the switch costs in the two Domains. The figure shows the correlation between participants' switch costs in the Semantic (y-axis) and Executive (x-axis) Domains.

The analysis on the children's accuracy substantially confirmed the results reported above, except for the Dom*PerfBase interactions, whose statistical significance was not confirmed (see Table 3).

Table 3 - Results of the LMM model assessing the domain-dependent switch effects on accuracy

	Estimate	SE	z	p	d_w
(Intercept)	1.4475	0.0362	40.01	< .0001	3.34
iRT preceding trial	-0.0319	0.0114	-2.81	.0050	-0.07
Trial	0.0095	0.0435	0.22	.8281	0.02
Age	0.6157	0.0302	20.39	< .0001	1.42
Post-error	-0.4922	0.0259	-19.02	< .0001	-1.14
Domain	-0.2955	0.0468	-6.31	< .0001	-0.68
Condition	-0.4960	0.0240	-20.65	< .0001	-1.15
Baseline Performance	0.0949	0.0170	5.59	< .0001	0.22
Trial:Age	-0.0506	0.0362	-1.40	.1624	-0.12
Age:Post-error	-0.1857	0.0234	-7.94	< .0001	-0.43
Domain:Condition	0.3339	0.0475	7.03	< .0001	0.77
Age:Domain	0.0208	0.0256	0.81	.4162	0.05
Age:Condition	-0.1634	0.0234	-6.99	< .0001	-0.38
Domain:Baseline Performance	0.0385	0.0295	1.31	.1918	0.09
Condition:Baseline Performance	-0.0416	0.0209	-1.98	.0472	-0.10
Age:Domain:Condition	0.0238	0.0465	0.51	.6091	0.06

Domain:Condition:Baseline Performance	0.0091	0.0419	0.22	.8285	0.02
--	--------	--------	------	-------	------

Notes: SE, standard error; d_w , effect size estimates calculated with Westfall's approach.

Aim 2 - Semantic Flexibility

Our second analysis aimed to characterize the development of semantic flexibility. As the log-likelihood ratio test confirmed that the model including the three-way interaction (M_SemFull_iRT) provided a significantly better fit to the data, we report the results from this more complex model (see Table 4).

First, the analysis confirmed a significant effect of Condition, indicating a robust switch cost within the semantic domain. We also found a significant effect of Semantic Distance, revealing that performance worsened as the semantic distance between consecutive stimuli increased.

Crucially, we found a significant three-way Age*Condition*Semantic Distance interaction. This interaction revealed that the performance worsening associated with greater semantic distance was more pronounced on Repeat trials, and that this effect became stronger with age (Figure 5). In older children especially, this resulted in a notable cost for repeating a task when they had to switch towards more semantically distant stimuli.

Table 4 - Results of the LMM model assessing semantic switch effect on iRTs

	Estimate	SE	df	t	p	d_s	d_w
(Intercept)	0.6192	0.0081	39	76.92	< .0001	12.26	2.71
iRT preceding trial	0.0325	0.0017	18570	19.63	< .0001	0.14	0.14
Trial	0.0402	0.0060	5441	6.75	< .0001	0.09	0.18
Age	-0.0019	0.0079	924	-0.25	.8051	-0.01	-0.01
Post-error	-0.0344	0.0040	18700	-8.57	< .0001	-0.06	-0.15
Condition	-0.0602	0.0081	12360	-7.45	< .0001	-0.07	-0.26
Baseline Performance	0.0856	0.0066	510	13.00	< .0001	0.58	0.37
Semantic Distance	-0.0784	0.0173	12090	-4.54	< .0001	-0.04	-0.34
Trial:Age	-0.0082	0.0055	18330	-1.50	.1350	-0.01	-0.04
Age:Post error	-0.0342	0.0042	18810	-8.22	< .0001	-0.06	-0.15
Age:Condition	-0.0459	0.0086	18330	-5.36	< .0001	-0.04	-0.20
Condition:Baseline Performance	-0.0056	0.0039	18310	-1.42	.1544	-0.01	-0.02
Condition:Semantic Distance	0.0824	0.0339	13200	2.43	.0151	0.02	0.36
Age:Semantic Distance	-0.0174	0.0173	18320	-1.01	.3149	-0.01	-0.08
Age:Condition:Semantic Distance	0.1346	0.0344	18330	3.91	.0001	0.03	0.59

Notes: SE, standard error; df, degrees of freedom computed with the Satterthwaite's approximation; d_s , effect size estimates calculated with Satterthwaite's approach; d_w , effect size estimates calculated with Westfall's approach.

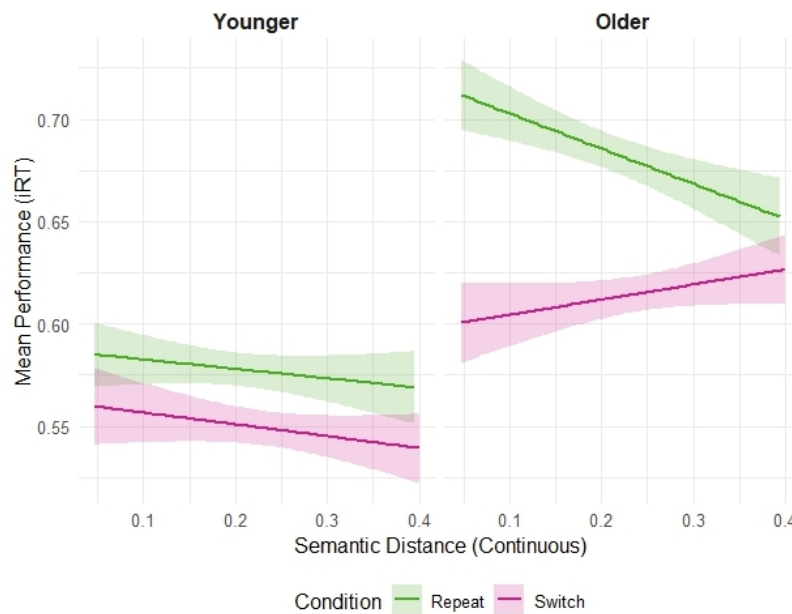


Figure 5. Results of the LMM on the semantic paradigm, three-way interaction between Age, Condition and Semantic Distance. The plot shows the behavioral performance (iRTs) corresponding to the marginal effects estimated by the linear mixed model as a function of Semantic Distance (x-axis), Condition (Repeat, green vs Switch, fuchsia lines), and Age (z-scored, left and right panels, corresponding to 4 years, 5 months and 10 years, 10 months, respectively). Shaded regions indicate the 95% confidence interval.

The analysis on children's accuracy largely mirrored the iRT results. Although the main effect of Semantic Distance was not significant (see Table 5), the key three-way interaction between Age, Condition, and Semantic Distance remained significant (see Table 5).

Table 5 - Results of the LMM model assessing semantic switch effect on accuracy

	Estimate	SE	z	p	dw
(Intercept)	1.3769	0.0598	23.02	< .0001	1.92
iRT preceding trial	0.0013	0.0172	0.08	.9391	< 0.01
Trial	0.0669	0.0618	1.08	.2793	0.09
Age	0.5874	0.0547	10.75	< .0001	0.82
Post-error	-0.3967	0.0368	-10.79	< .0001	-0.55
Condition	-0.4775	0.0831	-5.75	< .0001	-0.67
Baseline Performance	0.3134	0.0370	8.48	< .0001	0.44
Semantic Distance	-0.2366	0.1765	-1.34	.1800	-0.33
Trial:Age	-0.0804	0.0539	-1.49	.1359	-0.11
Age:Post-error	-0.1785	0.0344	-5.18	< .0001	-0.25
Age:Condition	-0.4177	0.0794	-5.26	< .0001	-0.58
Condition:Baseline Performance	-0.0413	0.0301	-1.37	.1698	-0.06
Condition:Semantic Distance	0.7035	0.3477	2.02	.0431	0.98
Age:Semantic Distance	-0.1509	0.1666	-0.91	.3650	-0.21

Age:Condition:Semantic Distance 1.2505 0.3309 3.78 .0002 1.74

Notes: SE, standard error; d_w , effect size estimates calculated with Westfall's approach.

Discussion

This study investigated the developmental patterns of executive and semantic flexibility in children using parallel versions of a cued task-switching paradigm to address a fundamental question raised in the adult cognitive control literature: whether these two forms of flexibility rely on distinct or shared mechanisms. By charting these abilities throughout childhood, we sought to determine if they follow distinct or shared developmental patterns. Our findings reveal a complex picture, characterized by evidence consistent with a strong shared component that is nonetheless modulated by domain-specific factors. These results provide a developmental perspective on executive and semantic control highlighting the intricate nature of cognitive development.

Our first key finding demonstrates a strong correlation between switch costs in the executive and semantic domains, consistent with the idea of shared control mechanisms. This result is consistent with findings in healthy adults²², which show a similar moderate-to-strong correlation. Finding this strong association in children suggests the link between these abilities is a fundamental aspect of development, not just a mature state. This aligns with theoretical models which argue that a core set of executive functions drive flexible behavior^{52,53}. The mental act of switching tasks also taxes working memory to maintain task sets and apply the relevant one, and inhibitory control to stop applying the old one. In essence, both our paradigms relied on this same underlying cognitive machinery for switching. Therefore, a child's developmental level in operating this machinery would naturally lead to similar performance across both domains, explaining the strong correlation.

While the correlation suggests a shared component, our findings are consistent with the asynchronous and multi-faceted nature of cognitive control development highlighted in our introduction^{24,52,54}. The significant Age by Condition interaction revealed that performance on Repeat trials improved more steeply with age than on Switch trials. This pattern points to a growing functional dissociation between more automatized task execution and effortful cognitive control. As children age, they become increasingly efficient at maintaining and executing an active task-set, leading to dramatic performance gains on Repeat trials. However, this emerging task automatization introduces a greater degree of task inertia. A more consolidated cognitive system, while more efficient, naturally offers more resistance to flexibility. Consequently, the process of switching to a newly relevant task-set necessarily requires a costly control intervention to override this emerging automaticity and implement a new goal-directed action. While this flexibility ability also matures, its developmental gains observed on Switch trials are intrinsically less pronounced than those afforded by the powerful process of task automatization on Repeat trials. Our results, therefore, suggest that a key

developmental achievement in this age range is the ability to form and execute efficient cognitive routines, which in turn makes the relative cost of volitionally breaking those routines – the switch cost – more apparent.

Our results of a significant Domain by Condition interaction also challenge a purely domain-general account by revealing domain-specific modulations. The larger switch cost in the executive domain than in the semantic domain is consistent both with findings with adults²² and developmental research showing that switching between highly interdependent perceptual dimensions (like color and shape) is particularly difficult for young children^{28, 54}. As discussed in our introduction, prior studies have shown children successfully switching between perceptual rules (shape) and semantic ones (function)²⁸. Our finding extends this, suggesting that a switch between two competing perceptual rules may be uniquely demanding.

Our most novel contribution is the finding of a three-way interaction between Age, Condition, and Semantic Distance, that speaks directly to the CSC framework¹⁰ and our second aim of characterizing semantic flexibility itself. The finding that older children showed a performance cost for larger semantic distance on Repeat trials, but not on Switch trials, is consistent with the interplay between the semantic "hub" (the knowledge store where conceptual distance is represented) and the "control" system¹⁰. This pattern is consistent with the possibility that semantic representations become more structured with age, thus becoming increasingly similar to adult-like representations²², although alternative explanations cannot be ruled out. It is important to note that semantic distance was computed using distributional semantic models (fastText), which capture both taxonomic and thematic relationships based on patterns of language use. However, these dimensions were not independently manipulated in the present study.

In addition, although our analyses focused on semantic distance as a continuous measure of stimulus-level variability, perceptual differences between stimuli may also have contributed to performance, particularly given that both tasks involved visual processing. In the executive task, perceptual variation was partially controlled by manipulating a set of discrete, task-irrelevant visual features (i.e., internal patterns and head proportions). While this ensured that the physical stimulus varied across trials even within task-set repetitions, these features are discrete and limited in complexity, preventing the derivation of a continuous 'perceptual distance' metric comparable to our semantic distance measure. Consequently, while we controlled for basic stimulus repetition, we cannot exclude that more complex perceptual similarities might contribute to performance. Future research could systematically manipulate perceptual similarity to better disentangle the contribution of perceptual versus conceptual factors to cognitive flexibility.

Moreover, although the stimuli were selected to be highly familiar and early-acquired, we did not explicitly control for variables such as age of acquisition or lexical frequency. These factors may influence semantic processing across development and could have contributed to the observed effects. Future research should aim to disentangle the role of different types of semantic relations and lexical properties in shaping semantic flexibility.

However, caution is warranted when inferring properties of semantic representations solely from behavioral response time data. As highlighted by the “maps and vehicles” problem⁵⁵, observed performance patterns may reflect not only the structure of the underlying semantic space (“the map”), but also the processes used to navigate it (“the vehicle”). In this framework, age-related differences in the effect of semantic distance could arise from changes in control processes, search strategies, or processing efficiency, rather than from changes in representational structure per se. Indeed, previous work has shown that controlled semantic retrieval is shaped by both representational and executive factors^{56,57}. Therefore, while our findings are compatible with the hypothesis of increasing semantic organization, they should be interpreted with caution, as they do not allow us to disentangle representational from process-based accounts.

As such, even when repeating the same task, traversing this space requires a form of flexibility to navigate through the stimuli in the semantic space (see stimulus-level flexibility)²². The masking of this effect on switch trials is what the CSC model would predict: the resource-intensive process of reconfiguring the entire control network leaves fewer resources for finer-grained, representation-level modulations. This echoes the broader theme that developmental trajectories are often not linear but are marked by increasing complexity and dynamics. Of course, this interpretation must be tempered by the possibility of a statistical floor effect. Performance on switch trials is already under such high load that our behavioral measures may lack the sensitivity to detect a more subtle, additional cost from semantic distance. Yet, this finding is consistent with the idea raised by research on early conceptual development²⁸: understanding the development of flexibility requires understanding the development of the knowledge it controls²⁸.

The potential confounds discussed above represent the most significant limitations of our study. We must explicitly state that while our model controlled for baseline performance, the larger executive switch cost could stem from differences in switch-specific task difficulty or rule type. Future work must aim to equate the interference demands between paradigms to isolate the specific effects of information content. Importantly, the present study adopts a cross-sectional design, and therefore the observed age-related differences reflect group-level comparisons rather than within-individual developmental change. Future longitudinal studies will be necessary to directly track how executive and semantic flexibility evolve within individuals across development.

Finally, while our behavioral data are informative, their greatest value may be in generating specific, testable hypotheses for future neuroimaging research. The strong behavioral correlation between switch costs is consistent with the hypothesis of a shared neural basis, which should be mirrored by correlated activation in the domain-general multiple-demand network that supports a wide range of effortful cognitive tasks¹⁴. Correspondingly, the unique variance in semantic task performance, particularly the modulation by semantic distance, is consistent with the possibility of a distinct neural basis uniquely explained by activity in semantic-specific control regions, such as the left inferior frontal gyrus and posterior middle temporal gyrus, consistent with the CSC framework^{8,10}. However, we must remain cautious regarding the interpretation of the shared variance. Given the strong structural similarity between our two experimental paradigms, this strong behavioral correlation could also be driven by the maturation of a more general ability to consciously represent and apply paired-rule structure and stimulus-response mappings, regardless of their content⁵⁴. A child who has mastered this fundamental logical operation will succeed at both tasks, while a child who has not will struggle, creating a strong correlation driven by general rule-use ability and mapping efficiency rather than a specific switch mechanism. While our data point towards a shared foundation, its precise nature – whether a specific flexibility mechanism or a more general rule-use capacity – remains an open question.

In conclusion, by employing the cued, trial-by-trial methodology, this study offers the first direct comparison of the group-level developmental patterns of executive and semantic flexibility. Our findings are consistent with a model where a shared cognitive foundation underpins both abilities, as observed in adult findings²², but where performance may be shaped by distinct developmental timelines for its subcomponents and modulated by domain-specific factors. These findings highlight that a complete understanding of adaptive behavior requires focusing on the dynamic interplay between maturing general control processes and specialized knowledge systems.

Acknowledgements: We thank Marina Mancuso and Chiara Maria Migliorin for their assistance with data collection. We are also deeply grateful to the children, parents, and teachers who participated in this study for their invaluable collaboration.

Competing interests: The authors declare no competing interests.

Data Availability Statement: The datasets generated and analyzed during the current study are available in the OSF repository, <https://osf.io/zr4dn/overview>.

Funding: This study was supported by the European Union (HORIZON-MSCA-2023-PF-01-01, CTRL-ALT-DEV, Grant 101150190 to MM and ERC-2021-STG, IN-MIND, Grant 101043216 to SB-V). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

1. Chiew, K. S. & Braver, T. S. Context Processing and Cognitive Control: From Gating Models to Dual Mechanisms. in *The Wiley Handbook of Cognitive Control* (ed. Egner, T.) 143–166 (Wiley, 2017). doi:10.1002/9781118920497.ch9.
2. Egner, T. Principles of cognitive control over task focus and task switching. *Nat Rev Psychol* **2**, 702–714 (2023).
3. MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*(2), 163.
4. Viviani, G., Visalli, A., Finos, L., Vallesi, A. & Ambrosini, E. A comparison between different variants of the spatial Stroop task: The influence of analytic flexibility on Stroop effect estimates and reliability. *Behav Res* **56**, 934–951 (2023).
5. Sulpizio, S., Job, R., Leoni, P. & Scaltritti, M. Prepotent task-irrelevant semantic information is dampened by domain-specific control mechanisms during visual word recognition. *Quarterly Journal of Experimental Psychology* **75**, 390–405 (2022).
6. Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.
7. Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874.
8. Jefferies, E. The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS. *Cortex* **49**, 611–625 (2013).
9. Montefinese, M., Hallam, G., Thompson, H. E. & Jefferies, E. The interplay between control processes and feature relevance: Evidence from dual-task methodology. *Quarterly Journal of Experimental Psychology* **73**, 384–395 (2020).
10. Lambon Ralph, M. A., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nat Rev Neurosci* **18**, 42–55 (2017).
11. Reilly, J. et al. What we mean when we say semantic: Toward a multidisciplinary semantic glossary. *Psychon Bull Rev* <https://doi.org/10.3758/s13423-024-02556-7> (2024) doi:10.3758/s13423-024-02556-7.

12. Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z. & Wagner, A. D. Dissociable Controlled Retrieval and Generalized Selection Mechanisms in Ventrolateral Prefrontal Cortex. *Neuron* **47**, 907–918 (2005).
13. Davey, J. *et al.* Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *NeuroImage* **137**, 165–177 (2016).
14. Duncan, J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences* **14**, 172–179 (2010).
15. Thompson, H. E. *et al.* The contribution of executive control to semantic cognition: Convergent evidence from semantic aphasia and executive dysfunction. *Journal of Neuropsychology* **12**, 312–340 (2018).
16. Almaghyuli, A., Thompson, H., Lambon Ralph, M. A. & Jefferies, E. Deficits of semantic control produce absent or reverse frequency effects in comprehension: Evidence from neuropsychology and dual task methodology. *Neuropsychologia* **50**, 1968–1979 (2012).
17. Montefinese, M., Hallam, G., Stampacchia, S., Thompson, H. E. & Jefferies, E. Deficits of semantic control disproportionately affect low-relevance conceptual features: evidence from semantic aphasia. *Aphasiology* **35**, 1448–1462 (2021).
18. Grégoire, C., & Majerus, S. (2023). Resisting Visual, Phonological, and Semantic Interference—Same or Different Processes? A Focused Mini-Review. *Psychologica Belgica*, 62(1), 44.
19. Invernizzi, S., Gilis, S., Bodart, A., Lefebvre, L., & Loureiro, I. S. (2026). The processing of ambiguous words without context: Semantic activation, inhibition, selection and controlled retrieval in younger and older adults. *Acta Psychologica*, 263, 106221.
20. Grégoire, C., Attout, L., Phillips, C., Rifon, L., Hody, L., & Majerus, S. (2025). The neural specificity of interference resolution in phonological, semantic, and visual domains at different ages. *Journal of Cognitive Neuroscience*, 37(2), 345-371.
21. Gonzalez Alam, T. R., Cruz Arias, J., Jefferies, E., Smallwood, J., Leemans, A., & Marino Davolos, J. (2024). Ventral and dorsal aspects of the inferior frontal-occipital fasciculus support verbal semantic access and visually-guided behavioural control. *Brain Structure and Function*, 229(1), 207-221.
22. Viviani, G. *et al.* Exploring semantic and executive flexibility interplay in task switching. *Sci Rep* **15**, 31609 (2025).
23. Ruffini, C., Traverso, L., Marzocchi, G. M. & Pecini, C. Enhancing preschoolers' cognitive flexibility and pre-literacy skills through dialogic reading: evidence from an Italian teacher-led intervention. *Early Child Development and Care* **195**, 847–868 (2025).

24. Buttelmann, F. & Karbach, J. Development and Plasticity of Cognitive Flexibility in Early and Middle Childhood. *Front. Psychol.* **8**, 1040 (2017).
25. Crone, E. A., Donohue, S. E., Honomichl, R., Wendelken, C. & Bunge, S. A. Brain Regions Mediating Flexible Rule Use during Development. *J. Neurosci.* **26**, 11239–11247 (2006).
26. Uddin, L. Q. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat Rev Neurosci* **22**, 167–179 (2021).
27. Schettini, E., Hiersche, K. J. & Saygin, Z. M. Individual Variability in Performance Reflects Selectivity of the Multiple Demand Network among Children and Adults. *J. Neurosci.* **43**, 1940–1951 (2023).
28. Deak, G. O. The Growth of Flexible Problem Solving: Preschool Children Use Changing Verbal Cues to Infer Multiple Word Meanings. *Journal of Cognition and Development* **1**, 157–191 (2000).
29. Deák, G. O., Ray, S. D. & Pick, A. D. Effects of age, reminders, and task difficulty on young children's rule-switching flexibility. *Cognitive Development* **19**, 385–400 (2004).
30. Chami, S., Charalambous, C., Knijnik, S. R. & Docking, K. Language and executive function skills as predictors of semantic fluency performance in pre-schoolchildren. *International Journal of Speech-Language Pathology* **24**, 626–635 (2022).
31. Ruffini, C., Osmani, F., Bigozzi, L. & Pecini, C. Semantic fluency in 3-6 years old preschoolers: which executive functions? *Child Neuropsychology* **30**, 563–581 (2024).
32. Memisevic, H., Bisevic, I. & Pasalic, A. Developmental trends in semantic fluency in preschool children. *Cogent Psychology* **4**, 1403064 (2017).
33. Kavé, G., Kigel, S. & Kochva, R. Switching and clustering in verbal fluency tasks throughout childhood. *Journal of Clinical and Experimental Neuropsychology* **30**, 349–359 (2008).
34. Betz, N. & Coley, J. D. Development of Conceptual Flexibility in Intuitive Biology: Effects of Environment and Experience. *Front. Psychol.* **11**, 537672 (2020).
35. Finger, H., Goeke, C., Diekamp, D., Standvoß, K. & König, P. LabVanced: A unified JavaScript framework for online studies. in 1–3 (University of Osnabrück., 2017).
36. *Handbook of Human Multitasking*. (Springer International Publishing, Cham, 2022). doi:10.1007/978-3-031-04760-2.
37. Jackson, R. L. The neural correlates of semantic control revisited. *NeuroImage* **224**, 117444 (2021).
38. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing. (2021).
39. Baayen, R. H., Davidson, D. J. & Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**, 390–412 (2008).
40. Baayen, R. H. & Milin, P. Analyzing reaction times. *Int. j. psychol. res.* **3**, 12–28 (2010).

41. Rabbitt, P. M. Errors and error correction in choice-response tasks. *Journal of Experimental Psychology* **71**, 264–272 (1966).
42. Westfall, J., Kenny, D. A. & Judd, C. M. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General* **143**, 2020–2045 (2014).
43. Bartoń, k. MuMIn: Multi-Model Inference. (2025).
44. Kim, S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *CSAM* **22**, 665–674 (2015).
45. Buchanan, E. M. *et al.* Measuring the semantic priming effect across many languages. *Nat Hum Behav* <https://doi.org/10.1038/s41562-025-02254-x> (2025) doi:10.1038/s41562-025-02254-x.
46. Montefinese, M., Zannino, G. D. & Ambrosini, E. Semantic similarity between old and new items produces false alarms in recognition memory. *Psychological Research* **79**, 785–794 (2015).
47. Joulin, A. *et al.* FastText.zip: Compressing text classification models. Preprint at <https://doi.org/10.48550/ARXIV.1612.03651> (2016).
48. Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. Enriching Word Vectors with Subword Information. Preprint at <https://doi.org/10.48550/ARXIV.1607.04606> (2016).
49. Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T. Learning Word Vectors for 157 Languages. Preprint at <https://doi.org/10.48550/ARXIV.1802.06893> (2018).
50. Brysbaert, M. & Stevens, M. Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition* **1**, 9 (2018).
51. Lakens, D. Equivalence Tests: A Practical Primer for *t* Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science* **8**, 355–362 (2017).
52. Diamond, A. Executive Functions. *Annu. Rev. Psychol.* **64**, 135–168 (2013).
53. Miyake, A. *et al.* The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology* **41**, 49–100 (2000).
54. Zelazo, P. D. Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Developmental Review* **38**, 55–68 (2015).
55. Hills, T. T., & Kenett, Y. N. (2022). Is the mind a network? Maps, vehicles, and skyhooks in cognitive network science. *Topics in Cognitive Science*, 14(1), 189–208.
56. Marko, M., & Riečanský, I. (2021). The structure of semantic representation shapes controlled semantic retrieval. *Memory*, 29(4), 538–546.
57. Michalko, D., Marko, M., & Riečanský, I. (2023). Response modularity moderates how executive control aids fluent semantic memory retrieval. *Memory*, 31(6), 871–878.

AUTHOR CONTRIBUTIONS STATEMENT

Giada Viviani: Data Curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - Original Draft, Writing - Review & Editing. **Ettore Ambrosini:** Data Curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing. **Annamaria Porru:** Resources, Writing - Review & Editing. **Silvia Benavides-Varela:** Funding acquisition, Writing - Review & Editing. **Erin M. Buchanan:** Writing - Review & Editing. **Irene Di Pietro:** Investigation, Writing - Review & Editing. **Daniela Lucangeli:** Writing - Review & Editing. **Maria Montefinese:** Conceptualization, Data Curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing - Original Draft, Writing - Review & Editing.

ARTICLE IN PRESS