Computers & Education 210 (2024) 104961

Contents lists available at ScienceDirect

Computers & Education

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

The cognitive effects of computational thinking: A systematic review and meta-analytic study

Chiara Montuori [a], Filippo Gambarota [a], Gianmarco Altoé [a], Barbara Arfé [a,b,*]

[a] Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy
[b] Human-Inspired Technology Research Centre, University of Padova, Padova, Italy

ARTICLE INFO

Keywords:
Coding
Computational thinking
Executive function
Problem solving
Children
Intervention

ABSTRACT

In this paper, we review and meta-analyze the findings of experimental studies published between 2006 and 2022 that examined the effects of coding and programming interventions on children's core and higher order executive functions (response inhibition, working memory, cognitive flexibility, planning and problem solving). The systematic review and meta-analysis aimed to address three research questions: 1) Which executive functions are most impacted by the teaching of CT? 2) Which instructional modality (educational robotics/virtual coding/unplugged coding) is most effective in enhancing executive function skills in learners aged 4–16 years? and 3) Does the cognitive effectiveness of coding vary with children's age? A total of 19 studies with 1523 participants met the selection criteria for the systematic review. The meta-analysis included 11 of those studies. The results reveal beneficial effects of structured virtual and tangible coding (educational robotics) activities for preschoolers and first graders, and significant effects of more unstructured virtual coding activities (e.g., Scratch-based) for older students. A multivariate fixed-effects model meta-analysis shows that the teaching of coding significantly improves *problem-solving* with the highest effect ($d_{ppc2} = 0.89$), but also *planning* ($d_{ppc2} = 0.36$), and *inhibition* and *working memory* with lower effects ($d_{ppc2} = 0.17$, $d_{ppc2} = 0.20$).

1. Introduction

Computation thinking (CT) is the mental ability to apply the concepts and reasoning typical of computing and computer science to solve problems (Wing, 2006). Thinking computationally entails developing four main component skills (a) the ability to analyze problems and decompose them into elements or parts (analytical thinking); (b) the ability to plan a sequence of actions or steps to get to the problem solution (algorithmic thinking); (c) the ability to monitor and correct errors in the execution of the plan (debugging, Flórez et al., 2017; Román-González et al., 2017; Strawhacker & Bers, 2019); and (d) the ability to identify the most relevant aspects of the problem and generalizable algorithms (abstraction), which allow applying to other problems what has been learned (Moreno-León et al., 2016; Román-González et al., 2017; Yaşar, 2018).

Since computational thinking encompasses not only proficiency in computer science methods but also domain-general problem-solving abilities, such as analyzing problems and planning, the development of computational thinking skills intersects with that of 21st-century competencies such as digital literacy skills(e.g., programming or coding) and with foundational cognitive skills like analytical thinking, planning, and the ability to inhibit impulsive responses (Arfé et al., 2020). The growing spread of CT and

* Corresponding author. University of Padova, Via Venezia 8, 35130, Padova, Italy.
E-mail address: barbara.arfe@unipd.it (B. Arfé).

https://doi.org/10.1016/j.compedu.2023.104961
Received 2 January 2023; Received in revised form 18 September 2023; Accepted 5 November 2023
Available online 18 November 2023

programming into compulsory education (Zhang & Nouri, 2019) is thus supported by the idea that learning to think computationally is important not only for preparing students in the field of computer science (Nardelli, 2019; Wing, 2006) but also, more broadly, for providing them with a general cognitive toolkit to approach and solve everyday problems (Chen et al., 2017; Feurzeig & Papert, 2011; Nardelli, 2019; Wing, 2006).

Computer programming and code writing are two means of CT and two instrumental skills through which CT is taught and practiced in schools (Lye & Koh, 2014). They include the skills to create, modify, and evaluate codes and the knowledge about programming concepts and procedures. In school settings, children are commonly taught CT through educational robotics (ER), or virtual coding activities, which both involve programming and coding by means of technologies, or through unplugged coding without the use of technology. ER consists in developing a program or code string to give instructions to a robot, so that it can perform specific actions or achieve goals in a physical environment (Di Lieto et al., 2020a,b; Chen et al., 2017; Keren & Fridin, 2014). Unplugged coding consists of tangible, paper and pencil or physical activity, by which children learn the basic concepts and procedures of CT and programming, whereas, virtual coding involves the development of a computer program to teach a computer, or a virtual character, to achieve a goal in a virtual environment (Arfé et al., 2019; Fessakis et al., 2013; Kalelioglu, 2015; Zhang et al., 2014). Although in this systematic review and meta-analysis we will refer to programming and coding as two means through which CT is introduced and taught in schools, this does not imply assuming that coding and programming equal computational thinking nor that they always involve it. However, when children in school learn to develop programs for "instructing" a computer, another child or a robot to solve a problem, they also learn to solve CT problems themselves, articulate and clarify their thoughts and plans, generate a clear sequence of commands (the code) and test their hypotheses, processes that can stimulate the development of other cognitive and metacognitive skills (Clements & Nastasi, 1999; Fessakis et al., 2013).

Past studies have reported positive cognitive effects of these CT activities especially in the domain of children's executive functions (EFs, Arfé et al., 2019; Di Lieto, Pecini, et al., 2020). EFs are a complex set of cognitive skills related to goal setting and the performance of goal-directed behaviors, and thus crucial for self-regulation and academic achievement. Although models of EFs may differ in the EF skills they focus on, there is general consensus on the distinction between core EFs (working memory, response inhibition and shifting or cognitive flexibility), and complex or higher-order EF skills, like planning and problem solving (Diamond & Ling, 2016).

Children's command over, or inhibition, of impulsive responses, their working memory (WM) capacities, encompassing their ability to monitor and update temporarily stored information, and their cognitive flexibility which entails switching perspectives, shifting attention between mental sets or tasks, and readily adjusting behavioral responses to different tasks or environments represent core executive functions (Anderson, 2002; Gioia et al., 2000; Miyake et al., 2000) from which higher-order executive functions, such as planning, problem-solving, and reasoning develop (Diamond & Ling, 2016; Miyake et al., 2000; Thayer & Lane, 2000). Welsh and Pennington (1988), for instance, focusing on higher order EF skills, describe executive functioning as the ability to maintain an appropriate problem-solving set for achieving a goal, and define it as a complex skill set that encompasses not only planning skills and a mental representation of the task and of the outcome, but also inhibitory control, that is, the capacity to inhibit or defer prepotent responses.

As higher order EF skills like problem solving involve core EFs, practicing and enhancing these higher order skills through computational thinking (CT) can impact on their core EF components too. In addition, since core EFs also underpin the ability to achieve academic outcomes in reading (Meixner et al., 2019; Nouwens et al., 2021), writing (Altemeier et al., 2008; Salas & Silvente, 2020), and mathematics (Welsh et al., 2010), improvements in problem solving skills could have transfer effects on academic skills (Scherer et al., 2019).

Although recent experimental studies provide evidence of a causal link between the teaching of CT and the improvement of EF skills in children (e.g., Arfé et al., 2019; Arfé et al., 2020; Di Lieto, Pecini, et al., 2020), there remains uncertainty regarding the consistency, robustness, and generalizability of these effects across different (higher order and core) executive functions. Moreover, it remains unclear whether the different instructional tools employed to teach CT skills, such as educational robotics (ER), virtual coding or unplugged coding, yield different effects. At this early stage of research, there are very few experimental studies that have directly compared the efficacy of these different types of intervention (e.g., Çınar & Tüzün, 2021). Thus, the only way to compare the efficacy of these types of CT programs is through systematic reviews or meta-analyses, which are now urgent because several countries are currently integrating CT into their school curriculum. Discussing evidence coming from experimental studies is important to allow educators and policymakers to make informed instructional decisions.

Attempts to conduct rigorous systematic reviews or meta-analyses on the cognitive effects of CT have been limited. What is even more important is that, to the best of our knowledge, none of the existing systematic reviews or meta-analyses have examined the causal relationship between CT and executive functions. Most of the current systematic reviews and meta-analyses have focused on one hand on studies that did not allow to draw robust inferences on causal effects, on the other hand they focused on a rather large and heterogeneous set of cognitive functions and academic skills (see Liao, 2000; Liao & Bright, 1991; Scherer et al., 2019; Scherer et al., 2020). The present study contributes to filling this literature gap and advancing our understanding of the beneficial effects of CT on executive functions, a set of cognitive skills that are especially crucial for academic performance and adaptive behavior. This investigation can increase our theoretical comprehension of CT, as observing transfer effects between different domains, like CT and EFs, provides insights into the shared cognitive foundations that underpin their connection. Determining the impact of CT interventions on children's EFs has also practical implications. By identifying the executive functions that are most influenced by computational thinking (CT), we can determine the optimal age or grade level for implementing CT programs within the school curriculum. Indeed, different EF skills have different windows of plasticity (e.g., Davidson et al., 2006; Ganesan & Steinbeis, 2022).

1.1. The causal link between CT and cognitive abilities

Recent years have witnessed a steady increase in the number of studies exploring the association between CT and children's cognitive abilities (e.g., Arfé et al., 2019; Finke et al., 2022; Polat et al., 2021; Román-González et al., 2017). The research is divided into correlational and intervention studies. Correlational studies identify an association between CT and cognitive functions, but only intervention studies can bring causal evidence on the cognitive effects of CT.

Unfortunately, most of the studies published on the topic are correlational. They have documented associations between CT and a wide range of cognitive and academic skills, including nonverbal intelligence (Marinus et al., 2018), visuo-spatial skills (Finke et al., 2022; Román-González et al., 2017; Tsarava et al., 2019), and mathematical cognition and mathematical skills (Gerosa et al., 2021; Liu et al., 2019; Román-González et al., 2018; Tsarava et al., 2019). The correlations observed are generally moderate (e.g., Román--González et al., 2017) or weak (e.g., Román-González et al., 2018). although stronger associations are reported between CT and problem-solving (Polat et al., 2021; Román-González et al., 2017). The limitation of these studies is that, although correlational studies are sometimes discussed as if they show causality, causality cannot be determined by correlations.

To investigate the causal effects of CT researchers have carried out intervention studies (see Arfé et al., 2019, 2020; Di Lieto, Pecini, et al., 2020). In general, these studies report positive effects of ER and virtual coding activities on first graders' EFs, specifically working memory (Di Lieto, Pecini, et al., 2020), inhibitory control (Arfé et al., 2019, 2020; Di Lieto, Pecini, et al., 2020), and planning (Arfé et al., 2020). However, findings vary between studies, particularly when higher order EFs, like problem-solving, are considered. For instance, Çakır et al. (2021) and Nam et al. (2019) found that experiencing coding improved preschoolers and first graders' problem solving, whereas Çiftci and Bildiren (2020) reported insignificant effects of coding on the problem-solving skills of 4–5-year-old preschool children.

One of the problems in comparing the results of these studies is, however, that the instructional tools used in the interventions were different: to train children's problem solving skills, Çakır et al. (2021) and Nam et al. (2019) used ER, whereas Çiftci and Bildiren (2020) employed game-based drag-and-drop exercises from the code.org platform, https://code.org/, that is, virtual coding. In ER interventions, children are engaged in tangible hands-on coding activities, and they materially interact with robots designing, assembling, and programming them to perform actions in a physical learning environment (Kazakoff & Bers, 2014). This experience is more closely related to children's everyday sensorimotor and concrete experience of the world than virtual coding, which involves children learning to program sprites (virtual objects) to perform actions in a virtual world presented on a screen, typically using block-based visual programming and a computing device (e.g., Kalelioğlu, 2015; Sáez-López et al., 2016). This second learning experience implies greater abstraction and perspective-taking skills: to drive the sprite toward the objective, the child must take the perspective of the virtual character on the bi-dimensional screen. From the age of 3 to 6, children's reasoning and problem solving are more closely related to their concrete experience of objects and the physical world (Barrouillet, 2015; Ping & Goldin-Meadow, 2008). However, with the beginning of schooling they increasingly use mental representations to test their hypotheses; therefores their problem solving becomes more abstract (Novack et al., 2014). These developmental changes may affect also their ability to learn CT concepts through different tools, allowing more or less tangible or concrete experiences.

Another dimension across which CT intervention programs vary is the degree to which coding or programming activities are structured (Lee et al., 2013; Socratous & Ioannou, 2021). Some instructional interventions ask children to solve structured computational problems, often consisting in logic games with a single correct solution (e.g., code.org, Arfé et al., 2019; Çiftci & Bildiren, 2020), whereas other interventions focus on unstructured and ill-defined problems (e.g., Erol & Çırak, 2022) such as creating a story or designing a video-game. These problems involve several possible solutions and steps (Zhang & Nouri, 2019).

It is unclear whether these different instructional methods have similar effectiveness on the development of children's EFs and whether it depends on children's age. Although all EFs have an extended window of developmental plasticity that span from early childhood to late adolescence, core and higher order EFs have different developmental trajectories and developmental peaks. Consequently, they could also be more sensitive to the effects of interventions during different time periods. For instance, the most intensive changes in core EFs-inhibitory control, working memory and cognitive flexibility- occur during the preschool period, from 3 to 6 years (Carlson, 2005; Scionti et al., 2020; Traverso et al., 2015) and in the transition to elementary school (Garon et al., 2008; Macdonald et al., 2014; Zelazo et al., 2003), although they continue to develop until adolescence (Brocki & Bohlin, 2004; Huizinga et al., 2006).

For higher order EFs, such as planning, the greatest developmental changes occur later. Planning, for example, begins to develop between the ages of 5 and 6 (Usai et al., 2014), with a steep developmental growth curve from the age of 6 to 9 (McGuckian et al., 2023) and a second remarkable developmental shift from the age of 9 to 15–17 years, related to the development of prefrontal regions (Luciana et al., 2009). Stimulating the child's EFs through CT within these developmental windows may lead to the largest effects. On the other hand, a certain development of EF skills may be necessary to perform CT tasks, or benefit from CT interventions. For instance, problems whose solution requires more steps, or longer algorithms, require sufficient command over impulsive responses and working memory skills.

1.2. Prior meta-analytic studies on the cognitive effects of CT/programming

The meta-analytic studies evaluating the cognitive effects of CT and programming are very few (Liao, 2000; Liao & Bright, 1991; Scherer et al., 2019, 2020). To our knowledge, only three meta-analyses have addressed the cognitive effects of coding/computer programming. Liao and Bright (1991) examined 65 studies targeting the relationship between computer programming and cognitive skills such as planning, reasoning skills, and metacognition, without restrictions on grade level (age) or study design (experimental or not). They included both experimental and nonexperimental studies in their meta-analysis and did not differentiate between cognitive functions, but considered only a general cognitive effect of computer programming. Their meta-analysis reports a moderate effect of learning computer programming on students' cognitive skills ($d = 0.41$). Liao (2000) performed a second, updated meta-analysis summarizing the results of 22 studies, with participants from preschoolers to college students, targeting a rather heterogenous set of cognitive skills, such as critical thinking, creative thinking, metacognitive skills, problem solving, spatial skills, and conceptual transfer. Their meta-analysis revealed large effects of computer programming on programming skills ($d = 2.48$), but only moderate effects on critical thinking, reasoning, and spatial skills ($d = 0.37$–0.58), and insignificant effects on creative thinking ($d = -0.13$). Scherer et al. (2019) obtained similar findings in a meta-analysis of 105 studies, selecting studies published since 1965. Also this meta-analysis considered a broad range of cognitive skills and age levels, covering an age span from preschool to university years. Based on the idea that learning computer programming can bring cognitive benefits in several domains, the authors examined the effects of programming on a broad range of skills, assessing both *near transfer* and *far transfer* effects. *Near transfer* effects are those of computer programming interventions on programming or coding, whereas *far transfer* effects are those of computer programming/coding interventions on cognitive skills less strongly related to coding, such as spatial skills, reasoning and metacognition, or academic achievements, like achievements in mathematics or literacy. Similar to Liao and Bright (1991), also in this meta-analysis the participants in the primary studies varied in agefrom preschool years to university. Moreover, although Scherer et. applied stricter criteria for study inclusion than Liao and Bright (1991), they considered eligible both standard experimental studies with a pre-posttest randomized control trial design and quasi-experimental studies that reported posttest only measures.

Scherer et al.'s (2019) meta-analytic study confirmed an averagely moderate effect of programming on cognitive abilities ($g = 0.49$). The findings revealed larger *near transfer* effects (e.g., to programming, $g = 0.75$) than *far transfer* effects (e.g., to different cognitive abilities, $g = 0.47$) . There considerable variation in *far transfer* effect sizes, which could be explained by the various cognitive skills assessed. Separate meta-analyses for each cognitive skill revealed the beneficial effects of learning computer programming, with effect sizes ranging from $g = 0.73$ for creative thinking to $g = 0.37$ for spatial skills. Moderator analyses revealed no significant moderating effect of age.

This meta-analysis has significantly contributed to advancing our knowledge on the effects of CT, revealing that certain cognitive functions are more responsive to coding/programming interventions compared to others. However, the inclusion of several quasi-experimental studies with nonequivalent pretest-posttest measures or posttest-only measures in the computation of the effect sizes may have affected the findings obtained. Without equivalent pretest and posttest measures, it is difficult to determine whether the observed improvement is a result of the intervention or is influenced by the different tasks performed or type of skills assessed in the posttest and pretest. It may indeed be that the posttest tasks or skills are simply easier than the pretest ones. These quasi-experimental studies provide weaker and less reliable evidence of the causal link between the learning of computer programming and cognitive development.

In a subsequent meta-analysis, Scherer and coll. (2020) explored also the effectiveness of different instructional approaches or programming tools (visual programming and robotics) on the acquisition of programming knowledge and skills. Yet, in this study they did not examine the cognitive effects of the different instructional methods.

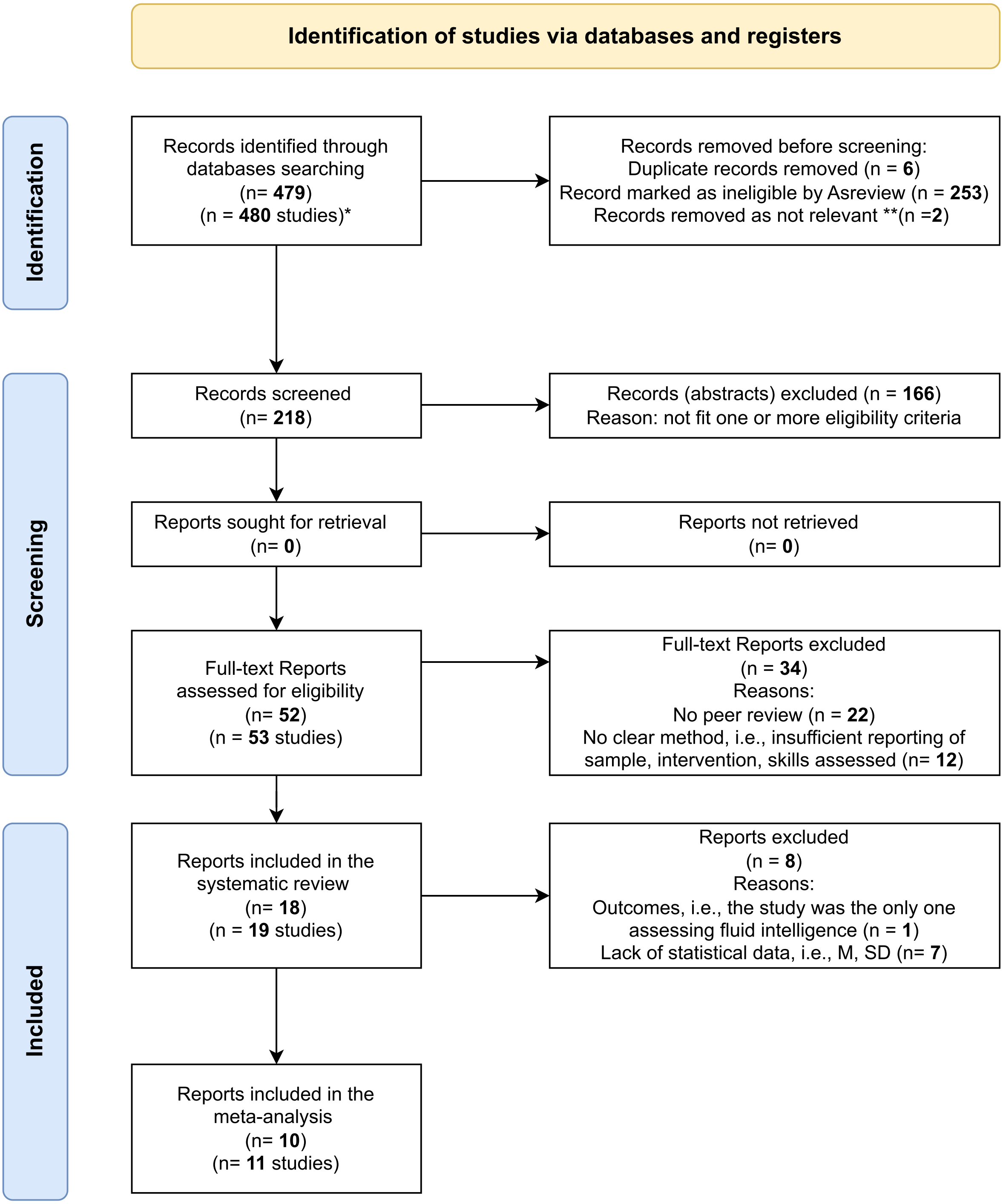
2. The systematic review and meta-analytic study

In instructional research, systematic-reviews and meta-analyses provide the best evidence to support instructional practice (Crocetti, 2016). Combining data from several independent studies increases the statistical power of the analysis and the precision of the effect estimates. The reliability of a meta-analysis relies, however, not only on the on the quantity of studies included but also on the robustness and reliability of their original findings from those studies. For instance, when the aim is to assess instructional effects, including in the meta-analysis studies that lack experimental rigor may represent, as noted earlier, an important limitation. If the research focus is on causality, rigorous hypotheses testing requires controlled experimental studies, consisting of randomized or cluster randomized trials (CRTs; CONSORT guidelines, Campbell et al., 2004).

Although a substantial number of studies in the field of CT report that learning to code improves CT (see, for example, Bers et al., 2014, or Özcan et al., 2021) and EFs (e.g., Arfé et al., 2019; Di Lieto, Pecini, et al., 2020; Çakır et al., 2021), only some of these are experimental or (cluster) randomized trials (Arfé et al., 2019; Di Lieto, Pecini, et al., 2020; Özcan et al., 2021). In contrast to Scherer et al. (2019), in the present systematic review and meta-analysis we thus consider only the experimental trials that effectively controlled for the potential impact of repeated testing and practice effects when evaluating the effectiveness of the intervention.

This systematic review and meta-analysis also focus on a narrower set of cognitive skills than those considered by Liao (2000) and Scherer et al. (2019): core and higher order EFs, like response inhibition, WM, cognitive flexibility, planning, and problem solving. Since these skills underpin children's performance across several complex cognitive and academic tasks, determining the effects of CT and programming on these EFs can help explain why the benefits of CT/programming documented by prior meta-analyses are so broad but also vary across tasks. A further goal of the present systematic review and meta-analysis is to compare the effects of using different instructional tools (e.g., ER and virtual coding) in the teaching of CT/programming.

Unlike previous meta-analytic studies that have explored the effects of computational thinking and coding across a wide age range, spanning from preschool to university years (Liao & Bright, 1991; Scherer et al., 2019), the focus of this systematic review and meta-analysis is on the preschool and school years (a range from 4 to 16 years of age), a time period which represents a critical window for investigating the effects of EF-focused interventions (Luciana & Nelson, 2002; Traverso, Viterbori, & Usai, 2015).



Note. * the number of studies does not correspond to the number of reports bacause one report has two studies;
** the study did not address the target topic of the present systematic review

Fig. 1. Flow diagram describing the literature search and the selection of eligible studies (adapted from the PRISMA Statement; Page et al., 2021).

2.1. Research questions

We addressed three research questions:

1) Which EFs are most impacted by the teaching of CT/coding?
2) Does the cognitive effectiveness of CT vary with children's age?
2) Which instructional modality (educational robotics/unplugged coding/virtual coding) is the most effective in enhancing EF skills in learners aged 4–16 years?

The first research question (*Which EFs are most impacted by the teaching of CT?*) was addressed both through a systematic-review and meta-analysis, whereas the second and third research questions (*Does the cognitive effectiveness of CT vary with children's age?* and *Which instructional modality is most effective?*) were addressed only by a systematic review of the literature, because the number of studies considered in the meta-analysis was insufficient to statistically test the effects of age and intervention type through moderator analyses.

2.2. Method

The systematic review and meta-analysis were conducted in accordance with the guideline of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020; Page et al., 2021).

2.2.1. Literature search, screening and eligibility criteria

Studies were selected according to the following keywords: ("executive functions" OR "cognitive control" OR "cognitive abilities" OR "planning" OR "inhibition" OR "working memory" OR "cognitive flexibility" OR "shifting" OR "problem solving") AND ("computational thinking"* OR "coding" * OR "programming" OR "educational robotics") using Scopus, ERIC, Web of Science, PsychInfo, Google Scholar, ACM Digital Library, and IEEE Xplore as databases. Both research articles and conference papers were included in the study, provided they were peer reviewed.

To perform our literature search and identify eligibility criteria we followed the Population, Intervention, Comparison, Outcome (PICO) guidelines (Methley et al., 2014). The following eligibility criteria were applied:

- Participants (P): the study participants had to be children or adolescents, between 4 and 16 years old.
- Interventions (I): CT and coding interventions could be based on tangible (ER, unplugged coding) or virtual coding (visual programming) instructional activities.
- Comparison (C): the study had to be an experimental, randomized trial or a cluster randomized trial (CRT), or a matched group trial. Thus, only studies involving equivalent pretests-posttests and an experimental and a control/comparison group randomly assigned to an experimental or control condition, or matched based on specified criteria, were considered eligible.
- Outcomes (O): dependent measures (outcomes) were executive functions (e.g., WM, response inhibition, shifting, planning or problem-solving). Fluid intelligence, which relies on cognitive flexibility, was also considered, but only for the systematic review because just one study was found for this outcome. Additional eligibility criteria were:
  - only papers written in English were considered.
  - only peer reviewed publications ranging from 2006 (when Wing provided the first definition of computational thinking) to 2022 were considered.
  - the studies should report sufficient information about the intervention tested (e.g., structure, kind of activities).
  - to be included in the meta-analysis all studies should specify sample size, participants' age, and provide effect sizes or means and standard deviations of the pretest and posttest performance.

The PRISMA diagram (Page et al., 2021) in Fig. 1 shows the literature selection process. A total of 479 records were identified based on our literature search. An open-source software for systematic reviews, ASReview (https://asreview.nl/), was used for the initial screening of this literature. Two examiners (the first author and a master's student) checked all the abstracts of the records removed by ASReview. After removing duplicates and the records marked as irrelevant by ASRview, the abstracts of all remaining records (n = 218) were independently screened by the same two judges. Based on this further screening, 166 publications were excluded as they did not meet one or more eligibility criteria. There were 52 papers (53 studies, as one paper presented two studies) assessed for eligibility, after the initial screening; 22 (mainly conference papers) were subsequently excluded because they were not peer-reviewed studies, and 12 were excluded due to insufficient information on participants (e.g., age or grade level), interventions or the outcomes assessed (specific measures, effect sizes means or standard deviations). From this selection process, 18 papers (19 studies) were included in the systematic review. Due to the strict eligibility criteria, only 10 publications (corresponding to 11 studies) were included in the meta-analysis. Of the eight studies included in the systematic review and excluded from the meta-analysis, seven were excluded due to lack of statistical data (i.e., means and standard deviations: Brown et al., 2008; Çakır et al., 2021; Çiftci & Bildiren, 2020; Çınar & Tüzün, 2021; Demir, 2021; Lai & Yang, 2011; Oluk & Saltan, 2015), and one study was excluded because it was the only one assessing fluid intelligence (Özcan et al., 2021). Although meta-analyses normally involve a larger number of studies, meta-analyses on few studies (10 studies or less) are not uncommon, especially in some research areas, like the assessment of new interventions (Mathes & Kuss, 2018).

Disagreements between the two examiners were solved via discussion with the last author of the study, to reach a consensus. There are various methods for calculating IRR (Belur et al., 2021); we used the percentage of agreement between coders (90% agreement) and the κ statistic measure ($\kappa = 0.85$).

2.2.2. Studies coding and data extraction

Coding for study descriptors and quality indicators was performed by the first author and, subsequently and independently by the last author. The first author read all 19 studies that met the inclusion criteria and classified the studies by intervention type into one of the three following categories: (1) virtual/plugged coding, (2) unplugged coding, and (3) educational robotics. Instructional activities were further coded as: (a) structured, when children were asked to solve well-defined problems with a clear, given goal and correct solution, e.g., giving instruction to a sprite or robot to reach a specific target, or (b) unstructured, when children solved open-ended (or creative) problems with no predefined solution, such as developing a project or designing a new game. There were only two cases of disagreement between the first and last author, which were resolved through discussion.

Each study was coded for thirteen study characteristics, as reported in Table 1: (1) study design (randomized control, matched design), (2) participants' grade level and (3) age group (younger or older), (4) sample size, (5) participants (i.e., with typical development or atypical development), (6) intervention modality (virtual coding, ER, unplugged coding), (7) intervention tool (e.g., code.org, Scratch, LEGO), (8) whether the intervention was structured or unstructured, (9) intervention activities (the specific activities performed during the intervention), (10) type of control group (i.e., passive or active control group), (11) control group activities (e.g., business-as-usual activities or other types of STEM or programming activities), (12) overall intervention length (in minutes), and (13) main outcomes (with effect sizes when available), coded in five categories: problem-solving, planning, working memory, inhibition, and cognitive flexibility. Some studies reported intervention duration in hours, some in minutes, and some in hours per week. Thus, all durations were converted to minutes to have the same unit of measurement across studies.

2.2.3. Quality assessment

Study quality was assessed by means of a checklist (see supplemental material, Quality indicators checklist in the extended table) across seven indicators: (1) quality of the experimental design, (2) presence/absence of an active control group, (3) pretest equivalence or pretest scores controlled in analyses, (4) same pre-posttest measures (5) reliability of the assessment tool (6) method clarity (7) intervention clarity (see Table 1). Each quality indicator was scored as 1 (met) or 0 (not met). If information concerning a specific quality indicator was not provided, it was assumed the indicator was not met, and the study scored 0. A synthesis of the quality assessment is reported in the *tables* folder of the supplemental material (Table quality assessment extended).

2.2.4. Meta-analysis: data extraction

Data extraction was performed by the two examiners who also performed the literature screening. Once the final pool of studies was

Table 1
Definitions for study characteristics and quality indicators.

	Definition
Study characteristics	
Study design	Type of experimental design: experimental or matched design
Grade level	Participants' grade level. If not provided, age was used to determine likely grade
Age group	Younger or older participants
Sample size per condition	Total number of participants in each condition
Participants	Participants' characteristics: typical or atypical development
Intervention modality	Type of intervention: i.e., virtual/plugged coding, unplugged coding, educational robotics
Intervention tool	Type of tool used in the intervention (e.g., code.org, Scratch, KIBO robot)
Interventionstructure	Structured or unstructured (e.g., structured games with one solution, or open ended or creative problem-solving activities)
Intervention activities	Specific activities carried out during each intervention session: e.g., game-design, code.org games
Type of control group	Type of control condition (i.e., passive or active)
Control groupactivities	Specific activities in which the control group was engaged: e.g., business-as-usual activities or other types of STEM or programming activities
Intervention length	Overall intervention length in minutes. Wherever minutes could not be determined, the most informative length unit, such as number of sessions, was used
Main outcome(s)	The dependent measure(s)
Quality indicators	
Quality of the experimental design	True random experiment or matched-group experiment (i.e., random assignment of individual participants or classes to conditions)
Control condition	Is an active or business-as-usual control group present?
Pretest equivalence	Intervention and control participants showed equivalence in the dependent measure (outcome) at the pretest or pretest measures were covaried
Pre-posttest measures	Intervention and control group's dependent measures were assessed at pre and posttest with the same assessment tools
Reliability of the assessment tool	Reliability coefficients for all dependent/outcome measures are reported or the reliability can be inferred by the use of a standardized and validated assessment tool
Method clarity	The study description provides sufficient information on the procedure and the instruments used for cognitive assessment
Intervention clarity	The intervention is explained in detail with reference to its duration, tools used, and the specific activities performed in each session

Table 2
Studies Included in the Systematic review and Meta-analysis (Effect Sizes or Significant Effects are Reported).

Author(s)/ year	Age group	Grade	Sample size	Intervention type	Coding tool	Structured/ Unstructured	Intervention length	Included SR/MA	Effects sizes or Sign. effects
Akcaoglu and Koehler (2014)	Older	Grades 5-8	44	Virtual coding	Microsoft Kubo	U	900 min	SR/MA	Problem-solving (d = 1.05)
Brown et al. (2008)	Older	Grades 5-6	113	Virtual coding	Scratch	U	180 min	SR	Problem-solving
Çınar and Tüzün (2021) [a]	Older	Grade 10	81	Educational robotics	LEGO Mindstorms NXT 2.0	S	2160 min	SR	Problem-solving n.s.
Demir (2021) [a]	Older	Grades 9-11	34	Coding unplugged	Algorithm cards	S	–	SR	Problem-solving
Erol and Çırak (2022) [a]	Older	Grade 6	34	Virtual coding	Scratch	U	1680 min	SR/MA	Problem-solving (η^2 = 0.26)
Lai and Yang (2011)	Older	Grade 6	130	Virtual coding	Scratch	U	–	SR	Problem-solving
La Paglia et al. (2017)	Older	Grades 5-6	60	Educational robotics	LEGO Mindstorms	S	1800 min	SR/MA	Problem-solving
Nam et al. (2010)	Older	Grade 6	60	Virtual coding	Scratch	U	480 min	SR/MA	Problem-solving
Oluk and Saltan (2015) [a]	Older	Grade 6	65	Virtual coding	Scratch	U	720 min	SR	Problem-solving n.s.
Özcan et al. (2021)	Older	Grade 4	174	Virtual coding	Code.org + Scratch	S + U	1200 min	SR	Fluid intelligence (cognitive flexibility) n.s.
Pardamean et al. (2011)	Older	Grade 5	85	Virtual coding	Logo programming	S	640 min	SR/MA	Problem-solving[b]
Arfé et al. (2019)	Younger	Grade 1	80	Virtual coding	Code.org	S	480 min	SR/MA	Inhibition (d = −0.65) Planning (d = 0.95) Problem-solving (**NT**, d = 1.62)
Arfé et al. (2019)	Younger	Grade 2	38	Virtual coding	Code.org	S	480 min	SR/MA	Inhibition (d = −1.05) Planning (d = 0.93) Problem-solving (**NT**, d = 1.91)
Arfé et al. (2020)	Younger	Grade 1	179	Virtual coding	Code.org	S	480 min	SR/MA	Inhibition (d = −0.71) Planning (d = 1.27) Problem-solving (**NT**, d = 1.31)
(Çakir et al., 2021)	Younger	Preschool	40	Educational robotics	LEGO WeDo 2.0	S	1920 min	SR	Problem-solving
Çiftci and Bildiren (2020)	Younger	Preschool	28	Virtual coding	Code.org	S	480 min	SR	Problem-solving n.s.
Di Lieto, Pecini, et al. (2020a,b)	Younger	Grade 1	187	Educational robotics	Bee Bot	S	1200 min	SR/MA	Inhibition (d = 0.69) Working memory (d = 0.65) Cognitive flexibility n.s.
Di Lieto, Castro, et al. (2020a,b)	Younger	Grade 1	42	Educational robotics	Bee Bot	S	1200 min	SR/MA	Inhibition Working memory n.s. Cognitive flexibility n.s.

(continued on next page)

Table 2 (*continued*)

Author(s)/ year	Age group	Grade	Sample size	Intervention type	Coding tool	Structured/ Unstructured	Intervention length	Included SR/MA	Effects sizes or Sign. effects
Nam et al. (2019)	Younger	Preschool	53	Educational robotics	TurtleBot	S	720 min	SR/MA	Problem-solving (η^2 = 0.17)

Note: Age group = the category by which results are reported; Effect sizes = Significant major effects; S = Structured; U = unstructured; NT = near transfer effects; n.s. = non-significant.

[a] The comparison group was an active control group exposed to coding/programming activities (e.g., visual programming tool, learning programming and algorithms during Information Computer Technology class).

[b] Non-equivalent pretest scores were not covaried in the analyses.

selected, one extracted all data (i.e., sample size, means and standard deviations or effect sizes) for the meta-analysis, and the other checked the data extraction subsequently.

In case of papers with missing data, we contacted the study authors before excluding them from the meta-analysis. One paper (Arfé et al., 2019) contained two studies with a different pool of participants. To simplify the data structure, the two experiments were treated as different studies (2019a and 2019b). In the *data* folder of the supplemental material, the full table summarizing the data and outcomes of all included studies is presented (i.e., meta table cleaned).

Table 2 summarizes all included studies, distinguishing between those included in the systematic review only, and those included both in the systematic review and meta-analysis. In the Results section, we provide detailed descriptions for each study.

2.3. Statistical analysis

2.3.1. Effect size computation

We used in the MA the d_{ppc2} effect size measure as Morris (2008) proposed, where the difference between pre-means and post-means for the experimental and control groups is standardized using the pretest standard deviations. The d_{ppc2} has been demonstrated to be the most appropriate measure for this research design. Despite being called *d*, the effect is computed by default (Morris, 2008) by applying the Hedges' correction for small samples. For measures where lower values correspond to better performance (e.g., number of errors) we changed the sign of the effect, thus, positive values always mean that the CT is effective. For computing the effect size sampling variance, we imputed the pre–post correlation because this was missing from the majority of included papers. We decided to use a pre–post correlation of 0.7. We also included a sensitivity analysis using different correlation values (0.5, 0.7, and 0.9) in the supplementary materials.

Between and within studies the same underlying psychological construct could be measured with different instruments. We thus decided to assign the same label to effects referring to the same underlying psychological construct. For example, measures derived from WM, the Backward Corsi Block Tapping subtest (BVN test; Bisiacchi et al., 2005) and Matrix Path (BVS Corsi; Mammarella et al., 2008) were all coded as WM. Similarly, scores at standardized tests, ad hoc problem solving tasks and self-report scales assessing problem solving were coded as problem solving. In this way, we classified the reported outcomes according to the underlying EF: inhibition, WM, shifting, planning, and problem solving. Thus, multiple effect sizes referred to the same construct were aggregated to obtain a single measure. Aggregating multiple statistically dependent effect sizes requires imputing the correlation between different measures. Following Borenstein et al. (2009, pp. 225–233), we used a correlation of 0.5, but with an inverse-variance weighted average (Viechtbauer, 2010). A sensitivity analysis using different correlation values (0.3, 0.5, and 0.7) is presented in the supplemental material.

2.3.2. Statistical model

When multiple outcomes are collected from the same pool of participants, a situation of statistical dependency emerges, which, if ignored, brings strongly biased meta-analytic estimations (Cai & Fan, 2020; Cheung, 2019; Van den Noortgate et al., 2015). The most appropriate approach is to consider the correlation between multiple outcomes using a multivariate meta-analytic (MA) model (Cai & Fan, 2020). Given the limited number of papers, we decided to use a *fixed-effect* model as Cai and Fan (2020) suggest. Despite the *random-effects* model allowing to generalize conclusions at population level, the between-papers variability estimation (i.e., the focus of the *random-effects* model) can be strongly biased with a limited number of papers (Cheung, 2013; Veroniki et al., 2016). The multivariate fixed-effect model estimates the average effect size for each outcome and requires considering the correlation between different measures. Again, the included papers did not report this correlation, thus, we imputed a value of 0.5. As before, we conducted a sensitivity analysis using different values (0.3, 0.5, and 0.7), again presented in the supplemental material.

The MA model was implemented in R (R Core Team, 2021) using the *metafor* package (Viechtbauer, 2010). For hypothesis testing, we used the *omnibus* Wald χ^2 test. For each outcome, we reported the average effect size, standard error, the 95% confidence interval, and the associated Wald z test. Further details about the analytic approach can be found in the supplemental material. All code and data to reproduce the analysis (supplemental material) are available online in the Open Science Framework repository (https://osf.io/uvbcd/).

3. Results

3.1. Characteristics of included studies

The sample comprised 19 studies included in the systematic review, for a total of 1527 participants. Of the primary 19 studies, 11 were also included in the meta-analysis (see Table 2 and Fig. 1). The overall sample size of the MA was of 862 participants ($n = 433$ in experimental conditions and $n = 429$ in control conditions).

The study samples included participants from 4 to 16 years, with a predominance of preschoolers or primary school children (12 studies, 63%). Five studies (26%) involved children from grade 5 to grade 8, and only one study involved older students (10th graders). The sample size of the included studies ranged from 28 to 187, with a majority of studies with samples > 50. Among the studies selected, 12 tested the effects of virtual coding interventions with children from preschool to grade 8, and seven tested the efficacy of tangible coding interventions (ER or unplugged coding activities) with preschoolers and students from grade 1 to grade 10. Only 31% of the interventions consisted of unstructured coding/programming activities (n = 6). They were all addressed to older students (grades 5 to 8) and involved virtual coding. Among the outcome measures, problem solving skills were the most frequently assessed (13 studies, 68%), inhibition skills were assessed in 5 studies, planning was assessed in 3, working memory in 2, cognitive flexibility in 2, and fluid intelligence in one study only.

3.1.1. Quality of the studies

A synthesis of the studies' quality is reported in the extended Table Studies characteristics and quality indicators in supplemental material. The overall quality was high. When all seven quality indicators were considered, the studies satisfied on average 78.20% of the quality criteria, with 8 studies meeting 100% of quality indicators.

3.2. Results of the systematic review

Table 2 summarizes the characteristics of the selected studies and their main outcomes; effect sizes are also reported if available. As in Scherer et al. (2019), we distinguished between *near transfer* effects (i.e., transfer between similar/closely related tasks or skills) and *far transfer* effects (i.e., the transfer between dissimilar tasks, which require different skills or strategies; Perkins & Salomon, 1992). As shown by Table 2, the majority of the studies assessed *far transfer* effects. In the following sections, the results are presented with reference to the three research questions of the study.

3.2.1. Which EFs are most impacted by the teaching of CT?

The systematic review revealed that CT interventions are generally effective in boosting children's EFs. Of the 19 studies examined in this systematic review, only four reported non-significant effects of CT interventions (Oluk & Saltan; 2015; Çiftci & Bildiren, 2020; Çınar & Tüzün, 2021; Özcan et al., 2021).

3.2.1.1. Problem solving. Sixteen out of 19 studies examined the effects of CT programs on children's problem solving, assessed by problem solving tests or ad hoc tasks (e.g., Akcaoglu & Koehler, 2014; Brown et al., 2008; Lai & Yang, 2011; Nam et al., 2010) or self-report problem solving measures (Erol & Çırak, 2022; Oluk & Saltan, 2015). Six of these studies tested the effects of game-design or project-development activities, while the remaining 10 studies explored the efficacy of more structured virtual coding or educational robotics programs.

3.2.1.2. Game-design and project-development interventions. Significant and positive effects of game-design and project-development interventions were found in studies in which students' problem solving skills were assessed by the Program for International Student Assessment (PISA) of the Organisation for Economic Co-operation and Development (OECD, 2013; Akcaoglu & Koehler, 2014; Nam et al., 2010). Akcaoglu and Koehler (2014) proposed to 11-14-year-old students a 15-h game-design activities in which students had to design digital games through Kodu Game Lab, a 3D game development environment for visual programming. The authors found a large effect of the intervention program, Cohen's $d = 1.05$, on participants' problem solving.

Similar findings are reported by Nam et al. (2010), despite their intervention program, based on project-development activities, was approximately half the time that of Akcaoglu and Koehler's (2014): 8-h versus 15-h. The authors engaged 12-year-old students in four weeks project-development activities based on Scratch, an open-source block-based visual programming tool used to create interactive stories and games, finding significant improvements of the participants on PISA problem solving assessment.

Project-development activities based on Scratch result effective also when compared with other experimental trainings (active control groups) and for students with diverse background. Lai and Yang (2011) assessed the efficacy of visual programming activities based on Scratch on sixth grade students' problem solving and reasoning skills assessed by a problem-solving test. During the intervention, the students received instruction on the basic operations and tools of Scratch, scaffolding in problem solving and subsequently developed their own project. The problem-solving skills of the students in the Scratch-based program improved significantly more than those of an active control group, participating in Adobe Flash learning activities.

Brown et al. (2008) tested the effects of four 45-min lessons (approximately 4-h program) based on Scratch on fifth and sixth graders' mathematical problem-solving. All students, predominantly African-American, were from disadvantaged backgrounds and low-income families. The Scratch lessons, designed to introduce the students to efficient and inefficient mathematical problem-solving methods, were focused on learning debugging and loops, two key operations in CT and programming. Students' problem-solving strategies were assessed by ad-hoc designed problem-solving exercises, consisting of mathematics problems that could be solved using an efficient (e.g., multiplication or loop) or inefficient (e.g., repeated addition) strategy. The results again showed that the students in the experimental group improved in solving mathematical problems more than the control group, addressed to standard instructional activities.

Other studies have examined the effects of CT programs on students' self-reported problem-solving. Erol and Çırak (2022) used the Problem-Solving Inventory for Children (PSIC, Serin et al., 2010) to assess changes in the problem-solving of a group of Turkish sixth graders after a 12-week game-design intervention (24-h) with Scratch. The students addressed to the game-design intervention were first introduced to programming basics (e.g., operations, control structures) and the use of Scratch, and then to semi-structured and free game-design activities. Their problem-solving scores were compared against those of an active control group who were also introduced to programming and asked to develop algorithms that solved problems during Information Technologies and Software classes. The findings revealed a significantly larger improvement in the students' approach to problem-solving for the experimental compared to the active control group ($\eta^2 = 0.26$).

Only one study, by Oluk and Saltan (2015), reported unsignificant effects of Scratch-based programs. The authors tested the effects of a 12-h algorithm development Scratch-based instructional program on sixth-grade students' self-reported problem solving. The students in the experimental group (31 participants) learned algorithms with the help of Scratch, whereas the control group (34 participants) received standard curricular instruction on algorithm development. Their approach to problem-solving was assessed by the same problem-solving inventory used by Erol and Çırak (the PSIC, Serin et al., 2010). Neither of the two groups showed improvement in problem solving scores. It must be noted, however, that differently from Erol and Çırak's study, the instructional program tested in this study did not focus on game-design or project-design activities, which are comprehensive problem-solving tasks, but targeted only one of the components of problem solving: algorithmic thinking.

3.2.1.3. Other visual programming/educational robotics and unplugged coding interventions. Other CT interventions based on different visual programming tools, such as *code.org* or *Logo programing*, *unplugged coding*, or *educational robotics* are more structured than game-design or project-development interventions, and for this reason were primarily addressed to younger children, preschoolers, or first or second graders. These interventions too have been proven effective in stimulating children's problem-solving skills.

Five studies tested the effects on problem solving of structured visual programming (i.e., virtual coding) activities (Arfé et al., 2019 Arfé et al., 2020; Çiftci & Bildiren, 2020; Pardamean et al., 2011). In three studies (Arfé et al., 2019; Arfé et al., 2020), involving a total of 297 children, 1st and 2nd graders, Arfé et al. found consistent and significant *near transfer* effects on problem solving of a short, 8-h, structured visual programming intervention based on *code.org*. Like Scratch, *code.org* is a visual programming platform in which children write their code by moving programming blocks. This way,they construct sequences of commands that give instructions to a sprite or a character (Angry bird, a bee, a zombie), which executes them. Differently from Scratch, however, the games typically have a predefined structure and aim at a given objective (e.g., getting to a target or performing a specific action). Although structured, in Arfé et al.'s studies the intervention was designed to cause children to switch between scenarios, programming functions (e.g., loops, debugging) and types of problems to force children to maintain a problem-solving approach. Children's problem-solving abilities were assessed by asking children to solve new *code.org* games, similar to those solved during the instructional program. The performance of the children who received the intervention was compared to that of a wait-list group receiving standard STEM instruction. Effect sizes were large, ranging between $d = 1.31$ to 1.91.

Different results were obtained by Çiftci and Bildiren (2020), who used a problem-solving skill scale (Aydoğan et al., 2012) to test the *far transfer* effects of a structured 8-h virtual coding intervention based on *code.org* on 28 4- to 5-year-old children's problem-solving abilities. The problem-solving scale aimed to assess the abilities shown by children when facing actual real-life problems. Children were shown pictures representing real life problems, explained by short stories that defined the problem, and were asked to find the best answer to the problem. The study showed no significant improvement of children's problem-solving skills. The intervention improved though their logical, nonverbal, cognitive abilities measured by the Raven Colored Progressive Matrices Test.

One of the studies examined (Pardamean et al., 2011), yielded unconclusive findings due to a methodological flaw. The authors assessed fifth graders' creativity and problem-solving skills following a 16 40-min lessons course based on Logo programming. Logo is a simple programming language by which children learn to program generating commands to control the movements of a turtle, a cursor, thereby creating drawings or geometric forms. In the study, the children in the Logo programming intervention worked in peers to solve geometric games. Their creative skills were assessed by a creative thinking figural test and their problem-solving skills were assessed by a logical word test and figural problem-solving test. The results revealed statistically significant effects of the intervention on children's creativity. At the posttest, children in the Logo program were also better in figural problem solving than the control group, who participated in standard ICT (Information and Computer Technology) curricular activities. However, as the experimental group outperformed the control group in figural problem solving also at the pretest and pretest scores were not covaried in the analyses, it is not possible to determine whether the outcome reflected true intervention effects or individual differences between the groups.

Other studies have examined the effects of educational robotics (ER) activities on children or early adolescents' problem-solving (Çakır et al., 2021; Çınar & Tüzün, 2021; La Paglia et al., 2017; Nam et al., 2019). La Paglia et al. (2017), involved 30 sixth graders (10–12 years old) in an extra-curricular ER laboratory of 10 3-h sessions. The laboratory activities employed a LEGO Mindstorms robot kit. Students participating in the laboratories worked in group to build a robot body and generate a program to assign it an artificial intelligence. The effects of the ER intervention were assessed by a metacognitive questionnaire assessing students' ability to apply metacognitive skills to problem solving. The performance of the experimental group was compared to that of a passive control group of 30 students not participating in any extra-curricular activity. The results revealed that the ER laboratory significantly improved the metacognitive control skills related to problem-solving of the experimental group.

When problem-solving abilities are assessed by self-report measures the results are less clear. Çınar and Tüzün (2021) compared the improvement in problem solving of two groups of 10th graders: one participating in ER activities, and the other in object-oriented visual programming (active control group). During the experimental ER intervention, which lasted 12-weeks (approximately 36 h), the students worked in groups using Lego Mindstorms NXT 2.0 to build robots and then manipulated the program to train their programming skills. The comparison, active control, group performed CT activities in a visual programming environment individually or in group. During the practical problem-solving sessions guideline questions were used to scaffold the problem solving of both groups. Problem solving was assessed through the Problem-Solving Inventory (PSI) developed by Heppner (1988), which provides a self-assessment of behaviors and approaches associated with successful problem solving. No significant changes were observed in the problem-solving scores of the two groups.

Only two studies have examined the effects of ER interventions on the problem-solving of preschoolers and first graders (Çakır et al., 2021; Nam et al., 2019). Nam et al. (2019) tested the effects of a 90 min -8-sessions- card-coded robotic course on 5–6-year-old children's mathematical problem solving. Children were divided in two groups: the experimental group participated in the card-coding robotic course, while the comparison group participated in daily school activities and performed problem-solving activities as outlined in the national school curriculum (e.g., board games). A TurtleBot, which is a card-coded robot, was used for the ER intervention. The instructional program consisted of several activities: problem identification, planning with the support of worksheets, coding with cards and observing or evaluating the program outcomes, followed by debugging. Mathematical problem-solving was assessed by an adapted version of the Ward (1993) problem solving instrument. The results revealed that the children in the ER group improved significantly more in problem-solving than controls: The dimension of the effect was large ($\eta^2 = 0.17$).

Similar findings are reported by Çakır et al., (2021), who evaluated the effect of a 32-h ER intervention in which LEGO Education WeDo 2.0 was used to enhance the problem-solving skills of preschoolers. Children in the experimental group were asked to first build a robot, and subsequently to write a code for programming the robot through drag-and-drop block-based programming activities. After completing and executing the program they had to verify the accuracy of the instructions given to the robot, reflecting on the functions of the different code blocks, and on the codes used for the activity. Children in the (active) control group were involved in joint book reading activities, reasoned about the stories read, and performed creativity and categorization activities. Children's problem-solving was measured by the Problem Solving Skill Scale (PSSS, Oğuz & Köksal Akyol, 2015) assessing problem-solving applied to real-life problems. The results showed a significantly greater improvement of problem-solving in the experimental compared to the active control group.

Only one study (Demir, 2021) tested the effects of unplugged coding activities on problem solving. Participants were 34, 14 to 18-year-old, students with mild intellectual disabilities. Demir examined the effects of unplugged coding activities on their problem solving. Special education classes, of approximately 4–6 students each, were randomly assigned to an experimental or control condition. The experimental intervention required the students to play structured games, based on unplugged activities, such as finding solutions to problems presented in stories or Tower of Hanoi problems. Problem-solving skills were assessed by asking to solve everyday like problems (e.g., washing dishes or making pasta problems). The results showed significant effects of the unplugged coding intervention.

3.2.1.4. Planning and core EFs. Six studies considered in this systematic review assessed other EF skills, such as core EFs (inhibition, working memory and cognitive flexibility or switching), planning or fluid intelligence. They involved primarily younger children, from preschool to grade 2, and tested the effects of structured visual programming or ER interventions.

In their studies, Arfé and colleagues (Arfé et al., 2019; Arfé et al., 2020) examined also the effects of *code-org* game-based coding program on first and second graders' cognitive inhibition and planning abilities, finding that the 8-h code.org based intervention produced significant and moderate to large effects on the planning and cognitive inhibition skills of first and second graders. Cognitive inhibition and planning skills were assessed through standardized neuropsychological tests. Inhibition skills were assessed by the square and circle NEPSY-II subtests (Korkman et al., 2007) and a Numerical Stroop test (Batteria Italiana ADHD, BIA, Marzocchi et al., 2010). Planning skills were assessed by the Elithorn maze test (Gugliotta et al., 2009) and the Tower of London, ToL test (Luciana et al., 2009). The effect sizes ranged from $d = 0.65$ for response inhibition to $d = 1.27$ for planning. These results confirmed that even a relatively short structured virtual coding intervention (8 h of code.org-based activities) can boost children's cognitive inhibition and planning abilities.

Virtual coding interventions seem however less effective in stimulating cognitive flexibility. In a randomized trial, Özcan et al.'s (2021) compared the effects of a 10-week (20 h) learn-to-code (virtual coding) program to two control instructional conditions: another STEM comparison treatment, based on mathematics, and a reading program control condition. One-hundred and seventy-four fourth graders from socioeconomically disadvantaged backgrounds were equally distributed and randomly assigned to the three conditions. Their fluid intelligence was assessed with the matrix reasoning task from the Wechsler Abbreviated Scale of Intelligence Measurement (Wechsler, 2011) at pre and posttest. The virtual coding program combined structured visual programming activities with Algo Digital (https://algodijital.com/) and code.org to project-based activities with Scratch. Although computational thinking scores improved significantly only for children in the learn-to-code treatment condition, children's fluid intelligence, a measure of cognitive flexibility (Colzato et al., 2006), improved equally in all groups, indicating that children's gains were unrelated to the intervention.

The few studies that have tested the effects of ER programs on children EFs have shown beneficial effects of ER on working memory and response inhibition skills in younger children (Di Lieto, Pecini et al., 2020). Di Lieto, Pecini, et al. (2020) examined the effects of a structured 20-h ER training on 5 and 6-year-old first graders' inhibition skills, WM, and cognitive flexibility. Cognitive inhibition was assessed by three standardized neuropsychological tests, the NEPSY-II circle and square subtest (i.e., Korkman et al., 2007), the Little Frogs subtest (i.e., BIA; Marzocchi et al., 2010), and Pippo Says test, a modified version of Simon-Says test assessing both inhibition and switching, that is cognitive flexibility (Marshall & Drew, 2014). Working memory was assessed by two visuospatial tasks: the Backward Corsi Block Tapping subtest (BVN test; Bisiacchi et al., 2005) and Matrix Path (BVS Corsi; Mammarella et al., 2008). Children were randomly assigned to an experimental (ER intervention, n = 96) or a control (wait list, n = 91) group condition, in which children participated in daily school activities and received the ER instructional program later. Children in the experimental group were introduced to programming through engaging coding activities with Bee-Bot, an interactive robot with a bee shape that can be programmed to execute movements using buttons on its back. After a familiarization phase with the Bee Bot, children were invited to solve complex visuospatial planning tasks with the robot to stimulate their working memory and inhibition skills. In the last sessions, task switching and inhibition tasks with Bee Bots were targeted. After the 20-h ER training, the intervention group showed significant improvements in WM and inhibition abilities with moderate effect size: $d = 0.63$ for visuo-spatial WM, and from $d = 0.43$ to $d = 0.69$ for inhibition skills, which is similar to the dimension of the effect reported for children of same grade level when the intervention involves virtual coding activities (Arfé et al., 2019; Arfé et al., 2020). Insignificant effects were instead found for children's cognitive flexibility (Di Lieto, Pecini, et al., 2020).

In a second study, Di Lieto et al. (Di Lieto, Castro et al., 2020) tested the efficacy of similar structured ER activities on the visuo-spatial WM and inhibition skills of 42 1st graders with special needs, including language disabilities, attention disabilities and cognitive or motor impairments. Like in the previous study, children were assigned to an experimental or wait list group condition. The 20-h ER program was adapted to meet the motor, cognitive, and social needs of the children. The assessment instruments were the same as in the previous study. Again, the results showed an improvement in response inhibition for the children who received the training. The training effects were instead insignificant for WM.

3.2.1.5. Synthesis of the research findings. In summary, the systematic review revealed broad and positive effects of CT interventions on children's problem solving, complex EFs, such as planning, as well as on core EFs, such as cognitive inhibition and working memory skills. Remarkably, these significant effects are found for long interventions, lasting up to 32 h (Çakır et al., 2021), as well as for short intervention programs of only 4 h (Brown et al., 2008). Moreover, these effects extend also to children with special needs (Demir, 2021; Di Lieto, Castro, et al., 2020) or disadvantaged backgrounds (Brown et al., 2008).

Effect sizes are typically larger for problem solving, both when problem solving skills are assessed by tasks similar to those proposed in the intervention programs (*near transfer*; effect sizes range from $d = 1.31$ to $d = 1.91$), and when they are assessed by different tasks, like PISA assessment tools (Akcaoglu & Koehler, 2014, $d = 1.05$) or self-report scales (Erol & Çırak, 2022, $\eta^2 = 0.26$). Notably, the studies are consistent in showing significant effects of CT interventions even in comparison to active control groups, assigned to other STEM or programming activities (e.g., Arfé et al., 2019; Erol & Çırak, 2022).

Exceptions to these findings are represented by intervention studies focused on single components of CT, such as algorithm development (Oluk & Saltan, 2015), studies in which intervention effects were assessed by self-reports or problem-solving questionnaires instead of performance on cognitive tests or ad hoc problem solving tasks (see Oluk & Saltan, 2015; Çiftci & Bildiren, 2020; Çınar & Tüzün, 2021), or studies assessing cognitive flexibility (Özcan et al., 2021).

CT programs appear very effective even when EF skills are concerned. The dimension of the instructional effect is large for complex EFs such as planning (e.g., Arfé et al., 2020, d = 1.27) and moderate to large for core EF skills. Again, the only exception is represented by shifting or cognitive flexibility. However, evidence is still limited in this latter case, as only two of the studies examined in this systematic review assessed cognitive flexibility.

3.2.2. Does the cognitive effectiveness of CT vary with children's age?

As shown in Table 2, we divided the studies in two age-categories based on the age of their participants. The first category comprised older children, from grade 4 to grade 10, the second included preschoolers, first and second graders. The distribution of the studies across the two age-categories was balanced. Eleven of the studies considered in this systematic review tested CT programs addressed to older students, in one case (Demir, 2021) with mild intellectual disabilities. The remaining eight studies involved children from preschool to grade two. Based on our systematic analysis of the literature, CT interventions resulted equally effective for older and younger children.

3.2.2.1. Older age-group. The CT programs appear to be generally effective for participants in the older age-group. Among the eleven studies including older participants, only three did not report significant positive effects of the CT programs. One was the study by Oluk and Saltan (2015), in which, as anticipated, the insignificant effects could be attributed to the narrow nature of the CT intervention that focused on a single component of CT, that is algorithm development. The other two studies reporting non-significant effects, by Çınar and Tüzün (2021) and Özcan et al. (2021), proposed structured ER and visual programming interventions to 10th and 4th graders respectively.

3.2.2.2. Younger age-group. Also for the younger participants, the studies reported significant positive effects of CT programs. The only exception is the study by Çiftci and Bildiren (2020), in which *far transfer* effects of a structured virtual coding intervention were assessed by a Problem-Solving Skill Scale (Aydoğan et al., 2012). The assessment tool used in this study could however explain the finding. The scale consisted of picture items representing various types of real-life problems. Children (4- and 5-year-old) had to understand the problem represented and find the best answer to it. The way problems were formulated and understood by the participant could have affected the results.

In synthesis, CT intervention programs yielded to significant and positive far transfer effects both in the older and younger groups. In the younger children significant transfer effects, from moderate to large were observed both for complex EF skills and core EFs, such as WM and cognitive inhibition skills (Arfé et al., 2019; Arfé et al., 2020; Di Lieto, Pecini, et al., 2020). For the older children, only complex EFs like problem solving and fluid intelligence were assessed, but again the effects were significant and effect sizes, when available, were large.

3.2.3. *Which instructional modality (educational robotics/unplugged coding/virtual coding) is most effective?*

Differences in the effectiveness of CT programs appear to be related to the nature (structured or unstructured, and comprehensive or not) of the intervention more than to its modality (virtual coding, ER, or unplugged coding). Both virtual coding (visual programming) and ER intervention programs resulted effective when they addressed the various components of CT, such as problem analysis, planning, evaluating and debugging. Conversely, CT interventions focused on just one component or on programming skills solely resulted less effective (Oluk & Saltan, 2015). CT interventions were also effective both when they were specifically tailored to boost specific EF skills, like in Di Lieto, Pecini, et al. (2020) and when they were more broadly targeting problem solving abilities (Arfé et al., 2019; Arfé et al., 2020).

Game design or project development programs, in which students are allowed to develop their own projects, have been found effective in enhancing the problem-solving skills of older participants (Akcaoglu & Koehler, 2014; Erol & Çırak, 2022). These intervention programs were not addressed to younger children, who typically received more structured interventions, based on tasks with specific and predefined objectives. In these structured programs both visual programming and ER tools have been used, resulting equally effective. For instance, Arfé (Arfé et al., 2019; Arféet al., 2020 and Di Lieto, Pecini, et al. (2020), found very similar effects of virtual coding and ER interventions on 5-6 year-old children's response inhibition skills: d = 0.65 and d = 0.71 (Arfè et al., 2019; Arfé et al., 2020) and d = 0.69 (Di Lieto, Pecini, et al., 2020).

It must be noted, however, that unstructured game-design or project-development programs and structured visual programming interventions, although proposed to children of different ages, yielded similar effects on students' problem solving or planning outcomes (Akcaoglu & Koehler, 2014; Arfé et al., 2019). Akcaoglu and Koehler (2014) and Arfé et al., for instance, reported similar large effect sizes of project-development and virtual coding interventions on the problem-solving skills of students from 5th to 8th grade and 1st and 2nd graders respectively.

In synthesis, when comprehensive structured CT interventions, that involve scaffolding and practicing different CT components, are proposed to younger children their transfer effects can be similar to that of problem-solving unstructured programs addressed to older participants. This does not mean, though, that the effects of CT interventions at a given age can be independent from their structured/unstructured nature. We do not have, indeed, evidence in support of this hypothesis. Conversely, the comparison between virtual coding and ER interventions provides evidence of the equivalence of these two CT tools in boosting younger children's problem solving and EF skills.

3.3. *Meta-analysis results: effects of CT/coding interventions on problem solving, planning and core EFs*

The meta-analysis allowed to derive quantitative estimates of the cognitive effects of CT/coding interventions across EFs. The *omnibus* test suggests that at least one outcome differs from zero ($\chi_5 = 195.693$, $p < 0.001$). Table 3 summarizes the results of the multivariate fixed-effect model. Fig. 2 depicts the multivariate forest plot.

The results of the meta-analytic study (Fig. 2) show that except for cognitive flexibility (accuracy), all cognitive outcomes improved significantly after children performed coding or programming activities. Problem solving is associated with the highest effect ($d_{ppc2} = 0.89$), planning with a moderate effect ($d_{ppc2} = 0.36$), and response inhibition and WM, despite being statically significant, associated with lower effect sizes ($d_{ppc2} = 0.17$, $d_{ppc2} = 0.20$).

As explained in the previous sections, due to lack of information from published papers and the impossibility to have always access

Table 3
Multivariate fixed-effect model summary.

Outcome	d_{ppc2}	SE	95% CI	z	*p*
Cognitive Flexibility Acc.	0.118	0.096	[-0.07, 0.306]	1.227	0.22
Inhibition Acc.	0.168	0.057	[0.057, 0.279]	2.956	0.003
Planning Acc.	0.364	0.072	[0.222, 0.505]	5.047	<0.001
Problem Solving	0.890	0.064	[0.764, 1.016]	13.816	<0.001
Working Memory Acc.	0.199	0.079	[0.045, 0.353]	2.530	0.011

Omnibus Test $\chi_5 = 195.7$ $p < 0.001$; $\rho_{pre\text{-}post} = 0.7$, $\rho_{agg} = 0.5$, $\rho_{multi} = 0.5$.
Note. Each outcome is summarized with the estimated mean effect (d_{ppc2}), the standard error, the 95% confidence interval, the *z* value, and the p-value.

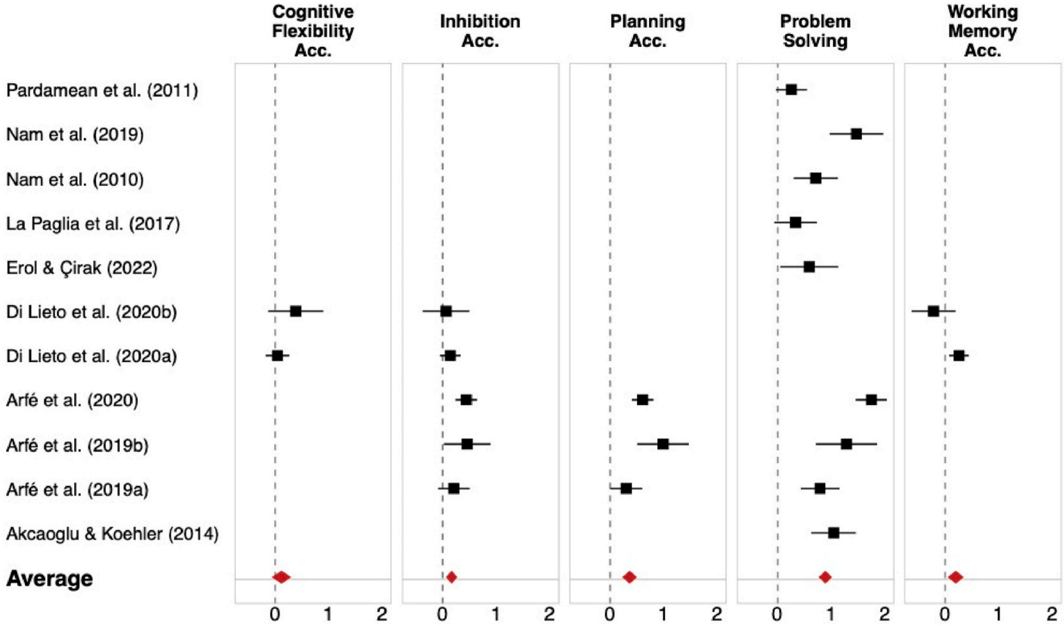


Fig. 2. Multivariate Forest Plot *Note.* Each individual effect is represented with the 95% confidence interval. The dotted line represents the null effect. For each outcome, the red diamonds depict the estimated average effect and the 95% confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

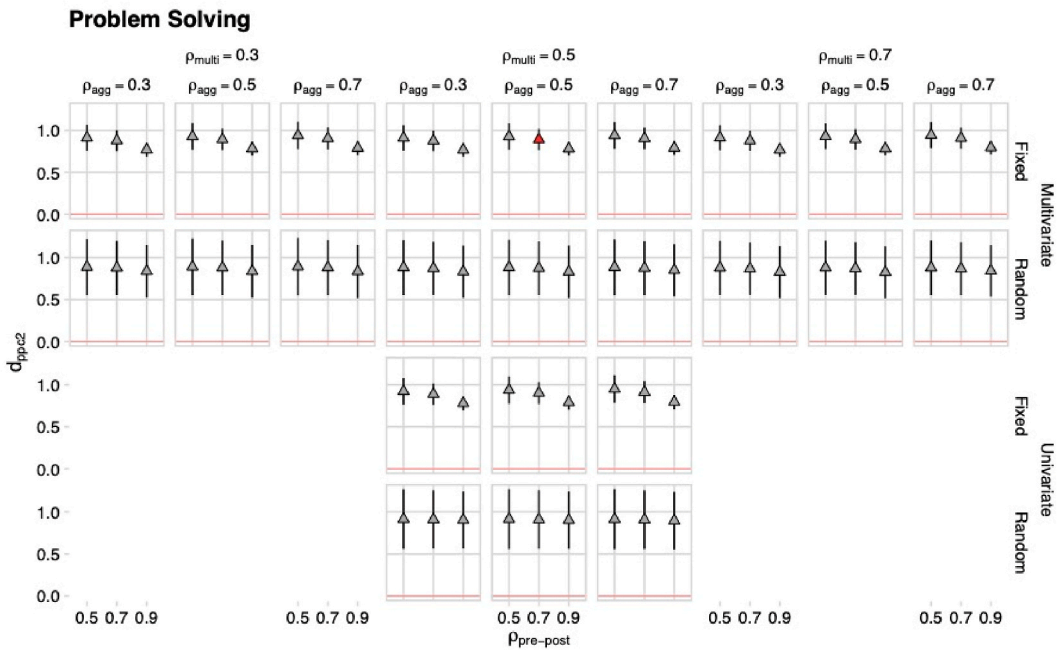


Fig. 3. Multiverse Analysis for Problem Solving *Note.* The x-axis represents the pre–post ($\rho_{pre\text{-}post}$) correlation. On the top, there is the correlation between multiple effect sizes for the same outcome (ρ_{agg}) and the correlation between different outcomes (ρ_{multi}). The triangle and the bars represent the estimated effect and 95% CI. The red shape is the chosen combination for the meta-analysis. On the right are the different MA models. For the univariate model, we computed a *fixed* or *random effect* model ignoring the statistical dependence between outcomes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

to raw data, we had to impute three correlations measures in order to compute the meta-analytic model. We also assessed the effect of imputing these correlations using a *multiverse* approach (Steegen et al., 2016). The main idea is to present a collection of different analytical possibilities. In the supplemental material, we present the effect of choosing a specific meta-analytic model (*fixed* vs. *random* *effect* and *univariate* vs. *multivariate*) varying also the imputed correlations. As an example, Fig. 3 presents the multiverse analysis for problem-solving. The example shows how choosing a certain model and correlations value does not affect conclusions about problem-solving. This strengthens the main meta-analytic result.

Moderator analyses to statistically test the influence of age and intervention type were not possible, due to the reduced number of studies for each outcome.

3.3.1. Publication bias

Typically, the required number of studies for a publication bias assessment using funnel plot-based analyses is large (Furuya-Kanamori et al., 2020). Sterne et al. (2011) suggest however a minimum of 10 studies, which may increase though in presence of heterogeneity. Using selection models (see Hedges, 1984) the number of studies should be event higher (Jin et al., 2015). For these reasons we report the publication bias assessment using the Egger regression test (Egger et al., 1997) only for the *problem solving* outcome that is associated with the highest number of effects (N = 9). We assessed the publication bias for the *univariate* fixed-effects model on *problem solving* with the chosen correlations' combination for the main meta-analysis model. The Egger regression is a meta-regression model estimating the relationship between sampling standard error and effect sizes. In the presence of publication bias (i.e., asymmetry in the funnel plot) the slope should be significantly different from zero. Furthermore, the intercept is usually interpreted as the estimated effect size with standard error close to zero. We computed the Egger test using the *regtest* function from the *metafor* package of R. The relationship between standard errors and effect sizes is not significantly different from zero ($z = 0.245,\ p = 0.807$) and the estimated *problem solving* effect with zero standard error is 0.833 (95% $CI = [0.268, 1.397]$). Overall, the Egger test suggests the absence of evidence for publication bias for the *problem solving* outcome.

4. Discussion

The systematic review and meta-analysis aimed to address three key questions regarding the teaching of computational thinking: 1) which specific EFs are most influenced by it? 2) does the impact of CT programs differ based on children's age? and 3) which instructional modality (educational robotics, unplugged coding, or virtual coding) is most effective in enhancing EF skills in children and adolescents? The first research question was answered both by the systematic review and meta-analysis. The second and third research questions were exclusively tackled through the systematic review of the literature.

Unlike previous meta-analytic studies (Liao, 2000; Scherer et al., 2019), we exclusively focused on experimental trials. This approach allowed us to assess the cognitive effectiveness of CT interventions while accounting for potential confounding factors such as repeated testing and practice effects. While this decision improved the quality of our overall conclusions regarding the cognitive effects of CT, it also restricted the number of studies available for conducting a statistical analysis of the moderating effects of age and type of intervention. As a result, the conclusions on age and type of intervention effects are based on a qualitative analysis of the literature.

4.1. Which EFs are most influenced by the teaching of CT?

The review of the experimental studies showed that CT interventions are generally very effective in boosting children and adolescents' higher order and core EF skills. The largest effects were observed on children's problem solving and complex EFs such as planning, but significant positive effects emerge also for core EFs, like cognitive inhibition and working memory.

The meta-analysis provided statistical evidence to the qualitative conclusions of the systematic review, comparing the effectiveness of CT interventions across cognitive skills. Its results confirmed that problem-solving and planning were the cognitive skills that CT programs most strongly affected, indicating that children's inhibition and WM skills were influenced too by these interventions, albeit to a lesser extent.

Most studies reported *far transfer* effects to problem-solving tasks that are different from those trained by the intervention. Our meta-analysis provides thus further support for the transferability of the skills acquired through CT programs to situations that require problem-solving abilities, finding however a larger transfer effect ($d_{ppc2} = 0.89$) than in previous meta-analyses (Scherer et al., 2019), both in comparison with the *near* and *far transfer* effects (respectively, $\bar{g} = .75$ and $\bar{g} = .47$). Planning is the other EF that reflects most the cognitive effects of CT interventions ($d_{ppc2} = 0.36$), and in this case, the effects are similar to those reported for other higher order EFs (i.e., reasoning) in Scherer et al.'s study ($g = 0.37$).

The impact of CT interventions on children's problem solving skills is large both when children's problem solving skills are assessed by tasks similar to those proposed in the intervention programs, such as coding problems (*near transfer,* e.g., Arfé et al., 2019), and by very different tasks, like the problem solving tasks of PISA assessment or self-report scales (*far transfer,* e.g., Akcaoglu & Koehler, 2014; Erol and Çırak, 2022). In addition, these effects emerge also in studies with active control groups, that is, when the effects of CT programs are compared to those of STEM or programming activities (e.g., Arfé et al., 2019; Erol & Çırak, 2022).

CT programs show also moderate *far transfer* effects on core EF skills such as cognitive inhibition and working memory, with effect sizes that are consistent across intervention studies employing different CT methods and tools (Arfé et al., 2020; Di Lieto, Pecini, et al., 2020). These transfer effect appear however weaker ($d_{ppc2} = 0.17$ for cognitive inhibition, $d_{ppc2} = 0.20$ for working memory) compared to those found for higher order EFs, such as planning and problem solving.

Planning and problem-solving are higher order EFs that closely relate to the kind of problem-solving activities involved in CT programs. Thus, they may better reflect improvements related to CT. As reported in other reviews and meta-analyses on the effectiveness of EF training programs (Diamond & Ling, 2016; Scionti et al., 2020), although the benefits of EFs training can transfer to untrained skills, the transfer effects appear to be typically narrow, and more pronounced for abilities similar to those trained.

On the other hand, foundational EFs such as inhibition and working memory may show less stability over time and be more influenced by temporary psychological and physiological states, like fatigue or stress, because they depend more on executive control processes than planning and problem solving. For these latter EF skills the acquisition of effective metacognitive strategies can be more critical (La Paglia et al., 2017). Other meta-analyses have shown that even when cognitive trainings are specifically focused on core EF skills, like WM and response inhibition, intervention effects can be small or non-significant, especially when far transfer effects are considered (Kassai et al., 2019; Melby-Lervåg & Hulme, 2013). Inhibitory control in particular seems resistant to EF interventions, showing non-significant effects to EF computer-based trainings (Cao et al., 2020), and small effect sizes ($g^{+} = 0.18$) even when the training focuses on inhibition skills (*near transfer,* Kassai et al., 2019).

4.2. *Does the cognitive effectiveness of CT vary with children's age?*

CT interventions seem equally effective for older and younger children. Beneficial effects of CT programs on problem solving are consistently reported across studies involving older (4th to 10th graders) or younger (preschoolers-2nd graders) participants who received comprehensive CT interventions targeting various components of CT, regardless of the specific type of intervention (virtual coding or ER; e.g., Erol & Çırak, 2022; La Paglia et al., 2017).

Older students participating in CT programs improved in performing problem-solving tasks (Akcaoglu & Koehler, 2014; Lai & Yang, 2011), in self-reported problem-solving (Erol & Çırak, 2022), and also developed metacognitive skills involved in problem solving (La Paglia et al., 2017). For the younger children, significant transfer effects, from moderate to large, were observed both for complex EF skills (Akcaoglu & Koehler, 2014; Nam et al., 2019) and core EFs, such as WM and cognitive inhibition skills (Arfé et al., 2019; Arfé et al., 2020; Di Lieto, Pecini, et al., 2020), although for the latter effect sizes were small.

Given that all the studies assessing inhibition and working memory in this systematic review and meta-analysis focused exclusively on younger children, aged 5–7, it cannot be excluded that age could account for the limited *far transfer* effects on these EFs. In our meta-analysis the moderating effects of age could not be tested owing to the small sample size. However, other recent meta-analytic studies that considered the effects of EF training programs on young children's (3 to 6-year-old, Scionti et al., 2020) or transfer effects of programming across a larger age-span (from prekindergarten to college students, Scherer et al., 2019) suggest that age is not a significant moderator of the effects observed. Despite these meta-analytic findings, age related effects remain a pedagogically relevant research question that deserves more attention, especially if we consider the different plasticity of core EFs across child development (McGuckian et al., 2023; Scionti et al., 2020).

4.3. *Which instructional modality (educational robotics/virtual coding) is most effective in enhancing children and adolescents' EF skills?*

Based on our systematic review the different effects of CT activities appear not to be strictly related to the method or programming tool used (virtual coding, unplugged coding or ER), but rather to the structured or unstructured nature of the intervention and the age of the participants.

4.3.1. *Structured versus unstructured interventions*

A qualitative comparison between the effectiveness of structured and unstructured CT intervention programs was possible only for the older age group (grades 4 to 10). Game-design or project-development intervention programs, that consist in open-ended problems, seem much effective in enhancing the problem-solving skills and self-reported problem-solving abilities of these older learners (Erol & Çırak, 2022; Nam et al., 2010). Structured interventions, whether alone or combined with unstructured project development, appear to be less effective (Çınar & Tüzün, 2021; Özcan et al., 2021). Exceptions are studies in which the intervention addressed students' mathematical problem solving (La Paglia et al., 2017) or was for students with intellectual disabilities (Demir, 2021).

Younger children received only structured interventions. For them, structured interventions, based on logic games with a single correct solution, proved effective both in boosting problem solving and EF skills (Arfé et al., 2020; Di Lieto et., 2020a). Among the structured programs that involved younger children, visual programming and ER tools appear to be equally effective. The specific effects of virtual coding and tangible (ER and unplugged) coding interventions are discussed below.

4.3.2. *Virtual coding*

The results of the systematic review show that virtual coding interventions, when appropriately designed, result effective for students in a large age span, from grade 1 to 8. Twelve of the experimental studies examined in this systematic review tested the cognitive effects of virtual coding interventions, and nine of them (75%) reported significant positive effects on problem-solving and other EFs.

CT involves a complex set of cognitive abilities, such as memory, self-regulation, and planning, which are also involved in problem-solving (Frensch & Funke, 1995; Keen, 2011). These skills, as well as the development of higher order EF, including problem-solving skills, develop significantly from elementary to middle school (Brocki & Bohlin, 2004; De Luca et al., 2003; Luciana & Nelson, 2002; Luna et al., 2004), and further on during adolescence (Hooper et al., 2004; Unterrainer & Owen, 2006; Zelazo et al., 1997). For younger students, who are still developing foundational EFs (Gathercole et al., 2004; Jonkman et al., 2003; Klenberg et al., 2001) structured virtual coding interventions enhance both problem-solving skills similar to those trained during the intervention (*near transfer*), and EFs less closely related to coding, such as planning and inhibition skills (*far transfer*). For late elementary school students and adolescents who have already developed the foundational EF skills (e.g., inhibitory control, working memory) and metacognitive abilities that are necessary to manage complex coding tasks (Best & Miller, 2010; Conklin et al., 2007; Gathercole et al., 2004; Jonkman et al., 2003; Klenberg et al., 2001), practicing with ill-defined, or unstructured, virtual coding or programming problems, such as those involved in game-design or creative problem solving activities, results most effective. The finding that unstructured virtual coding activities seem to work better with older students should not be surprising. As Diamond and Ling (2016) have emphasized EFs need to be challenged, not just used, to promote improvement.

4.3.3. Educational robotics

Based on a concrete sensory-motor experience, ER or unplugged coding activities, are more often addressed to younger children, that is, preschoolers or first graders, or students with cognitive disabilities (Demir, 2021; Di Lieto et al., 2020b; Nam et al., 2019). Out of the 19 studies included in this systematic review, only two experimental studies (Çınar & Tüzün, 2021; La Paglia et al., 2017) specifically examined the effects of ER interventions on late elementary school or middle school students, producing contrasting findings. La Paglia et al. (2017), found positive *far transfer* effects of ER on fifth graders' problem-solving skills. Çınar and Tüzün (2021) found no significant *far transfer* effects of ER to tenth graders' self-reported problem-solving.

In the younger, 5–6 year-old children, ER interventions seem to have general positive *far transfer* effects on EF skills, improving significantly children's problem-solving skills (Çakır et al., 2021; Nam et al., 2019), as well as core EFs (response inhibition and working memory, Di Lieto, Pecini, et al., 2020; Di Lieto, Castro, et al., 2020). The benefits of ER seem however reduced for children with special educational needs (children with sensory, motor, or cognitive disabilities, or attention deficit hyperactivity disorder, and/or specific learning disorders, Di Lieto, Castro, et al., 2020). For these children, tangible, unplugged coding activities, seem more effective (Demir, 2021). A possible explanation is that although ER can sustain EF through concrete and tangible activities, it requires basic computer skills and memory and cognitive resources that students with intellectual disabilities may lack (Demir, 2021; Di Lieto et al., 2020b).

Since all studies focused on ER interventions also involved structured problem-solving activities, it is also difficult to determine whether their beneficial effects for younger children were due to the tangible (ER) or structured nature of the intervention. In fact, when narrowing our analysis to younger, 5-6 year-old children, structured virtual coding and ER activities seem to have equivalent *far transfer* effects on core EFs (Arfé et al., 2019, 2020; Di Lieto, Pecini, et al., 2020) and higher order EF skills, like problem solving (Arfé et al., 2019; Nam et al., 2019). These findings seem in contrast with those of Scherer et al. (2020), who report larger effects of instructional programs based on physicality, such as robotics ($\bar{g} = 0.72$) than of virtual coding ($\bar{g} = 0.44$), suggesting that the medium used in the instructional intervention can play a role. Direct comparisons between structured virtual coding and ER interventions are needed to test this hypothesis and could be a goal of future studies.

5. Conclusions

Over the last 20 years, computer scientists, experts in education, and psychologists have explored the cognitive effectiveness of CT-based activities or instructional programs, primarily in the domain of children's EFs (e.g., Arfé et al., 2020; Brown et al., 2008; Di Lieto, Pecini, et al., 2020; Lai & Yang, 2011). However, there have been very few systematic reviews and meta-analyses that synthetized this literature (Liao, 2000; Liao & Bright, 1991; Scherer et al., 2019, 2020). Systematic reviews and meta-analyses constitute complementary and fundamental tools of evidence-based practice, particularly when the focus of interest is on the evaluation of intervention effects, like in this study.

Overall, the systematic review and meta-analysis presented in this paper confirm that CT programs, that most countries are recently integrating in their school curriculum (Lye & Koh, 2014), can be a powerful tool to boost and support the development of EFs, particularly higher order EFs like planning and problem solving, but also core EFs, like response inhibition and WM that underpin and predict early and late academic achievements (Clark et al., 2010; Jacob & Parkinson, 2015; Spiegel et al., 2021).

Despite variation in outcomes across EFs, 14 of the 19 studies considered in this systematic review (74%) reported evidence of EF benefits. This is a higher percentage than that reported by Diamond and Ling (2016) for cognitive trainings (from 20% to 60%) and aerobic exercises (43%). While, it is remarkably similar to the effectiveness of other school programs, such as Montessori and Tools of the Mind (75%). As for these programs, an early integration of CT/coding in the school curriculum may be strategic for children's future academic accomplishments.

School embedded programs are comparatively more effective than cognitive training interventions targeting EFs (Diamond & Ling, 2016). When the training is part of children's daily school curriculum its activities become meaningful to children, and children may better perceive the importance and utility of the abilities learnt, their social value, and their transferability to other learning situations.

The finding of a beneficial effect of CT, especially on higher order EFs, is also important for another reason. Although a number of intervention studies have focused on improving EFs (e.g., Passolunghi & Costa, 2016; Pozuelos et al., 2019; Schmitt et al., 2015), few have demonstrated effects on higher order EFs (Diamond & Ling, 2016; Scionti et al., 2020). Enhancing these higher-order cognitive skills, also bears on a child's adaptive capacities to the social environment (Barkley, 2001; Huepe et al., 2011).

Our systematic review found that the effects of CT/coding did not seem to vary by the training method or tool (virtual coding/unplugged/ER). Comparisons between training methods is however limited by the fact that different types of interventions have been typically addressed to different age-groups. The lack of direct comparisons between virtual coding and ER or structured and unstructured intervention conditions stands out as a prominent limitation in the existing literature. Another notable gap lies in the scarcity of studies examining the efficacy of CT activities in enhancing core executive functions (working memory, response inhibition, cognitive flexibility) in older students. CT programs may be particularly suitable to support the development of core EF skills even during late elementary schools and adolescence.

5.1. Limitations and future directions

In addition to the limitations inherent in the existing literature, it is important to acknowledge some limitations of the present meta-analysis. The first concerns the lack of complete data from published studies. For computing the appropriate meta-analytic model, we needed several correlation values often omitted in the papers selected. We addressed this problem by employing the multiverse-like sensitivity analysis that clearly shows when a certain imputed correlation has a relevant impact on the meta-analysis model.

A second limitation of this systematic review and meta-analysis is that, by applying rigorous inclusion criteria, we could include a limited number of studies. For instance, as one e of the inclusion criteria for this systematic review was the peer-reviewed status of publications, we did not consider grey literature. Consequently, we were not able to robustly assess publication bias or consider important moderators, such as age and type of intervention, in the meta-analysis. With the growing number of experimental studies in the field of CT, future meta-analyses will be able to compare the effectiveness of different types of CT intervention, or how it varies with students' age. Finally, other variables could influence the effectiveness of a coding training and should be considered in future studies; examples are motivational factors like self-efficacy (Tsai, 2019), or gender (Montuori et al., 2022). These variables may have an important role in moderating children's response to coding interventions. The spread of experimental research in the area of CT studies, and the interdisciplinary collaboration between computer scientists and psychologists will soon allow to address also these factors.

Credit author statement

Chiara Montuori: conducted the systematic review and meta-analysis; produced the original draft of the paper in collaboration with Barbara Arfé; Filippo Gamabarota and Gianmarco Altoé: conducted the meta-analysis and contributed all statistical details in the paper; read and commented the original and revised version; Barbara Arfé: supervised the systematic review and meta-analysis. Produced the first draft in collaboration with Chiara Montuori; Revised the manuscript.

Declaration of competing interest

We have no known conflict of interest to disclose.

Data availability

Data are available at the supplementary materials link.

Acknowledgments

The authors wish to thank Elisa Scandella for helping with the initial literature screening.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2023.104961.

References

Akcaoglu, M., & Koehler, M. J. (2014). Cognitive outcomes from the game-design and learning (GDL) after-school program. *Computers & Education, 75*, 72–81. https://doi.org/10.1016/j.compedu.2014.02.003

Altemeier, L. E., Abbott, R. D., & Berninger, V. W. (2008). Executive functions for reading and writing in typical literacy development and dyslexia. *Journal of Clinical and Experimental Neuropsychology, 30*(5), 588–606. https://doi.org/10.1080/13803390701562818

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724

Arfé, B., Vardanega, T., Montuori, C., & Lavanga, M. (2019). Coding in primary grades boosts children's executive functions. *Frontiers in Psychology, 10*. https://www.frontiersin.org/article/10.3389/fpsyg.2019.02713.

Arfé, B., Vardanega, T., & Ronconi, L. (2020). The effects of coding on children's planning and inhibition skills. *Computers & Education, 148*, Article 103807. https://doi.org/10.1016/j.compedu.2020.103807

Aydoğan, Y., Ömeroğlu, E., Büyüköztürk, Ş., & Özyürek, A. (2012). *Problem çözme becerileri*.

Barkley, R. A. (2001). The executive functions and self-regulation: An evolutionary neuropsychological perspective. *Neuropsychology Review, 11*(1), 1–29.

Barrouillet, P. (2015). Theories of cognitive development: From Piaget to today. *Developmental Review, 38*, 1–12. https://doi.org/10.1016/j.dr.2015.07.004

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research, 50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education, 72*, 145–157. https://doi.org/10.1016/j.compedu.2013.10.020

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

Bisiacchi, P., Cendron, M., Gugliotta, M., Tressoldi, P. E., & Vio, C. (2005). *BVN 5-11—batteria di valutazione neuropsicologica per l'età evolutiva*. Erickson.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. https://doi.org/10.1002/9780470743386

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology, 26*(2), 571–593. https://doi.org/10.1207/s15326942dn2602_3

Brown, Q., Mongan, W., Kusic, D., Garbarine, E., Fromm, E., & Fontecchio, A. (2008). *Computer aided instruction as a vehicle for problem solving: Scratch programming environment in the middle years classroom*. In ASEE Annual Conference and Exposition, Conference Proceedings. https://doi.org/10.18260/1-2—3826, 10.18260, 13.319.1 - 13.319.16.

Cai, Z., & Fan, X. (2020). A comparison of fixed-effects and random-effects models for multivariate meta-analysis using an SEM approach. *Multivariate Behavioral Research, 55*(6), 839–854. https://doi.org/10.1080/00273171.2019.1689348

Çakir, R., Korkmaz, Ö., İdil, Ö., & Uğur Erdoğmuş, F. (2021). The effect of robotic coding education on preschoolers' problem solving and creative thinking skills. *Thinking Skills and Creativity, 40*, Article 100812. https://doi.org/10.1016/j.tsc.2021.100812

Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004). CONSORT statement: Extension to cluster randomised trials. *BMJ, 328*(7441), 702–708. https://doi.org/10.1136/bmj.328.7441.702

Cao, Y., Huang, T., Huang, J., Xie, X., & Wang, Y. (2020). Effects and moderators of computer-based training on children's executive functions: A systematic review and meta-analysis. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.580329

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., & Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & Education, 109*, 162–175. https://doi.org/10.1016/j.compedu.2017.03.001

Cheung, M. W.-L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(3), 429 454. https://doi.org/10.1080/10705511.2013.797827

Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review, 29*(4), 387–396. https://doi.org/10.1007/s11065-019-09415-6

Çiftci, S., & Bildiren, A. (2020). The effect of coding courses on the cognitive abilities and problem-solving skills of preschool children. *Computer Science Education, 30*(1), 3–21. https://doi.org/10.1080/08993408.2019.1696169

Çınar, M., & Tüzün, H. (2021). Comparison of object-oriented and robot programming activities: The effects of programming modality on student achievement, abstraction, problem solving, and motivation. *Journal of Computer Assisted Learning, 37*(2), 370–386. https://doi.org/10.1111/jcal.12495

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*(5), 1176–1191. https://doi.org/10.1037/a0019672

Clements, D. H., & Nastasi, B. K. (1999). Metacognition, learning, and educational computer environments. *Information Technology in Childhood Education Annual, 1999*(1), 3–36.

Colzato, L. S., van Wouwe, N. C., Lavender, T. J., & Hommel, B. (2006). Intelligence and cognitive flexibility: Fluid intelligence correlates with feature "unbinding" across perception and action. *Psychonomic Bulletin & Review, 13*(6), 1043–1048. https://doi.org/10.3758/BF03213923

Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental Neuropsychology, 31*(1), 103–128. https://doi.org/10.1080/87565640709336889

Crocetti, E. (2016). Systematic reviews with meta-analysis: Why, when, and how? *Emerging Adulthood, 4*(1), 3–18. https://doi.org/10.1177/2167696815617076

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J.-A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the cantab. I: Development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology, 25*(2), 242–254. https://doi.org/10.1076/jcen.25.2.242.13639

Demir, Ü. (2021). The effect of unplugged coding education for special education students on problem-solving skills. *International Journal of Computer Sciences and Engineering Systems, 4*(3). https://doi.org/10.21585/ijcses.v4i3.95. Article 3.

Di Lieto, M. C., Castro, E., Pecini, C., Inguaggiato, E., Cecchi, F., Dario, P., Cioni, G., & Sgandurra, G. (2020a). Improving executive functions at school in children with special needs by educational robotics. *Frontiers in Psychology, 10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02813.

Di Lieto, M. C., Pecini, C., Castro, E., Inguaggiato, E., Cecchi, F., Dario, P., Cioni, G., & Sgandurra, G. (2020b). Empowering executive functions in 5- and 6-year-old typically developing children through educational robotics: An RCT study. *Frontiers in Psychology, 10*. https://www.frontiersin.org/article/10.3389/fpsyg.2019.03084.

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience, 18*, 34–48. https://doi.org/10.1016/j.dcn.2015.11.005

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Erol, O., & Çırak, N. S. (2022). The effect of a programming tool scratch on the problem-solving skills of middle school students. *Education and Information Technologies, 27*(3), 4065–4086. https://doi.org/10.1007/s10639-021-10776-w

Fessakis, G., Gouli, E., & Mavroudi, E. (2013). Problem solving by 5–6 years old kindergarten children in a computer programming environment: A case study. *Computers & Education, 63*, 87–97. https://doi.org/10.1016/j.compedu.2012.11.016

Feurzeig, W., & Papert, S. A. (2011). Programming-languages as a conceptual framework for teaching mathematics. *Interactive Learning Environments, 19*(5), 487–501. https://doi.org/10.1080/10494820903520040

Finke, S., Kemény, F., Sommer, M., Krnjic, V., Arendasy, M., Slany, W., & Landerl, K. (2022). Unravelling the numerical and spatial underpinnings of computational thinking: A pre-registered replication study. *Computer Science Education, 32*(3), 313–334. https://doi.org/10.1080/08993408.2022.2079864

Flórez, B. F., Casallas, R., Hernández, M., Reyes, A., Restrepo, S., & Danies, G. (2017). Changing a generation's way of thinking: Teaching computational thinking through programming. *Review of Educational Research, 87*(4), 834–860. https://doi.org/10.3102/0034654317710096

Frensch, P. A., & Funke, J. (1995). *Complex problem solving. The European perspective*. Psychology Press.

Furuya-Kanamori, L., Xu, C., Lin, L., Doan, T., Chu, H., Thalib, L., & Doi, S. A. R. (2020). P value–driven methods were underpowered to detect publication bias: Analysis of Cochrane review meta-analyses. *Journal of Clinical Epidemiology, 118*, 86–92. https://doi.org/10.1016/j.jclinepi.2019.11.011

Ganesan, K., & Steinbeis, N. (2022). Development and plasticity of executive functions: A value-based account. *Current Opinion in Psychology, 44*, 215–219. https://doi.org/10.1016/j.copsyc.2021.09.012

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

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 Years of age. *Developmental Psychology, 40*, 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gerosa, A., Koleszar, V., Tejera, G., Gómez-Sena, L., & Carboni, A. (2021). Cognitive abilities and computational thinking at age 5: Evidence for associations to sequencing and symbolic number comparison. *Computers and Education Open, 2*, Article 100043. https://doi.org/10.1016/j.caeo.2021.100043

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). TEST REVIEW behavior rating inventory of executive function. *Child Neuropsychology, 6*(3), 235–238. https://doi.org/10.1076/chin.6.3.235.3152

Gugliotta, M., Bisiacchi, P., Cendron, M., Tressoldi, P. E., & Vio, C. (2009). *BVN 12-18: Batteria di valutazione neuropsicologia per l'adolescenza*. Erickson.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics, 9*(1), 61–85. https://doi.org/10.2307/1164832

Heppner, P. P. (1988). *The problem-solving inventory: Manual*. Consulting Psychologists Press.

Hooper, C. J., Luciana, M., Conklin, H. M., & Yarger, R. S. (2004). Adolescents' performance on the Iowa Gambling Task: Implications for the development of decision making and ventromedial prefrontal cortex. *Developmental Psychology, 40*(6), 1148–1158. https://doi.org/10.1037/0012-1649.40.6.1148

Huepe, D., Roca, M., Salas, N., Canales-Johnson, A., Rivera-Rei, Á. A., Zamorano, L., Concepción, A., Manes, F., & Ibañez, A. (2011). Fluid intelligence and psychosocial outcome: From logical problem solving to social adaptation. *PLoS One, 6*(9), Article e24858. https://doi.org/10.1371/journal.pone.0024858

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 85*, 512–552. https://doi.org/10.3102/0034654314561338

Jin, Z.-C., Zhou, X.-H., & He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine, 34*(2), 343–360. https://doi.org/10.1002/sim.6342

Jonkman, L. M., Lansbergen, M., & Stauder, J. E. A. (2003). Developmental differences in behavioral and event-related brain responses associated with response preparation and inhibition in a go/nogo task. *Psychophysiology, 40*(5), 752–761. https://doi.org/10.1111/1469-8986.00075

Kalelioğlu, F. (2015). A new way of teaching programming skills to K-12 students: Code.org. *Computers in Human Behavior, 52*, 200–210. https://doi.org/10.1016/j.chb.2015.05.047

Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near- and far-transfer effects among children's executive function skills. *Psychological Bulletin, 145*(2), 165–188. https://doi.org/10.1037/bul0000180

Kazakoff, E. R., & Bers, M. U. (2014). Put your robot in, put your robot out: Sequencing through programming robots in early childhood. *Journal of Educational Computing Research, 50*(4), 553–573. https://doi.org/10.2190/EC.50.4.f

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual Review of Psychology, 62*(1), 1–21. https://doi.org/10.1146/annurev.psych.031809.130730

Keren, G., & Fridin, M. (2014). Kindergarten social assistive robot (KindSAR) for children's geometric thinking and metacognitive development in preschool education: A pilot study. *Computers in Human Behavior, 35*, 400–412. https://doi.org/10.1016/j.chb.2014.03.009

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20*(1), 407–428. https://doi.org/10.1207/S15326942DN2001_6

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: A developmental neuropsychological assessment*. The Psychological Corporation. https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft15125-000.

La Paglia, F., La Cascia, C., Francomano, M. M., & La Barbera, D. (2017). Educational robotics to improve mathematical and metacognitive skills. *Annual Review of CyberTherapy and Telemedicine, 15*.

Lai, A. F., & Yang, S. M. (2011). The learning effect of visualized programming learning on 6th graders' problem solving and logical reasoning abilities. In *2011 international conference on electrical and control engineering* (pp. 6940–6944). https://doi.org/10.1109/ICECENG.2011.6056908

Lee, K. T. H., Sullivan, A., & Bers, M. U. (2013). Collaboration by design: Using robotics to foster social interaction in kindergarten. *Computers in the Schools, 30*(3), 271–281. https://doi.org/10.1080/07380569.2013.805676

Liao, Y. (2000). A meta-analysis of computer programming on cognitive outcomes: An updated synthesis. In *Paper presented at the proceedings of world conference on educational multimedia, hypermedia and telecommunications, Montreal, québec, Canada*. Retrieved from https://www.learntechlib.org/primary/p/16132/.

Liao, Y. K. C., & Bright, G. W. (1991). Effects of computer programming on cognitive outcomes: A meta-analysis. *Journal of Educational Computing Research, 7*(3), 251–268. https://doi.org/10.2190/E53G-HH8K-AJRR-K69M

Liu, Y., Ma, Z., & Qian, Y. (2019). Developing Chinese elementary school students' computational thinking: A convergent cognition perspective. *Proceedings of the ACM Conference on Global Computing Education, 238*. https://doi.org/10.1145/3300115.3312514

Luciana, M., Collins, P. F., Olson, E. A., & Schissel, A. M. (2009). Tower of London performance in healthy adolescents: The development of planning skills and associations with self-reported inattention and impulsivity. *Developmental Neuropsychology, 34*(4), 461–475. https://doi.org/10.1080/87565640902964540

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the cambridge neuropsychological testing automated battery: Performance in 4- to 12-year-old children. *Developmental Neuropsychology, 22*(3), 595–624. https://doi.org/10.1207/S15326942DN2203_3

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development, 75*(5), 1357–1372. https://doi.org/10.1111/j.1467-8624.2004.00745.x

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior, 41*, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Macdonald, J. A., Beauchamp, M. H., Crigan, J. A., & Anderson, P. J. (2014). Age-related differences in inhibitory control in the early school years. *Child Neuropsychology, 20*(5), 509–526. https://doi.org/10.1080/09297049.2013.822060

Mammarella, I. C., Toso, C., Pazzaglia, F., & Cornoldi, C. (2008). *BVS - Corsi—Batteria per la valutazione della memoria visiva e spaziale*. Erickson.

Marinus, E., Powell, Z., Thornton, R., McArthur, G., & Crain, S. (2018). Unravelling the Cognition of Coding in 3-to-6-year Old: The development of an assessment tool and the relation between coding ability and cognitive compiling of syntax in natural language. *Proceedings of the 2018 ACM Conference on International Computing Education Research*, 133–141. https://doi.org/10.1145/3230977.3230984

Marshall, P. J., & Drew, A. R. (2014). What makes Simon Says so difficult for young children? *Journal of Experimental Child Psychology, 126*, 112–119. https://doi.org/10.1016/j.jecp.2014.03.011

Marzocchi, G. M., Re, A. M., & Cornoldi, C. (2010). *BIA. Batteria italiana per l'ADHD*. Erickson.

Mathes, T., & Kuss, O. (2018). A comparison of methods for meta-analysis of a small number of studies with binary outcomes. *Research Synthesis Methods, 9*(3), 366–381. https://doi.org/10.1002/jrsm.1296

McGuckian, T. B., Wilson, P. H., Johnston, R. D., Rahimi-Golkhandan, S., Piek, J., Green, D., Rogers, J. M., Maruff, P., Steenbergen, B., & Ruddock, S. (2023). Development of complex executive function over childhood: Longitudinal growth curve modeling of performance on the Groton Maze Learning Task. *Child Development, 94*(3), 648–658. https://doi.org/10.1111/cdev.13888

Meixner, J. M., Warner, G. J., Lensing, N., Schiefele, U., & Elsner, B. (2019). The relation between executive functions and reading comprehension in primary-school students: A cross-lagged-panel analysis. *Early Childhood Research Quarterly, 46*, 62–74. https://doi.org/10.1016/j.ecresq.2018.04.010

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

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research, 14*(1), 579. https://doi.org/10.1186/s12913-014-0579-0

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Montuori, C., Ronconi, L., Vardanega, T., & Arfé, B. (2022). Exploring gender differences in coding at the beginning of primary school. *Frontiers in Psychology, 13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.887280.

Moreno-León, J., Robles, G., & Román-González, M. (2016). Comparing computational thinking development assessment scores with software complexity metrics. In *2016 IEEE global engineering education conference (EDUCON)* (pp. 1040–1045). https://doi.org/10.1109/EDUCON.2016.7474681

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Nam, D., Kim, Y., & Lee, T. (2010). The effects of scaffolding-based courseware for the scratch programming learning on student problem solving skill. In *Proceedings of the 18th international conference on computers in education (ICCE)*.

Nam, K. W., Kim, H. J., & Lee, S. (2019). Connecting plans to action: The effects of a card-coded robotics curriculum and activities on Korean Kindergartners. *The Asia-Pacific Education Researcher, 28*(5), 387–397. https://doi.org/10.1007/s40299-019-00438-4

Nardelli, E. (2019). Do we really need computational thinking? *Communications of the ACM, 62*(2), 32–35. https://doi.org/10.1145/3231587

Nouwens, S., Groen, M. A., Kleemans, T., & Verhoeven, L. (2021). How executive functions contribute to reading comprehension. *British Journal of Educational Psychology, 91*(1), Article e12355. https://doi.org/10.1111/bjep.12355

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science, 25*(4), 903–910. https://doi.org/10.1177/0956797613518351

OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing.* https://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf. https://doi.org/10.1787/9789264190511-en

Oğuz, V., & Köksal Akyol, A. (2015). Problem-solving skills scale (PSSS) validity and reliability. *Cukurova University Faculty of Educational Journal, 44*(1). https://avesis.ankara.edu.tr/yayin/ea04cac4-92de-4348-8408-4797177d4dc8/problem-solving-skills-scale-psss-validity-and-reliability.

Oluk, A., & Saltan, F. (2015). Effects of using the scratch program in 6th grade information technologies courses on algorithm development and problem solving skills. *Participatory Educational Research (PER), 2*(5), 10–20. https://doi.org/10.17275/per.15.spi.2.2

Özcan, M.Ş., Çetinkaya, E., Göksun, T., & Kisbu-Sakarya, Y. (2021). Does learning to code influence cognitive skills of elementary school children? Findings from a randomized experiment. *British Journal of Educational Psychology, 91*(4), 1434–1455. https://doi.org/10.1111/bjep.12429

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Pardamean, B., Evelin, E., & Honni, H. (2011). The effect of logo programming language for creativity and problem solving. *Proceedings of the 10th WSEAS International Conference on E-Activities*, 151–156.

Passolunghi, M. C., & Costa, H. M. (2016). Working memory and early numeracy training in preschool children. *Child Neuropsychology, 22*(1), 81–98. https://doi.org/10.1080/09297049.2014.971726

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In *International encyclopedia of education* (Vol. 2, pp. 6452–6457). Pergamon Press.

Ping, R., & Goldin-Meadow, S. (2008). Hands in the air. *Developmental Psychology, 44*(5), 1277–1287. https://doi.org/10.1037/0012-1649.44.5.1277

Polat, E., Hopcan, S., Kucuk, S., & Sisman, B. (2021). A comprehensive assessment of secondary school students' computational thinking skills. *British Journal of Educational Technology, 52*(5), 1965–1980. https://doi.org/10.1111/bjet.13092. Scopus.

Pozuelos, J. P., Combita, L. M., Abundis, A., Paz-Alonso, P. M., Conejero, Á., Guerra, S., & Rueda, M. R. (2019). Metacognitive scaffolding boosts cognitive and neural benefits following executive attention training in children. *Developmental Science, 22*(2), Article e12756. https://doi.org/10.1111/desc.12756

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior, 72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

Román-González, M., Pérez-González, J.-C., Moreno-León, J., & Robles, G. (2018). Can computational talent be detected? Predictive validity of the computational thinking test. *International Journal of Child-Computer Interaction, 18*, 47–58. https://doi.org/10.1016/j.ijcci.2018.06.004

Sáez-López, J.-M., Román-González, M., & Vázquez-Cano, E. (2016). Visual programming languages integrated across the curriculum in elementary school: A two year case study using "scratch" in five schools. *Computers & Education, 97*, 129–141. https://doi.org/10.1016/j.compedu.2016.03.003

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

Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2019). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. *Journal of Educational Psychology, 111*, 764–792. https://doi.org/10.1037/edu0000314

Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2020). A meta-analysis of teaching and learning computer programming: Effective instructional approaches and conditions. *Computers in Human Behavior, 109*, Article 106349. https://doi.org/10.1016/j.chb.2020.106349

Schmitt, S. A., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2015). Strengthening school readiness for Head Start children: Evaluation of a self-regulation intervention. *Early Childhood Research Quarterly, 30*, 20–31. https://doi.org/10.1016/j.ecresq.2014.08.001

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

Serin, O., Bulut Serin, N., & Saygılı, G. (2010). İlköğretim düzeyindeki çocuklar için problem çözmeenvanteri'nin (ÇPÇE) geliştirilmesi [developing problem solving inventory for children at the levelof primary education (PSIC)]. *İlköğretim Online, 9*(2), 446–458. http://ilkogretim-online.org.tr/index.php/io/article/view/1774/1610.

Socratous, C., & Ioannou, A. (2021). Structured or unstructured educational robotics curriculum? A study of debugging in block-based programming. *Educational Technology Research & Development, 69*(6), 3081–3100. https://doi.org/10.1007/s11423-021-10056-x

Spiegel, J. A., Goodrich, J. M., Morris, B. M., Osborne, C. M., & Lonigan, C. J. (2021). Relations between executive functions and academic outcomes in elementary school children: A meta-analysis. *Psychological Bulletin, 147*(4), 329–351. https://doi.org/10.1037/bul0000322

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, d4002. https://doi.org/10.1136/bmj.d4002

Strawhacker, A., & Bers, M. U. (2019). What they learn when they learn coding: Investigating cognitive domains and computer programming knowledge in young children. *Educational Technology Research & Development, 67*(3), 541–575. https://doi.org/10.1007/s11423-018-9622-x

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders, 61*(3), 201–216. https://doi.org/10.1016/S0165-0327(00)00338-4

Traverso, L., Viterbori, P., & Usai, M. C. (2015). Improving executive function in childhood: Evaluation of a training intervention for 5-year-old children. *Frontiers in Psychology, 6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00525.

Tsai, C. Y. (2019). Improving students' understanding of basic programming concepts through visual programming language: The role of self-efficacy. *Computers in Human Behavior, 95*, 224–232. https://doi.org/10.1016/j.chb.2018.11.038

Tsarava, K., Leifheit, L., Ninaus, M., Román-González, M., Butz, M. V., Golle, J., Trautwein, U., & Moeller, K. (2019). Cognitive correlates of computational thinking: Evaluation of a blended unplugged/plugged-in course. *Proceedings of the 14th Workshop in Primary and Secondary Computing Education, 1–9*. https://doi.org/10.1145/3361721.3361729

Unterrainer, J. M., & Owen, A. M. (2006). Planning and problem solving: From neuropsychology to functional neuroimaging. *Journal of Physiology Paris, 99*(4), 308–317. https://doi.org/10.1016/j.jphysparis.2006.03.014

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

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods, 47*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P. T., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods, 7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Ward, C. S. D. (1993). *Developmental versus academic mathematics education: Effects on problem-solving performance and attitudes toward mathematics in kindergarten children.* Nashville: Peabody College for Teachers of Vanderbilt University.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of intelligence-second edition (WASI-II)*. San Antonio, TX: NCS Pearson.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*(1), 43–53. https://doi.org/10.1037/a0016738

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

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Yaşar, O. (2018). A new perspective on computational thinking. *Communications of the ACM, 61*(7), 33–39. https://doi.org/10.1145/3214354

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1*(2), 198–226. https://doi.org/10.1037/1089-2680.1.2.198

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., & Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3). https://doi.org/10.1111/j.0037-976x.2003.00260.x. vii–137.

Zhang, J. X., Liu, L., Pablos, P. O., & She, J. (2014). The auxiliary role of information technology in teaching: Enhancing programming course using alice. *International Journal of Engineering Education, 30*(3), 560–565.

Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education, 141*, Article 103607. https://doi.org/10.1016/j.compedu.2019.103607