



Metacognitive control during problem solving at early ages in programming tasks using a floor robot

Javier Del Olmo-Muñoz¹ · Pascual D. Diago² · David Arnau² · David Arnau-Blasco³ · José Antonio González-Calero¹

Accepted: 31 July 2024
© The Author(s) 2024

Abstract

This research, following a sequential mixed-methods design, delves into metacognitive control in problem solving among 5- to 6-year-olds, using two floor-robot environments. In an initial qualitative phase, 82 pupils participated in tasks in which they directed a floor robot to one of two targets, with the closer target requiring more cognitive effort due to the turns involved. The results of this phase revealed that younger students often rationalised their decisions based on reasons unrelated to the difficulty of the task, highlighting limitations in children's language and abstract thinking skills and leading to the need for a second quantitative study. In this subsequent stage, involving 117 students, a computerised floor-robot simulator was used. The simulator executed the students' planned movements and provided feedback on their validity. Each participant had three attempts per problem, with the option to change their target each time. The simulator stored the information pertaining to the chosen resolution path, design of the plan, and re-evaluation of decision making based on the results and feedback received. This study aims to describe the criteria upon which students base their metacognitive control processes in decision making within problem-solving programming tasks. Additionally, through a comparative analysis focusing on age and gender, this research aims to assess the relationship between metacognitive processes and success in problem-solving programming tasks.

Keywords Metacognitive control · Problem-solving · Floor robot · Programming tasks · Early ages

1 Introduction and literature review

In this study we utilise a computer floor robot to pose programming tasks that involve metacognitive control. By observing how students pre-estimate the difficulty of the tasks and how they modify these decisions based on success and the task constraints, we intend to contribute to the understanding of metacognitive control in early ages. Importantly, the present study also acknowledges the role of individual differences, such as age and gender, in problem-solving and decision-making processes. Prior research has shown that these factors can influence the strategies students use and their overall performance in problem-solving tasks (e.g., Fessakis et al., 2013). In the following section, we

offer a pertinent literature review on metacognition and floor robots as a metacognitive control study aid. Subsequently, we outline the research objectives and the theoretical framework elucidating the foundations of metacognition, age and gender differences, and programming tasks. Finally, we present the methodology of our study, along with the results and discussion corresponding to the research objectives.

1.1 Metacognition in mathematical problem solving

Formulating an exact definition of metacognition has proven to be a challenging task due to its multifaceted nature, which at times lacks clear connections between its various interpretations. According to Garofalo and Lester (1985, p. 164) “cognition is involved in doing, whereas metacognition is involved in choosing and planning what to do and monitoring what is being done”. More specifically, Stillman (2020) defines metacognition as “any knowledge or cognitive activity that takes as its object, or monitors, or regulates any

✉ Pascual D. Diago
pascual.diago@uv.es

¹ University of Castilla-La Mancha, Albacete, Spain

² University of Valencia, Valencia, Spain

³ Pompeu Fabra University, Barcelona, Spain

aspect of cognitive activity; that is, knowledge about, and thinking about, one's own thinking" (p. 445).

However, as Desoete and De Craene (2019) point out, there remains an ongoing discussion regarding the true nature of metacognition and the methods by which it can be comprehended, evaluated, and enhanced. In the midst of this discussion, we can identify a certain consensus in dividing metacognition into two subdomains: "knowledge" about one's cognitive processes and "regulation" of those processes (Flavell, 1976). This dichotomy is well-recognised in metacognitive literature, with various terminologies used interchangeably to describe these two core dimensions of metacognition (Lucangeli et al., 2019; Mahdavi, 2014). "Metacognitive knowledge", sometimes referred to as "knowledge of cognition", encompasses an individual's attitudes, awareness, and emotional insights, regarding how the mind operates, aiding in the comprehension of various cognitive processes. In contrast, "metacognitive control", sometimes referred to as executive control or regulation of cognition, represents a series of deliberate actions, such as planning, monitoring, or evaluation, taken by individuals to supervise and direct their own thinking or learning effectively.

In addition, there is also a lack of agreement on whether metacognition should be conceptualised as a general domain-independent construct, or whether metacognitive processes are domain-specific (Desoete & De Craene, 2019). Beyond the general or particular nature of metacognition, the existence of an important relationship between metacognition and mathematical competence, in general, and in mathematical problem solving, in particular, has been demonstrated (Kuzle, 2018). However, metacognition as an object of study in mathematics education was introduced with a certain delay compared to other disciplines. According to Garofalo and Lester (1985), most research at that time on mathematical problem solving had been focused on the study of processes, heuristics and strategies. As an example, Schoenfeld (1992) took into account five aspects as part of a framework for exploring mathematical knowledge in problem solving processes: the knowledge base, problem solving strategies, monitoring and control, beliefs and affects, and practices. Later, this author associated monitoring and control with self-regulation, included both processes within the more general term of metacognitive control, and highlighted the importance of monitoring and control in processes of decision making and in the design and re-elaboration of the plan to solve a problem.

1.2 Challenges in the study of young children metacognition

A factor already mentioned, not exclusive to these educational levels, is the absence of a universally accepted

standard for assessing metacognition, that makes it difficult to compare research outcomes (Desoete & De Craene, 2019). Besides, from the point of view of mathematical activity, Garofalo and Lester (1985) point out that metacognition is an activity that is difficult to identify and analyse, since the experimental techniques used to observe metacognition could influence problem-solving processes.

In the study of metacognition in kindergarten, several challenges need to be addressed. These difficulties have been reflected in a disproportion of studies at different educational levels. Thus, while metacognition's contributions to learning have been extensively explored in elementary, secondary, and higher education, it has been somewhat overlooked in the context of elementary education (Chen & McDunn, 2022). These authors point out the important limitations of older studies carried out with preschool students where they attempted to draw conclusions about their metacognitive processes from think-aloud protocols. The limitations in the verbal skills of early-age students have been widely described in the literature, as seen in studies such as that of Donaldson and Balfour (1968). Therefore, it is necessary to create mathematical tasks where, in addition to the possible verbalization of the students, metacognitive aspects can emerge implicitly through the students' performances (Chen & McDunn, 2022).

1.3 Floor-robots as a metacognitive control study aid

As recommended by Whitebread and Coltman (2010), in order to generate situations conducive to evoking and supporting young children's articulation of emerging metacognitive control skills, we will seek to situate tasks in contexts that make sense to children. In line with the Spanish curriculum, which emphasises computational thinking as a fundamental component of early mathematics education, our tasks will be designed to reflect the kind of problem-solving and logical reasoning activities that are encouraged at these educational levels. With that purpose, this study builds on the foundational work of Papert (1980), who set the stage for the use of programming tasks and floor-robot simulators for children to develop their mathematical and metacognitive skills. As one of the creators of the Logo programming language in the 1960s, Papert introduced the concept of "turtles", physical floor-roaming robots that could be programmed with Logo which children could control and move. This hands-on approach allowed children to visualise and interact with mathematical concepts in a novel and engaging way, thereby facilitating problem-solving (Feurzeig et al., 1970).

2 Research objectives

In the context of this study, we designed a mixed-method experiment aimed at studying the criteria of decision making in early-age students when solving mathematical problems. As a general aim, we sought to determine whether students at these levels are capable of considering aspects related to the complexity of the task as an element of metacognitive control, both in the initial planning of the resolution and, if needed, in the reformulation of the plan. Bearing this in mind, we designed a sequential mixed-method study, combining qualitative and quantitative techniques. The first phase of the study, with a predominant qualitative approach aimed to identify the criteria on which students based their decision making when they had to choose between two problem-solving strategies. For the second phase, some aspects observed in the previous qualitative phase were taken into consideration, and programming tasks were designed to allow for the massive collection of data, with a quantitative approach. This second phase aimed to elucidate the connections between the factors influencing metacognitive processes and students' success in solving programming tasks, considering both gender and educational level. In order to evaluate plan-design and decision-making behaviours within the context of problem-solving programming tasks, the study aim to address the following objectives:

- **O1.** To describe and categorise the criteria on which students based their metacognitive control processes in decision making in problem-solving programming tasks.
- **O2.** To assess the relationship between students' metacognitive control processes when devising and reformulating a plan and the proficiency in problem-solving programming tasks based on age and gender.

3 Theoretical framework

In the following we delve on the elements that underpin the analysis of the data collected in our study: metacognition, decision making, and the consideration of age and gender differences, all of which are relevant to problem-solving in early mathematics education. Additionally, we revise the metacognitive control processes involved in problem-solving programming tasks through robotics.

3.1 Metacognition, decision making and problem solving in early mathematics education

The study of metacognition in the context of mathematics education has been a subject of interest for many

researchers. Schoenfeld (1992) discussed the importance of understanding the nature of problem solving, providing a comprehensive overview of metacognition in the context of mathematical problem solving. Alan Schoenfeld stressed that problem solving, in the spirit of George Pólya, is about learning to cope with new and unfamiliar tasks when the relevant solution methods are not known. He also delineated metacognition into three distinct categories: (1) the conscious understanding individuals have about their own cognitive processes, (2) the self-regulatory strategies, including monitoring and real-time decision making, and (3) the influence of personal beliefs and emotions on an individual's performance. The present study is concerned with monitoring and real-time decision making. These metacognitive abilities are used when planning and assessing a problem-solving process. Both self-regulatory strategies, which would be included within the metacognitive control or metacognitive skills (Mahdavi, 2014; Veenman & van Cleef, 2019), describe a perspective that closely matches Pólya's (1957) classical stages of mathematical problem solving. In this vein, in an overview of the current state of research in this area, Desoete and De Craene (2019) note that metacognition is one of the key predictors of mathematical performance. Although the activation of metacognitive processes in children has been a topic of debate, there is a consensus that early years are critical for nurturing metacognitive skills (Temur et al., 2019), in a process that begins to develop during preschool or early school years and reaches its full bloom around the age of 12.

A reason that may have contributed to this lack of consensus is the difficulty to identify and measure metacognitive processes. According to Veenman and van Cleef (2019), in the study of metacognitive processes we can differentiate between on-line and off-line methods. The off-line methods are mainly based on the analysis of questionnaires that are administered before or after the task. The on-line methods are carried out on observations of the performances when facing the tasks that are collected by the observer in real time or stored in computer-logfile registrations. Although, according to these authors, on-line methods are preferable, their design should avoid interfering with decision-making processes when solving mathematical problems. In the study of Lingel et al. (2019), different measurement procedures were compared to identify the ability to carry out metacognitive monitoring using different configurations of off-line methods when a group of German children in grade 7 solve mathematical tasks. The results show students' overconfidence when assessing the difficulty of a task. In addition, these judgments become more accurate in retrospective evaluations, although maintaining an excess of confidence.

3.2 Age and gender differences in metacognition

The influence of age and gender on metacognitive control in education has been the subject of numerous investigations. In a study conducted by Zimmerman and Martinez-Pons (1990), they investigated self-regulated learning in students, considering factors like grade, gender, and giftedness. The study included boys and girls from 5th, 8th, and 11th grades. In general, 11th graders excelled, followed by 8th graders, and then 5th graders in self-regulated learning measures. Girls were found to have used self-regulated learning strategies more frequently than boys, particularly in areas such as record-keeping, monitoring, environmental structuring, goal setting, and planning.

In order to assess the relationship between gender, age and metacognition, Mok et al. (2007) conducted a study involving 8948 students aged 9–17 years, which revealed significant gender differences in metacognition. Compared to boys, girls reported greater knowledge of metacognitive strategies, use of learning strategies, regulation of learning and evaluation of learning. The research affirmed that both sexes use metacognitive strategies for learning, but girls tend to use them more frequently. In addition, the study observed a decline in students' metacognitive competences with age, from 9 to 17 years old.

In relation to this matter, Moè (2009) analysed gender differences in decision-making processes among individuals aged 15–22, revealing that women tend to show greater risk aversion and less confidence in their decisions compared to men. However, she notes that these differences could be attributed more to social expectations and stereotypes than to innate qualities, suggesting the malleability of these traits. Moè proposes to further investigate possible mediating factors, such as the tendency to guess or to be more cautious in the face of uncertainty.

In summary, although existing research has provided information on age and gender differences in metacognition among older students, extending this research to a younger age group may offer a more comprehensive understanding of metacognitive skills abilities and metacognitive control.

3.3 Programming tasks in young learners through robotics

Following Papert's pioneering work, a large-scale study by Clements et al. (2001) further examined the impact of the Logo programming language on children's cognitive development. This study involved children from kindergarten through to sixth grade using Logo and found that these children scored significantly higher on tests of mathematics, reasoning, and problem-solving. An earlier study by Clements (1987) specifically focused on kindergarten children using Logo. It was observed that these children

demonstrated sustained attention and self-direction during their interactions with Logo. These studies, in conjunction with Papert's seminal work, emphasise the value of programming tasks like Logo in fostering cognitive and metacognitive skills in children, laying the groundwork for the use of modern programming tasks and floor-robot simulators in current research. Since then, the application of robotics and programming tasks in elementary education has been gaining momentum due to its potential in enhancing young children's (aged 7 and under) cognitive and metacognitive skills, including problem solving and decision making (Bers, 2020).

Atmatzidou et al. (2018) delved into age and gender differences in metacognition and problem-solving processes within the educational robotics context (using Lego NXT robotics tools). They implemented various modes of guidance among two groups of students aged 11–12 and 15–16 years. The study revealed that strong problem-solving teacher guidance (vs. minimal guidance) can positively influence students' metacognitive and problem-solving skills. In addition, it was found that regardless of their age or gender, students eventually reach the same level of metacognitive and problem-solving skill development.

In early ages, robots coded through tangible programming environments are one of the most notable robotic initiatives. These robots offer a combination of computer programming experience in and with the physical world (Horn et al., 2012). A common characteristic of these robots is the presence of objects to build computer programs without computer screens or keyboards. Tangible robotics are specifically aimed at teaching a particular subset of mental tools to young children, including problem solving and decision-making skills (Bers, 2020). Further research by Sullivan and Bers (2016) explores gender differences in children's performance on robotics and programming tasks. Their study, focusing on kindergarten to 2nd grade students, found no significant gender differences in basic robotics and simple programming tasks. However, on more complex programming tasks that required a deeper understanding of mathematical concepts related to number sense (such as cardinality and counting), boys outperformed girls.

In parallel with these developments, the Bee-Bot, a floor robot designed for use by early years students, emerged as another practical tool to introduce young learners to programming concepts. Highfield et al. (2008) studied how young children's use of Bee-Bot can promote the development of mathematical problem-solving and meta-cognitive processes. They also wanted to observe what forms of mathematical reasoning and strategic thinking are involved when children plan and program a Bee-Bot to solve problems. Throughout the research, the children exhibited problem-solving strategies and relational thinking, enabling them to plan, code, and manipulate the robot along a complex

pathway. Carrying this concept further into the realm of virtual environments, the work of Diago et al. (2019) presents a virtual counterpart to the physical Bee-Bot. By offering a floor-robot simulator for children to engage with programming tasks and problem-solving scenarios, this environment provides a versatile platform to explore decision making and other cognitive skills. The data gathered from the software extends beyond mere success or failure metrics; it offers valuable insights into the cognitive processes involved in student monitoring and management of various resolution plans.

4 Methodology

To address the stated objectives, a sequential mixed design was proposed. Specifically, the design followed an exploratory sequential design, in which qualitative data are collected first, followed by quantitative data, from a larger sample, used to analyse the phenomenon systematically and generalise the findings (Cohen et al., 2018). Thus, the study was composed of two sequential phases: an initial phase, in which both the data and the predominant analysis were qualitative in nature (hereinafter, Qualitative Phase, QUALP) and a second quantitative phase (hereinafter, Quantitative Phase, QUANP).

As part of the sequential mixed-method design of this study, the results obtained in QUALP study served to determine the study variables and the design of the programming tasks employed in QUANP study. Specifically, qualitative arguments related to the trajectory the robot should follow, the number of turns included in the path, or the distance/proximity to the target-flower were taken into account, as we will describe below. Thus, QUALP study along with the data collected in QUANP enabled the addressing of objective 1. Likewise, QUANP study enabled the addressing of objective 2, analysing success rates, route choices, and the effect of modifying the initial plan in floor-robot programming tasks, considering the age and gender factors mentioned in the objectives.

4.1 Participants

A convenience sampling method was used in the study. Two different samples of students were used for QUALP and QUANP studies, since the objectives were intended for young students being confronted for the first time with programming tasks with floor robots. Otherwise, students in the second study might have employed skills or knowledge acquired during the first study, and the results derived from the second study could be biased.

In QUALP, 82 students from two different Spanish schools participated. Among them, 51 subjects were in the

final year of kindergarten (KG), and 31 subjects were in the first year of primary education (PE). All the participants were interviewed individually and, to delve deeper into the students' reasoning, 21 of them (13 KG and 8 PE students) were randomly selected for a more extended interview.

In QUANP, 117 children from a third Spanish public school participated. As in QUALP, they belonged to two different grade levels: 60 in the first year of PE and 57 in the last year of KG. In the groups used as samples in either QUANP or QUALP, there were no students who had been diagnosed as students with special educational needs or had repeated any grade, so no adaptation in the study was necessary.

4.2 Procedure

4.2.1 Procedure in QUALP

QUALP was conducted during school hours in the classroom. The intervention lasted between 30 and 40 min for each student. During this session, an initial direct instruction was provided, demonstrating the basic operation of the floor robot to all participants. After the direct instruction stage, data collection was conducted individually with the path-selection problems in a specific room. An extended interview was carried with each of the participants. In this study, the tangible user interface made up of Bee-Bot and a card-box system (Fig. 1) was used with the trajectories traced with red tape. In what follows, we expand upon this information.



Fig. 1 Bee-Bot and the card-box system used in QUALP

Procedures for direct instruction and problem solving with single-path task (instruction stage)

Prior to the administration of the target problems in this study, direct instruction was conducted. Figure 2 shows the problems that were used for direct instruction (I1–I5). During the instruction stage, students, as a whole group, were provided with the opportunity to observe a researcher's problem-solving approach while tackling the five problems depicted. The latter problem example served the purpose of instructing participants on the process of flower selection within the subsequent two-flower tasks that would appear later in the intervention. This stage lasted approximately 20 min.

Procedures for two-flower tasks (experimental stage)

After the instructional stage, participants were individually presented with the task depicted in Fig. 1. Intentionally, the route to reach the nearest flower included the necessity to turn in comparison to the flower further away, which offered a straight route. Firstly, participants were asked to select a flower to which they wanted the robot to move, and they were requested to explain their choice. Subsequently, and prior to programming the robot, participants prepared the instructions arranging the set of cards in the programming box (Fig. 1). Finally, the students programmed the robot with the arranged instructions to complete the chosen path. Participants only had one attempt, as we were only interested in the reasons regarding the chosen route and not in the proficiency. The duration of this experimental stage was approximately 10 min for each student.

Finally, based on the availability of time allocated to the experiment, 21 participants were selected to undergo a more detailed interview. Once the task on Fig. 1 was presented, participants were asked to select the flower they wanted to reach. Next, they had to plan the resolution by ordering the set of cards in the programming box and, later, program the sequence in the robot. After that, they were interviewed to explain their decisions about the choice of the flower, which one was more difficult to reach and why. With an average duration of about 10 min, these interviews were video recorded and later transcribed.

4.2.2 Procedure in QUANP

This phase took place in a single session, during school hours, in which participants, in groups of 20–25 subjects, attended to a classroom equipped with computers to use the floor-robot simulator. Two stages were involved in this experiment, a direct instruction stage lasting about 30 min, devoted to explaining the simulator; and an experimentation stage, lasting about 15 min, in which the target problems were solved by the participants.

Overview of the computer floor-robot simulator and programming tasks used

A customised graphical software was developed to simulate a tangible scenario involving Bee-Bot and the card-box system implemented in QUALP. The design principles of this software are extensively detailed in Diago et al. (2019). Each student was provided with a computer with a touch screen to interact with the software. As in QUALP, all the programming tasks involved a bee-shaped robot that must travel a given path until it reaches a flower (Fig. 3). The findings derived from QUALP (outlined in Sect. 5.1) guided the development of the programming

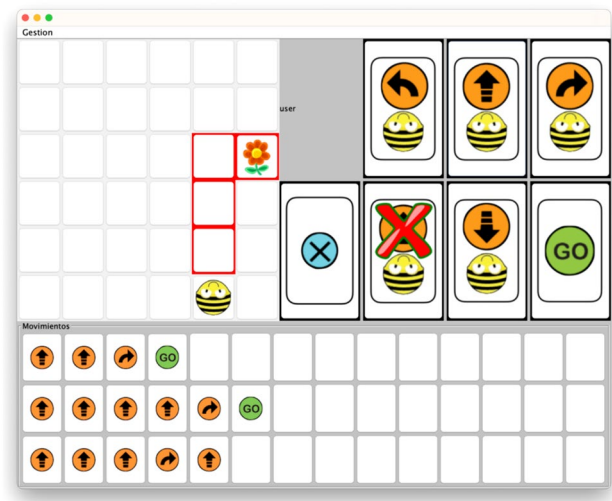


Fig. 3 Graphical interface of the floor-robot simulator

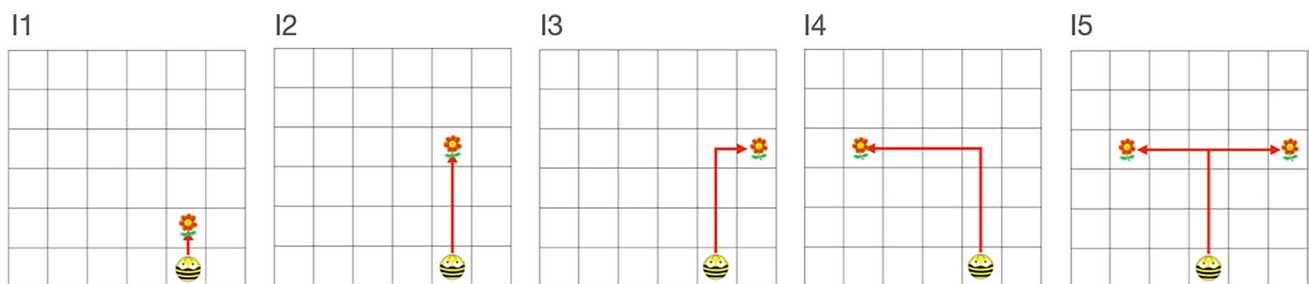


Fig. 2 Instruction programming tasks

tasks utilized in QUANP. Specifically, certain observations made during QUALP influenced the design of tasks, incorporating elements such as the trajectory of the robot, the number of turns in the path, and the distance to the target-flower. Consequently, this investigation employs two distinct types of programming tasks: activation problems and selection problems.

Activation problems entailed the provision of a singular path for the robot to follow, ultimately leading it to the flower. Conversely, selection problems introduced two distinct paths of varying complexities, compelling the solver to engage in a preliminary evaluation of the available options and subsequently select a path based on predetermined criteria, such as logical reasoning, mathematical analysis, idiosyncratic considerations, etc.

As illustrated in Fig. 3, the interface of the application effectively presents information to the user through graphical language (i.e., the arrows and symbols). The software stores information on each user's action enabling the monitoring of the resolution path taken by the student. The user interface comprises three distinct areas:

1. The problem-statement area (top-left section): This area shows the path (coloured boxes) to be followed by the robot.
2. Programming blocks for robot movements (top-right section): This area displays the available programming blocks corresponding to the robot movements and includes options to delete individual instructions or the entire plan.
3. The path-planner (bottom section): In this area, the user can view and sequence the movement blocks that will be executed by the robot when the "GO" instruction is received. The path-planner is divided into three lines, representing the maximum number of attempts allowed for the user to solve the problem (currently set at three). The aim of this area is to facilitate both the planning and the debugging solution paths foreseeing the movements of robot, and checking the movements based on the user choices by tracking the robot's movement.

Procedures for direct instruction and problem solving with single-path task (instruction stage)

Prior to the administration of the target problems in this research, direct instruction was conducted to familiarize the participants with each area of the graphical interface. Special attention was dedicated to elucidating the overarching objective they were to encounter: "guide the bee to the flower, regardless of which of the two flowers is presented in the two-flower problems". The protocol and problems used in the direct instruction phase for QUANP were the same that in the QUALP (see Fig. 2), but in the simulator environment.

After the direct instruction stage, the intervention began. Students started using the application by addressing activation problems (denoted as A1 to A4) shown in Fig. 4. For each problem, students were granted three attempts to devise a resolution path. When an unsuccessful path was executed, the attempt remained visible and locked, thereby affording the user an opportunity to review and formulate a new plan for subsequent attempts. Upon the successful determination of the solution path, the robot effectively reached the corresponding flower, triggering the launch of the next problem within the preloaded collection of tasks. If all three attempts were completed, the program automatically proceeded to load the next problem in the collection.

Procedures for two-flower tasks (experimental stage)

Upon the completion of the fourth activation problem, the application automatically launched the subsequent collection of selection problems. Participants were then presented with a total of four programming tasks featuring two flowers, illustrated in Fig. 5 and denoted as S1–S4.

The procedural steps and application usage during the selection problems phase remained consistent with the approach adopted for the activation problems. The primary distinction was the inclusion of an additional step prior to commencing each attempt. Specifically, the program prompted the selection of a flower, temporarily disabling the panel containing programming blocks for robot movements (as depicted in the left image of Fig. 6). Once a flower was freely chosen, the alternative flower became deactivated (in

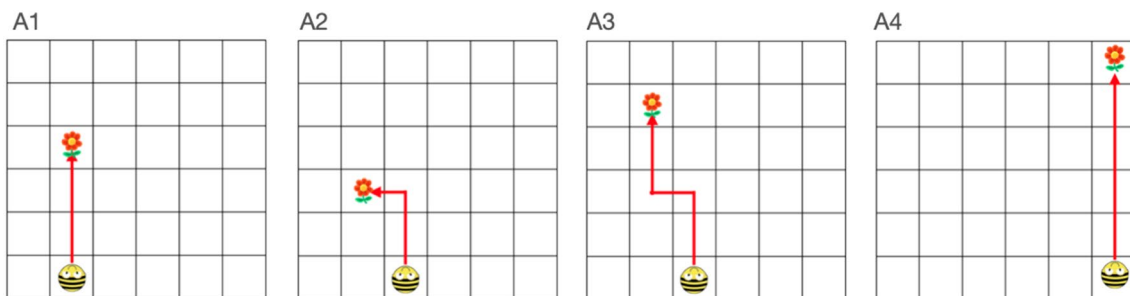


Fig. 4 Activation programming tasks

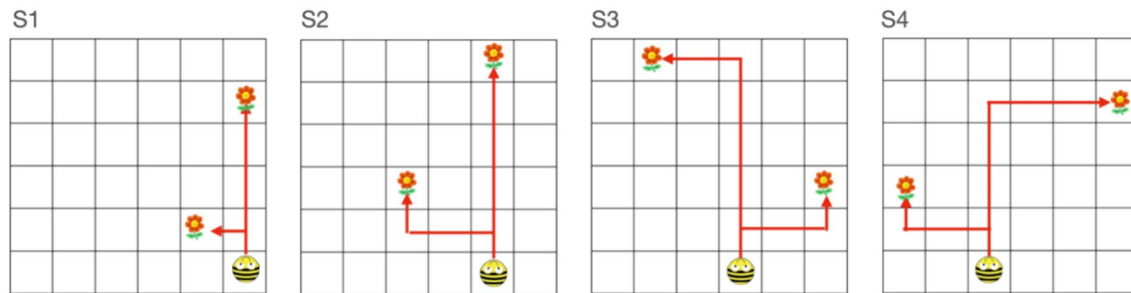


Fig. 5 Selection programming tasks



Fig. 6 Flower selection procedure in the program: movement buttons locked while the flower is not selected (left panel); buttons active when the flower is already selected (right panel)

grey, see Fig. 6), and the robot movements panel was subsequently enabled (as illustrated in the right image of Fig. 6). Subsequently, users could proceed with programming the movements in a manner analogous to the approach employed in the preceding problems. It is important to underline that participants could modify their decision to go to one flower or another because, as mentioned earlier, the application was designed in such a way that, in each attempt, the user was given the option to choose a flower.

4.3 Data collection and analysis methods

In QUALP, for the categorization of the arguments that participants put forward when choosing one flower over another, the content analysis technique was applied. Thus, starting from the transcription of the interviews, a coding of participants' responses was carried out, aimed at identifying patterns and classifying responses into emergent categories. As a result of this process, four dominant categories were identified, sufficient to encompass all the motivations provided by the participating students, namely: (1) *task-related arguments*, for explanations explicitly based on trajectory

characteristics (e.g., "because [the other] turned here" or "because it was straighter"); (2) *distance-related arguments*, when explanations are grounded in the distance between the starting point and the chosen flower (e.g., "because it's closer"); (3) *personal arguments*, when the explanation originated from criteria external to the situation (e.g., "because I like it"); and (4) *I don't know/no answer*, when participants did not provide any response or claimed not to know why they had chosen one flower over another.

In QUANP, concerning the data collection methods, for each participant, the application recorded anonymised information in CSV format related to the specific problem, the attempt number, the programmed sequence of instructions per attempt, the success or failure of the programmed path, as well as demographic data such as gender and educational stage. This data was subsequently utilised to compute success rates per problem for each attempt, as well as an overall binary score that considered the success ("1") or failure ("0") for each problem. Additionally, in the case of selection problems, the application also stored the selected flower for each attempt, thus providing insights into the decision-making process of the students. Following data

analysis, two students who did not complete the full intervention sequence were excluded from the study, resulting in a final sample size of 115 students.

Regarding the analysis method, in QUANP, and considering the stipulated allowance of three attempts per problem, for each activation or selection problem, we compute the maximum score obtained in the initiated attempts. In order to complete the aim of this study (O2), we analysed the average success rates in the target problems taking into account the factors age, gender, and type of problem (activation or selection). In addition, we introduced a finer distinction within the activation and selection problems: the “route type” factor. Specifically, we differentiated four levels in all target problems: (a) *One flower—Straight*, when there was only one flower and it could be reached without the need for a turn; (b) *One flower—Turn*, with one flower where at least one turn was required; (c) *Two flowers—Straight or turn*, with two flowers where the less complex flower could be reached with a straight path and the second flower required at least one turn; (d) *Two flowers—Turn or turn*, where both flowers required the execution of a turn, with the more complex one requiring two turns and the simpler one requiring one turn. Accordingly, problems A1 and A4 were *One flower—Straight*-type; and A2 and A3 were *One flower—Turn*-type. On the other hand, S1 and S2 were *Two flowers—Straight or turn*-type and S3 and S4 were *Two flowers—Turn or turn*-type. By making this differentiation, we aimed to gain deeper insights into the specific types of decisions made by the students. All statistical analyses were conducted using the R software, with a predetermined significance level of 0.05. Effect sizes were evaluated using the explanatory measure. The normality of the data was assessed using the Shapiro–Wilk test. In instances where the assumption of normality was violated, the nonparametric Mann–Whitney test was employed in order to compare means as different participants take part in each condition.

5 Results

5.1 Results from QUALP

Table 1 summarises the information gathered from the experimental phase in QUALP, displaying, for each educational level, the frequency of participants broken down according to two variables: the chosen path (straight trajectory or trajectory involving turns) and the justification for the choice of the flower (*task-related arguments*, *distance-related arguments*, *personal arguments* or *I don’t know/no answer*). In addition to the frequency in each category, the percentage relative to the total number of participants in each educational level is provided. Thus, for example, a value of 3.92% in the second row represents the percentage of participants in KG who have chosen the turn trajectory flower and whose choice is due to reasons related to intrinsic task characteristics.

The information gathered in the interviews of the 21 students allowed us to observe, beyond eventual differences in programming skills, distinct patterns of performance and reasoning between PE and KG students. Although simple choice predominates in both groups, there is a greater tendency among KG students to choose the flower involving turns (35.29%) compared to PE students (25.81%). Similarly, delving into the differences between educational levels, there is clear evidence of KG students tending to make decisions based on non-task-related criteria (72.55% versus 27.45%; this percentage consists of the categories “distance-related”, “personal” and “no answer” from Table 1). Thus, it becomes evident that a significant number of KG students did not plan their actions based on an anticipation of the difficulties or characteristics involved in solving the task (e.g., EC-S73: “Because it looks like a slide”; EC-S85, EC-S03: “Because I like it more”). In some cases, the arbitrariness of the participants’ decisions became apparent. This is highlighted in the following dialogue between the interviewer (I) and one KG participant EC-S09:

Table 1 Summary of flower choices organised by educational level and type of reasoning in QUALP

Argument	Task-related		Distance-related		Personal		I don’t know/no answer		Total	
	N	%	N	%	N	%	N	%	N	%
Stage/flower										
KG	14	27.45	19	37.25	12	23.53	6	11.76	51	100.00
Turn	2	3.92	9	17.65	5	9.80	2	3.92	18	35.29
Straight	12	23.53	10	19.61	7	13.73	4	7.84	33	64.71
PE	9	29.03	17	54.84	2	6.45	3	9.68	31	100.00
Turn	0	0.00	7	22.58	1	3.23	0	0.00	8	25.81
Straight	9	29.03	10	32.26	1	3.23	3	9.68	23	74.19
Total	23	28.05	36	43.90	14	17.07	9	10.98	82	100.00

I:	Why did you choose that flower? [S09 had indicated that they chose the simple trajectory (without turns)]
EC-S09:	Because I like those flowers (sic)
I:	They are the same (pointing to one and the other), why do you like that one more?
EC-S09:	Because it's very cute
I:	But it's the same as the other
EC-S09:	But I like it more

On the contrary, among PE participants, there was a greater ability to take into consideration relevant elements of the trajectory that affect the complexity of programming the robot's path in the decision-making process, e.g.:

PE-S14:	"because you only have to go straight ahead".
I:	You chose the straight path, why?
PE-S11:	Because it looked easier.
I:	Why did it look easier?
PS-S11:	Because it's only 5 lines, and that's it (referring to five robot advances).
PE-S39:	"Because it's easier this way, and I don't have to turn".

Another noteworthy aspect emerging from the case study is the tendency to choose the trajectory involving turns by arguing a shorter linear distance between the starting point and the flower, without considering that programming the complex path requires programming turns in two different directions, e.g.:

I:	Why did you choose this flower here (points to the turn flower) and not this one (points to the straight one)?
EC-S07:	Because this one (the turn flower) is very close, and that one (the straight one) is very far away.
EC-S74:	Because it's the shortest path.
EC-S83:	It's lower down.
I:	What do you mean?
EC-S83:	Because the path is shorter.

5.2 Results from QUANP

5.2.1 Findings related to students' proficiency due to gender, education level and task features

Regarding the participants' proficiency considering gender and educational stage factors, as well as the characteristics of

the problems, Table 2 presents descriptive statistics for the eight target problems stratified according to those factors.

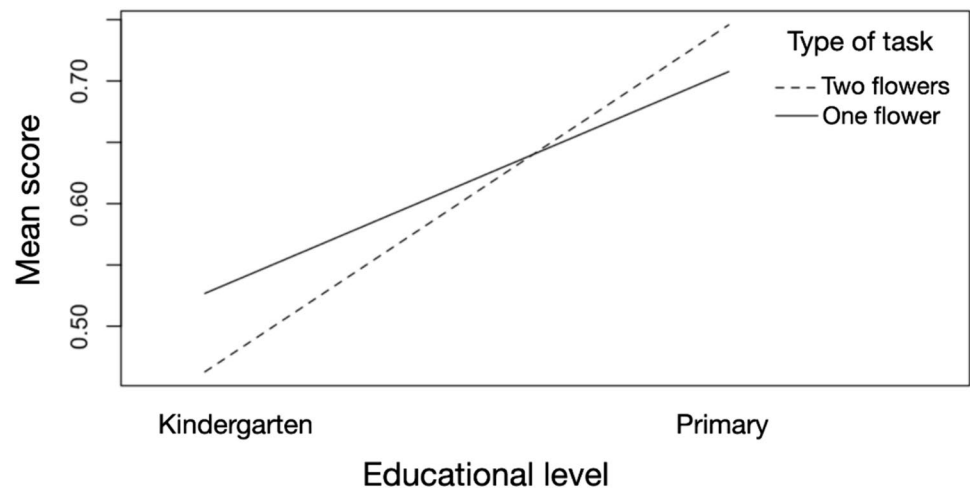
Thus, in order to evaluate potential differences based on participant characteristics (educational stage and gender) and activity type (one-flower or two-flower tasks), a mixed three-way ANOVA was used. The results revealed a significant main effect of the educational stage ($F(1, 111) = 30.75$, $p < 0.0001$), indicating that PE students outperformed KG students. Additionally, a significant interaction between educational stages and the type of tasks (one or two flowers) was observed, $F(1, 111) = 6.08$, $p = 0.0151$, suggesting that the availability of a second flower had a differential impact on students' proficiency depending on their educational level. No other significant main effects or interactions were found. As reflected in Fig. 7, this is due to the fact that participants in KG showed better performance in tasks involving a single flower than in tasks involving two flowers, while on the contrary, PE students obtained better results when faced with tasks that allowed them to choose which flower to approach.

Concerning the main effect, as observed in Table 2, there was a notable disparity in the global success rate between the two groups. Specifically, the elementary students exhibited a significantly higher average success rate compared to the KG students. This difference in means was found to be statistically significant ($W = 80,988$, $p < 0.0001$), with a small effect size ($r = -0.24$). These findings confirm that PE students outperformed the KG students in terms of success rates in the problem-solving programming tasks.

In order to further analyse the task characteristics and their influence on the complexity experienced by solvers, considering differences based on age and gender, a three-way mixed ANOVA was conducted once again. In this case, the within-variable refers to the characteristics of the route required to reach the flower. In particular, the route type factor consists of four levels: *One flower—straight*; *One flower—turn*; *Two flowers—straight or turn*; and, *Two*

Table 2 Descriptive analysis disaggregated by gender and educational level in QUANP

Educational stage	KG		PE	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
All problems	0.49	0.50	0.73	0.47
Female	0.47	0.50	0.70	0.46
Male	0.53	0.50	0.76	0.49
Activation (one flower)	0.53	0.50	0.71	0.46
Female	0.49	0.50	0.67	0.47
Male	0.58	0.50	0.74	0.44
Selection (two flowers)	0.46	0.50	0.75	0.44
Female	0.45	0.50	0.72	0.45
Male	0.48	0.50	0.78	0.42

Fig. 7 Interaction plot for educational stage and activity type**Table 3** Descriptive analysis on the “route type” factor by educational level

Route type problem	KG		PE	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
One flower—straight (A1, A4)	0.839	0.369	0.992	0.092
One flower—turn (A2, A3)	0.214	0.412	0.424	0.496
Two flowers—straight or turn (S1, S2)	0.732	0.444	0.941	0.237
Two flowers—turn or turn (S3, S4)	0.189	0.393	0.551	0.500

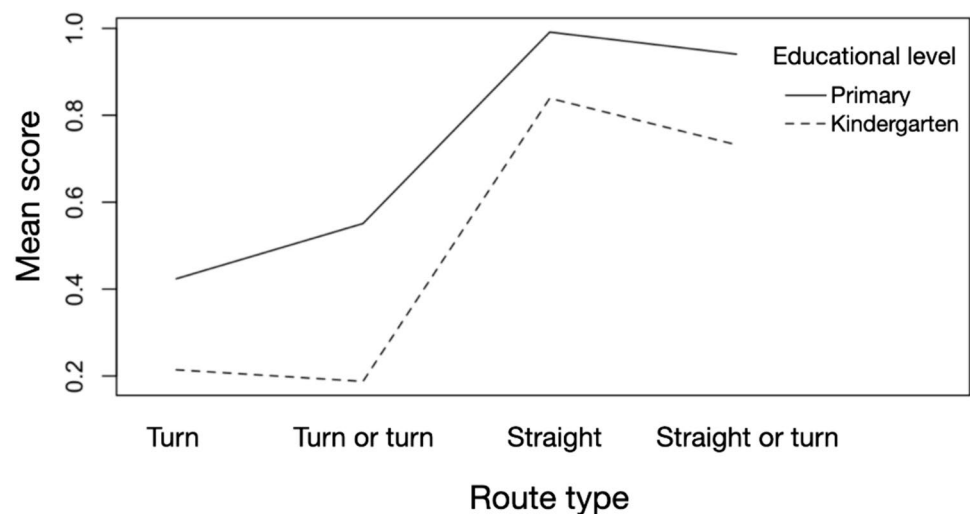
flowers—turn or turn. Table 3 shows the descriptive statistics broken down according to this factor.

Once again, the results indicated a significant interaction between educational level and the structure of paths associated with the task, ($F(3, 333) = 3.85, p = 0.0099$), what suggests again that the inclusion of a second flower affects the complexity of the tasks differently depending on the students' grade. Similarly, significant main effects were found

for both educational level ($F(1, 111) = 30.91, p < 0.0001$) and type of tasks ($F(3, 333) = 172.48, p < 0.0001$). No significant effect was found for gender. Figure 8 indicates that differences in competence between PE and KG students occurred across all types of tasks. Similarly, it confirms how tasks that involve the need to program turns are more challenging for the participants.

5.2.2 Findings related to the choice of resolution path, design and re-evaluation of the plan

The second objective, related to examining the metacognitive control processes in problem-solving programming tasks, is addressed by analysing two critical facets in decision making. On one hand, the initial choice made by the participants to identify the least complex route in selection tasks is evaluated. On the other hand, their ability to reformulate their initial plan in case of an error, considering the difficulty of the routes, is also assessed.

Fig. 8 Interaction plot for educational stage and route type

Concerning the initial choice, Fig. 9 shows, for each educational stage, the first chosen path in terms of percentages relative to the total number of students in each level. A closer examination of this relationship reveals that PE students showed greater competence in decision-making compared to KG students, particularly evident in the first two selection problems (S1 and S2), where one of the paths lacked turns. However, this ability to perceive the more complex path was diminished in the subsequent two problems (S3 and S4), in which both paths appeared similar (both involved turns) despite differing numbers of turns. Consequently, the distribution of choices made by PE students in their first attempts clearly favoured the simpler route. In contrast, the distribution of choices made by KG students was more uniformly distributed, indicating a more erratic decision-making process possibly influenced by the proximity of the flower in the more complex path.

Based on the observations made previously regarding flower choices, and once we have confirmed a trend of PE students making better choices in terms of complexity (devising a plan) than KG students, we proceed to take a deeper look at this issue considering students' gender. For this purpose, we will focus on studying tasks involving straight paths or turns (S1 and S2), as they exhibit a more pronounced difference in complexity, and this aspect could be reflected to a greater extent in decision-making processes. Thus, specifically, we will analyse the choices made in these tasks, broken down according to the educational level and gender of the participants. Table 4 clearly shows how girls

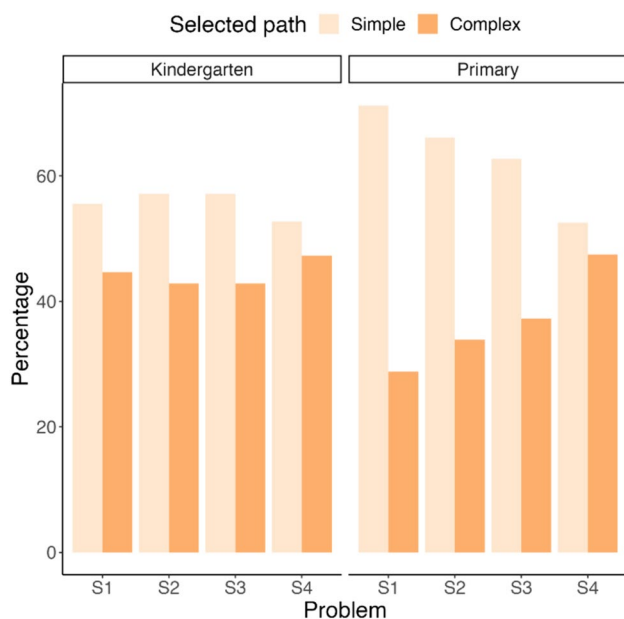


Fig. 9 Percentages of path election in the first attempt by selection problem

Table 4 First choice in problems S1 and S2, by educational stage and gender

Educational stage	Gender	Task S1		Task S2	
		Straight	Turn	Straight	Turn
KG	Female	21	13	24	10
	Male	10	12	8	14
PE	Female	23	6	23	6
	Male	19	11	16	14

tend to choose the simpler option to a greater extent than boys, regardless of the educational level.

However, this greater ability to choose an optimal route (devising a plan) does not translate into a higher score (executing the plan) compared to boys in S1. This is evidenced by the results of girls in KG ($M_{S1}=0.41$, $SD_{S1}=0.50$) and PE ($M_{S1}=0.69$, $SD_{S1}=0.47$), which are not significantly different from boys in the same educational stage (KG, $M_{S1}=0.50$, $SD_{S1}=0.51$; PE, $M_{S1}=0.70$, $SD_{S1}=0.47$), when comparing the success rates on the first initiated attempt (i.e., on their initial path choice) to solve task S1. In contrast, in task S2, which is characterised by a higher complexity in terms of turns, the best initial decision does lead to a higher success in problem-solving by females when computing the data from the initial path choice (KG-girls, $M_{S2}=0.44$, $SD_{S2}=0.50$; PE-girls, $M_{S2}=0.72$, $SD_{S2}=0.45$; KG-boys, $M_{S2}=0.22$, $SD_{S2}=0.43$; PE-boys, $M_{S2}=0.53$, $SD_{S2}=0.51$).

From the perspective of monitoring and evaluation processes, the changes in strategies that the solver may execute as a result of failure in the first attempt play an important role. Table 5 shows, for those students who did not solve the task correctly in their first attempt, the comparison between their initial choice and their final choice, indicating the changes in strategy as a result. The results show that, in general, students demonstrated the ability to modify their strategy and opt for a simpler route when faced with failure in their initial approach to the activity. However, the analysis of flower choices indicates that PE participants are more efficient in correcting their decision (revising the plan) compared to KG students, who tend to maintain their choice or even worsen it to a greater extent. This situation

Table 5 Last choice in problems S1 and S2, by educational stage and gender

Educational stage	Gender	Task S1		Task S2	
		Straight	Turn	Straight	Turn
KG	Female	25	9	30	4
	Male	17	5	16	6
PE	Female	26	3	28	1
	Male	21	9	28	2

is particularly observed in problem S2, which involves a complex route with two turns. While it was selected as the first attempt by many PE participants, especially boys, it was ultimately discarded by almost all of them.

The change in decisions (revising the plan) is reflected in a clear increase in achievement when completing the task. In fact, only one PE student was unable to program the task in problems S1 and S2, after choosing the simpler route. Thus, in PE, the gender differences observed in problem S2 disappear as a result of the widespread change in strategy among boys to the simpler route, resulting in success rates exceeding 90% in both S1 and S2 tasks (girls, $M_{S1}=0.89$, $SD_{S1}=0.31$; $M_{S2}=1.00$, $SD_{S2}=0.00$; boys, $M_{S1}=0.93$, $SD_{S1}=0.25$; $M_{S2}=0.93$, $SD_{S2}=0.25$). These mean values were derived from the success rates collected on the final attempt (in which some students had the opportunity to change their path choice) for tasks S1 and S2. However, in KG, differences in success rates are observed for each problem. In problem S1, where a similar proportion of participants from each gender ultimately chose the optimal flower, boys outperform girls in their ability to successfully program the robot (girls, $M_{S1}=0.65$, $SD_{S1}=0.49$; boys, $M_{S1}=0.82$, $SD_{S1}=0.39$). However, in problem S2, girls demonstrate a better ability to identify and choose the simpler route, resulting in better performance than boys ($M_{S2}=0.79$, $SD_{S2}=0.41$; boys, $M_{S2}=0.68$, $SD_{S2}=0.48$), when analysing the success rates on the last choice.

6 Discussion and conclusions

The first objective was to evaluate the metacognitive control processes that pupils use when having to choose between two problem-solving paths. Basically, a greater tendency in both QUALP and QUANP is observed among KG students to choose the path that requires them to make turns compared to PE students, which would mean a lower ability to integrate the difficulty of the task as an element of judgment in decision making. When justifying their decision-making process, KG students tend to rely more on personal arguments. The results seem to indicate that, when these types of justifications are presented, the choice of the solving path is arbitrary.

In both KG and PE students, it is observed in QUALP that those who argument their decision on the characteristics of the task mostly choose the one that requires a smaller number of turns. However, there does not seem to be a relationship between the distance-related and personal arguments and the choice of one or another path. This could be related to the weakness of using thinking-aloud protocols to analyse the cognitive and metacognitive processes at these ages, as claimed by Chen and McDunn (2022), which would be a limitation of the interviews conducted in our study.

The second objective focused on evaluating the performance of the participants considering educational stage and gender factors. The results from QUANP show a higher success rate among PE students. This result, to some extent logical, is complemented by a relevant result from the point of view of metacognitive control processes. PE students had higher success rates in two-flower tasks than in one-flower tasks, opposite to KG students. This fact, unrelated to coding skills—validates what was pointed out in the results from QUALP, based on the arguments given by students when choosing a path—emphasises the importance of the phases of devising and reformulating the plan in programming tasks. Regarding gender, no differences were observed between boys and girls in overall success in tasks neither in KG nor in PE. These results contribute to the study of gender differences in relation to success in solving floor-robot tasks. While some studies point to a higher performance of boys in certain robotics programming tasks (e.g., Sullivan & Bers, 2016), there are dissenting results where no initial differences between genders are revealed (e.g., Atmatzidou et al., 2018), to which the findings of the present study add.

The results demonstrate a greater ability of PE students to initially identify the path that requires a smaller number of turns while devising a plan. When it comes to gender differences, irrespective of their educational background, girls tend to exhibit a more prudent initial choice, often opting for the straightforward strategy represented by a simple path. However, this initial advantage in decision making does not necessarily translate into a higher success rate than that of boys when it comes to tasks of modest complexity (S1). Nevertheless, it does manifest its significance in situations demanding a greater degree of sophistication (S2), where the ability to make the right choice becomes notably more pivotal. Hence the importance of planning and evaluation. These differences can be linked to research findings on gender differences on risk aversion and on overvaluation capacities. Moè (2009) demonstrated that women often exhibit a higher degree of risk aversion and tend to have less confidence in their decision-making abilities compared to men. This latter possibility was studied by Lubinski et al. (2021), whose results showed an overconfidence of boys when evaluating the difficulty of a task.

In those cases where the revision of the plan is necessary, PE students show a superior ability to modify their initial plan and switch to the optimal flower. This means that these participants achieve a very high success rate. In KG, the results do not evidence this ability to replan the initial plan. Regarding the age factor, boys and girls in KG, in both tasks (S1 and S2) venture more to choose the complex path on their first attempt without achieving success, showing a lower planning capacity. This does not happen with boys

and girls in PE. Again, the reason may go hand in hand with the greater cognitive and metacognitive control capacity of PE students.

These findings have important implications for educational practice and policy. First, this work reveals the feasibility and potential of floor-robot programming contexts—widely popular in early ages—for developing, beyond coding skills, problem-solving abilities. In this regard, this study addresses the need highlighted by various authors to tackle the development of metacognitive skills from an early age (e.g., Temur et al., 2019). In addition, the study provides valuable evidence into the influence of age and gender on success in problem-solving programming tasks and decision-making processes. For PE students, having the possibility to select between different lines of resolution provided them with a wider margin to achieve success. Conversely, for KG students, the mere necessity of making a choice introduced an added layer of complexity to the task. On the other hand, the fact that girls tend to choose the least complex route suggests greater metacognitive control. However, this fact, which might seem positive, does not necessarily have to be so, as the enhancement of metacognitive control could be behind a tendency to avoid taking risks. In fact, this tendency to take risks has been considered a determining factor in the gender gap when students confront challenging mathematical problems according to Lubienski et al. (2021). Further research is needed on the differences between boys and girls at early ages in the aforementioned aspects—risk aversion and overestimation of skills—in this type of problems, as this is something that the data from this study do not allow us to clarify.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Funding was provided by University of Castilla-La Mancha and the European Regional Development Funds (2022-GRIN-34039), MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR (TED2021-131557B-I00, TED2021-129485B-C42), Generalitat Valenciana (AICO/2021/019), Ministerio de Universidades (FPU19/03857), Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana (GV/2019/146).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Atmatzidou, S., Demetriadis, S., & Nika, P. (2018). How does the degree of guidance support students' metacognitive and problem-solving skills in educational robotics? *Journal of Science Education and Technology*, 27(1), 70–85. <https://doi.org/10.1007/s10956-017-9709-x>
- Bers, M. U. (2020). *Coding as a playground*. Routledge. <https://doi.org/10.4324/9781003022602>
- Chen, S., & McDunn, B. A. (2022). Metacognition: History, measurements, and the role in early childhood development and education. *Learning and Motivation*, 78, 101786. <https://doi.org/10.1016/j.lmot.2022.101786>
- Clements, D. H. (1987). Longitudinal study of the effects of Logo programming on cognitive abilities and achievement. *Journal of Educational Computing Research*, 3(1), 73–94. <https://doi.org/10.2190/RCNV-2HYF-60CM-K7K7>
- Clements, D. H., Battista, M. T., & Sarama, J. (2001). Logo and Geometry. *Journal for Research in Mathematics Education Monograph*. <https://doi.org/10.2307/749924>
- Cohen, L., Manion, L., & Morrison, K. R. B. (2018). *Research Methods in Education* (8th ed.). Routledge. <https://lcn.loc.gov/2017015256>
- Desoete, A., & De Craene, B. (2019). Metacognition and mathematics education: An overview. *ZDM*, 51(4), 565–575. <https://doi.org/10.1007/s11858-019-01060-w>
- Diago, P. D., González-Calero, J. A., & Arnau, D. (2019). Fundamentos de diseño de un entorno tecnológico para el estudio de las habilidades en resolución de problemas en primeras edades escolares. *Research in Education and Learning Innovation Archives*, 22, 60. <https://doi.org/10.7203/realia.22.14113>
- Donaldson, M., & Balfour, G. (1968). Less is more: A study of language comprehension in children. *British Journal of Psychology*, 59(4), 461–471. <https://doi.org/10.1111/j.2044-8295.1968.tb01163.x>
- 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., Bloom, M., Grant, R., & Solomon, C. (1970). Programming-languages as a conceptual framework for teaching mathematics. *ACM SIGCUE Outlook*, 4(2), 13–17. <https://doi.org/10.1145/965754.965757>
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231–235). Lawrence Erlbaum. <https://cir.nii.ac.jp/crid/1570291225421995776>
- Garofalo, J., & Lester, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education*, 16(3), 163–176.
- Highfield, K., Mulligan, J., & Hedberg, J. (2008). Early mathematics learning through exploration with programmable toys. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepulveda (Eds.), *Proc. 32th Conf. of the Int. Group for the Psychology of Mathematics Education* (pp. 169–176). PME.
- Horn, M. S., Crouser, R. J., & Bers, M. U. (2012). Tangible interaction and learning: The case for a hybrid approach. *Personal and Ubiquitous Computing*, 16(4), 379–389. <https://doi.org/10.1007/s00779-011-0404-2>
- Kuzle, A. (2018). Assessing metacognition of grade 2 and grade 4 students using an adaptation of multi-method interview approach during mathematics problem-solving. *Mathematics Education Research Journal*, 30(2), 185–207.
- Lingel, K., Lenhart, J., & Schneider, W. (2019). Metacognition in mathematics: Do different metacognitive monitoring measures

- make a difference? *ZDM*, 51(4), 587–600. <https://doi.org/10.1007/s11858-019-01062-8>
- Lubienski, S. T., Ganley, C. M., Makowski, M. B., Miller, E. K., & Timmer, J. D. (2021). “Bold problem solving”: A new construct for understanding gender differences in mathematics. *Journal for Research in Mathematics Education*, 52(1), 12–61. <https://doi.org/10.5951/jresmetheduc-2020-0136>
- Lucangeli, D., Fastame, M. C., Pedron, M., Porru, A., Duca, V., Hitchcott, P. K., & Penna, M. P. (2019). Metacognition and errors: The impact of self-regulatory trainings in children with specific learning disabilities. *ZDM*, 51, 577–585.
- Mahdavi, M. (2014). An overview: Metacognition in education. *International Journal of Multidisciplinary and Current Research*, 2, 529–535.
- Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, 19(1), 21–27. <https://doi.org/10.1016/j.lindif.2008.02.002>
- Mok, Y. F., Fan, R. M., & Pang, N. S. (2007). Developmental patterns of school students’ motivational- and cognitive-metacognitive competencies. *Educational Studies*, 33(1), 81–98. <https://doi.org/10.1080/03055690600948281>
- Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*. Basic Books Inc.
- Pólya, G. (1957). *How to solve it* (2nd ed.). Princeton University Press.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). MacMillan.
- Stillman, G. (2020). Metacognition. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 608–610). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_166
- Sullivan, A., & Bers, M. U. (2016). Girls, boys, and bots: Gender differences in young children’s performance on robotics and programming tasks. *Journal of Information Technology Education: Innovations in Practice*, 15, 145–165.
- Temur, Ö. D., Özsoy, G., & Turgut, S. (2019). Metacognitive instructional behaviours of preschool teachers in mathematical activities. *ZDM*, 51(4), 655–666. <https://doi.org/10.1007/s11858-019-01069-1>
- Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: Students’ self-reports versus on-line assessment methods. *ZDM*, 51(4), 691–701. <https://doi.org/10.1007/s11858-018-1006-5>
- Whitebread, D., & Coltman, P. (2010). Aspects of pedagogy supporting metacognition and self-regulation in mathematical learning of young children: Evidence from an observational study. *ZDM*, 42(2), 163–178. <https://doi.org/10.1007/s11858-009-0233-1>
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82(1), 51–59. <https://doi.org/10.1037/0022-0663.82.1.51>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.