



# A Bebras Computational Thinking (ABC-Thinking) program for primary school: Evaluation using the competent computational thinking test

María Zapata-Cáceres<sup>1</sup> · Pedro Marcelino<sup>2</sup> · Laila El-Hamamsy<sup>3,4</sup> · Estefanía Martín-Barroso<sup>1</sup>

Received: 5 June 2023 / Accepted: 19 December 2023  
© The Author(s) 2024

## Abstract

Bebras tasks are considered to develop Computational Thinking (CT) and are currently used for this purpose in many studies. However, the relationship between Bebras tasks and CT is recent and, given the scarcity of validated instruments for assessing CT that are not associated with a specific learning environment, the expected development of CT through Bebras tasks has not been sufficiently evaluated, especially in Primary School. For this reason, we designed the ABC-Thinking program for the development of CT in which a set of Bebras tasks were selected and categorised according to the CT skills they were related to, specific lesson plans were designed to integrate these tasks in a 12-week curriculum, and teachers were trained to apply the program. Using the competent Computational Thinking test (cCTt) in a pre-post-test experimental design, we looked to determine whether Primary School students developed specific CT skills after the program, and, therefore, whether Bebras could be considered suitable for the development of this competence. The results show a significant development in specific CT skills, interesting findings in the gender differences in this development, and between high and low performers. Finally, qualitative data indicate the suitability of the ABC-Thinking program with respect to teachers' practice.

**Keywords** Computational thinking · Primary education · Bebras tasks · Teacher professional development · Computational thinking test (cCTt) · Educational program

## 1 Introduction

Computational Thinking (CT) is widely recognized as a fundamental cognitive skill essential for adapting to our technologically driven society and it is increasingly integrated into early education curricula worldwide. However, despite the

---

Extended author information available on the last page of the article

various learning strategies available to develop CT from an early age, there is no clear evidence regarding which strategies are most suitable for this purpose (Hsu et al., 2018). Unplugged learning is one of the most commonly employed strategies when it comes to fostering CT in young children (Caeli & Yadav, 2020).

Bebras tasks ([bebras.org](http://bebras.org)) consist of sets of small problems or challenges, which can be regarded as unplugged activities since they do not require specific technological devices to be used in schools. These tasks were originally designed to teach programming and stimulate students' interest in computer science from early stages. Recently, it has been demonstrated that they can also enhance CT (Dagiene & Stupuriene, 2016). However, it is crucial to establish a proper connection between Bebras tasks and CT. The aim is to select a set of Bebras activities that is balanced and can include all the skills into which CT is decomposed. Various categorizations for this correspondence exist in the literature, although further research in this area is necessary. In this study, we will employ Bebras tasks, and the two-dimensional categorization (Dagienė et al., 2017), in a comprehensive program specifically designed for this research, referred to as the ABC-Thinking program, to foster students' CT skills.

Given the recent implementation of CT education in schools internationally, it is essential to provide educators with comprehensive programs that encompass training, activities, and assessment. In this research, the ABC-Thinking program has been developed as a comprehensive CT development program. It includes pre-training sessions for teachers, a 12-week program of activities with students employing a balanced CT Bebras tasks set along with games and gamification elements, and an assessment of students' CT development. The program has been implemented in the primary education stage, involving students aged between 8 and 10 years.

Throughout the program implementation, the entire process has been monitored, and both qualitative and quantitative data have been collected to assess whether the program is effective. Students' improvement in different categories has been measured using the validated competent Computational Thinking test (cCTt) (El-Hamamsy et al., 2022), through a pre-post-test experimental design. These data provide insights into the aspects in which students improve their CT skills based on the chosen categorization. Moreover, qualitative data have been gathered from the teachers responsible for the program implementation to determine the program's suitability for their daily practice based on their perspectives.

The Bebras tasks could also be used to assess students' knowledge and skills in Computer Science. However, they have not been specifically validated as a CT assessment instrument, although they have been used in some studies for this purpose. Through the ABC-Thinking program, a first effort has been made to introduce the Bebras tasks as an assessment instrument, in addition to being a tool for learning CT. The categorization that has been carried out for the ABC-Thinking program to relate the different Bebras tasks to the different CT skills, as well as the confirmation that CT can be developed through these tasks, could be a first step towards validating these tasks as an assessment tool in future research.

## 2 Related work

### 2.1 Fostering computational thinking skills

Computational Thinking (CT) is considered to be an essential cognitive skill to adapt to today's technological society which is increasingly being included in school curricula from an early age and at an international level (Shute et al., 2017). While there is still no universal definition of CT, there are various decompositions of CT such as the one suggested by Selby and Woollard (Selby & Woollard, 2013), which includes the following skills: abstraction, algorithmic thinking, decomposition, evaluation, and generalisation. On the other hand, one of the most cited and empirically used CT operational frameworks in the literature is the 3D framework (Brennan & Resnick, 2012), which classifies CT according to three dimensions: (1) computational concepts (concepts that programmers use); (2) computational practices (problem-solving practices that are needed/produced in the programming process); and (3) computational perspectives (perspectives that designers form about themselves and the world around them).

There are various learning strategies to develop Computational Thinking, particularly at early ages, however, there is no clear evidence as to which strategies are most appropriate for this development. The most recent research highlights the following strategies, both for their positive effect on learning and for being the most widely used for CT development: problem-based learning, collaborative learning, project-based learning, and game-based learning (Hsu et al., 2018).

In terms of the environments used for CT development, digital games and applications stand out, although it has been shown that, in the case of early childhood and primary education, unplugged activities, i.e., those that do not require electronic devices to be carried out, enhance this development significantly (Brackmann et al., 2017), such as using tangible robots (Zapata-Ros, 2019), comics (Suh et al., 2020), character creation through algorithms (Waite et al., 2019), or sequential narratives and stories (Acha et al., 2021; Kazakoff et al., 2013). These unplugged activities provide students with the opportunity to explore the fundamental ideas of Computer Science and CT without needing the technical knowledge required for programming (Bell et al., 2009). In this regard, Bebras tasks can be seen as a CT development activity that does not require specific material resources and is familiar to teachers, making it more likely that they will accept the activity in their long-term practice and therefore can be implemented across a broad demographic group.

### 2.2 Computational thinking development through Bebras tasks

Since its origin, the aim of the Bebras International Challenge has been to promote interest and learning of Computer Science and CT for students worldwide. The Bebras challenges are composed of a set of short problems called Bebras tasks which rely in particular unplugged activities (Cartelli et al., 2010; Dagiene et al., 2014; Dagiene & Futschek, 2008), in such a way that students are confronted with

real and meaningful problems and seek the solution using CT. These tasks are independent of a particular hardware, software, or environment, and do not require prior programming experience. Thus, each Bebras task includes at least one computational concept, engages children's attention through story, image, or interactivity, is short (fits on one page) and requires no specific technical knowledge (see example in Fig. 1).

There is evidence that Bebras tasks can be used in school curricula as effective tools to promote problem-solving skills and CT in children and young people (Dagiene & Stupuriene, 2016). However, since the introduction of CT in the school curricula is recent, more information, analyses, and solid evidence on the development of this competence through Bebras tasks in primary education is needed (Dagiene & Stupuriene, 2016; Dagiene & Dolgopolas, 2022).

It is possible to select a set of Bebras tasks to configure a program suitable for different school contexts. However, in order to select a balanced set of tasks in terms of CT, it is necessary to know what specific CT skills are being developed, for which it is therefore required to establish a correspondence between the categories of Bebras tasks used and the CT that each one of them addresses. In this regard, there are several categorizations of Bebras tasks, such as that of Dagiene & Futschek, 2008, or that of Kalas & Tomcsanyiova, 2009. These however have significant limitations, for example, in terms of overlap and imbalance between

### ***Cleaning the Lawn***

After a concert, the cleaning robot picks up the trash that some of the audience has left on the lawn:



The cleaning robot moves to the closest piece of trash and picks it up. It then moves to the next closest piece of trash and picks it up. The robot continuously does this till all the pieces of trash are picked up.

### **Question / Challenge**

**Which piece of trash will the robot pick up *last*?**

Fig. 1 Example of Bebras task used: "Cleaning the lawn"

categories, in their too general nature, and, in particular, in their correspondence with CT (Dagienė et al., 2017).

Based on the previous categorizations, in 2017, the Bebras community introduced five categories of tasks, oriented to the development of programs for learning computer science (Dagiene & Stupuriene, 2016): ALP: Algorithms and Programming, DSR: Data, Data Structures and Representations, CPH: Computer Processes and Hardware, COM: Communications and Networks, ISS: Interactions, Systems and Society. Finally, to complement this last classification, (Dagienė et al., 2017) propose a two-dimensional categorization by adding as a second dimension the CT skills suggested by (Selby & Woollard, 2013): abstraction, algorithmic thinking, decomposition, evaluation, and generalisation. In this research, the two-dimensional categorization has been used (Dagienė et al., 2017), as it provides an adequate correspondence between the skills associated with CT and the current categorization of Bebras tasks.

### 2.3 Adequately assessing the impact of Bebras tasks on CT skills

Besides the annual Bebras International Contest ([bebras.org](http://bebras.org)), there is an increasing use of Bebras tasks in formal education contexts. For example, (Lehtimäki et al., 2022) created CT learning resources based on Bebras tasks. These were distributed to over 14,000 students and 250 teachers in Ireland but unfortunately did not undergo extensive validation beyond acquiring teachers' perception of the content. Thus, assessing, both quantitatively and qualitatively, students' development of CT through Bebras tasks is indispensable for the introduction of this competence in school curricula.

Bebras tasks have been used as validated computer science assessment instruments (Bellettini et al., 2015; Hubwieser & Mühling, 2015; Kalelioğlu et al., 2015). Moreover, Bebras tasks have also been used as a CT assessment tool, e.g., (Lockwood & Mooney, 2018), assessed the development of CT in secondary school students (15-17 years old) using Bebras tasks, and Jun and Soojin et al. in 2018 (Jun et al., 2018), used Bebras tasks as an assessment instrument in primary school students. In all cases, the validity of Bebras tasks as a CT assessment instrument needs to be confirmed by comparing the results with those of a reliable and validated CT assessment instrument (Dagiene & Stupuriene, 2016). However, there is a lack of psychometric analysis of the Bebras tasks as instruments for the assessment of CT (Román-González et al., 2017). To this regard, it is essential to analyse the effectiveness of Bebras tasks, as a CT-learning strategy which also provides a first step towards establishing the validity of Bebras Tasks as CT assessment tools.

As stated previously, there is neither a consensus on which strategies are the most appropriate for assessing CT at an early age, nor an agreement on what should be assessed. This makes it difficult to measure the effectiveness of CT learning interventions in a reliable and valid way (Grover & Pea, 2013; Shute & Moore, 2017). In this regard, in recent years there has been an international effort to develop a series of unplugged CT assessment instruments (El-Hamamsy et al., 2022; Román-González et al., 2018; Román-González et al., 2019; Tsarava et al., 2022; Zapata-Cáceres et al.,

2020) that cover a wide range of target ages, from early childhood to university, and that can be used independently of any learning environment. Of these instruments, the competent Computational Thinking test (cCTt) (El-Hamamsy et al., 2022) is aimed at primary school students, and has a proven reliability for grade 3-4 students. The cCTt has 25 questions which target the CT-concepts defined by (Brennan & Resnick, 2012): sequences, loops, if-else statements, and while statements. The instrument underwent psychometric validation with data from 1519 students and demonstrated adequate reliability, difficulty and discriminability for the target age groups. One of the interesting aspects of the cCTt is that it relies entirely on graphs and symbols, so that no literacy skills are required to solve the questions. On the other hand, Bebras tasks require considerable reading comprehension from students in order to understand the challenge posed to them, which could interfere with or mask CT skills, especially at early stages, in this case in Primary Education. Thus, the confirmation, through the cCTt, of the Bebras tasks as a CT development tool, is even more pertinent.

A limitation of traditional test-based assessments, such as the cCTt, is that they do not cover all CT Brennan and Resnick's framework dimensions. According to Grover and Pea (Grover & Pea, 2015), a "system of assessments" is needed to assess deeper learning by combining different data measures (e.g., traditional tests, student surveys, and teachers' perceptions). In this regard, to include qualitative data for a more comprehensive evaluation, the teachers monitored the implementation of the tasks by filling in forms while directly observing the experience. In addition, questionnaires and interviews with teachers and coordinators about the experience were also carried out. This data collection could also provide insights into teachers' perceptions of a Bebras CT program (ABC-Thinking program), as one of the major problems identified in previous research is that teachers do not know exactly how to teach or assess CT in a motivating way (Zapata-Cáceres & Martín-Barroso, 2021). Moreover, we explore the possibility of combining CT assessment through the cCTt, together with the Bebras tasks, and qualitative data.

### 3 Research questions

Considering the research that has already been carried out, we are interested in the following main research question: Does the Bebras tasks-based CT development program contribute to student CT-learning in Primary Education? In order to answer this main research question, the following sub-questions have been formulated: (RQ1) What specific CT skills can be developed and therefore assessed through the Bebras tasks? and (RQ2) What are teachers' perspectives on the CT development program suitability with respect to their practice?

### 4 Methodology

A 12-week CT development program based on Bebras tasks (ABC-Thinking program) was proposed to teachers to implement in their classrooms (see Section 4.1). To evaluate the impact of the program and understand how it can foster

specific CT skills (RQ1) and how the teachers’ perceived the program (RQ2), 131 students, from 3rd and 4th grades (Primary School), from 4 schools in Portugal and 15 teachers participated in the study (see Section 4.2). More specifically, we employed a “system of assessments” which combined quantitative data from students through a pre-post-test experimental design using the cCTt (see Section 4.2.1) and qualitative data from teachers (see Section 4.2.2) to gain more comprehensive insight into the adequacy of the program with respect to their practice.

The methodology section is divided into the following two sub-sections (see Fig. 2):

- Activities design: where it is specified which Bebras activities will be implemented in the classroom during the 12 weeks of the program. This section details how these activities have been selected, how a lesson plan or protocol has been established so that teachers can properly implement the activities, and how the activities have been gamified.
- Participants and data collection: the profile of the students and teachers who participated in the research is specified, as well as the teachers training and the instruments used for data collection, both in the case of students, through a validated CT test; and teachers, through forms and interviews.

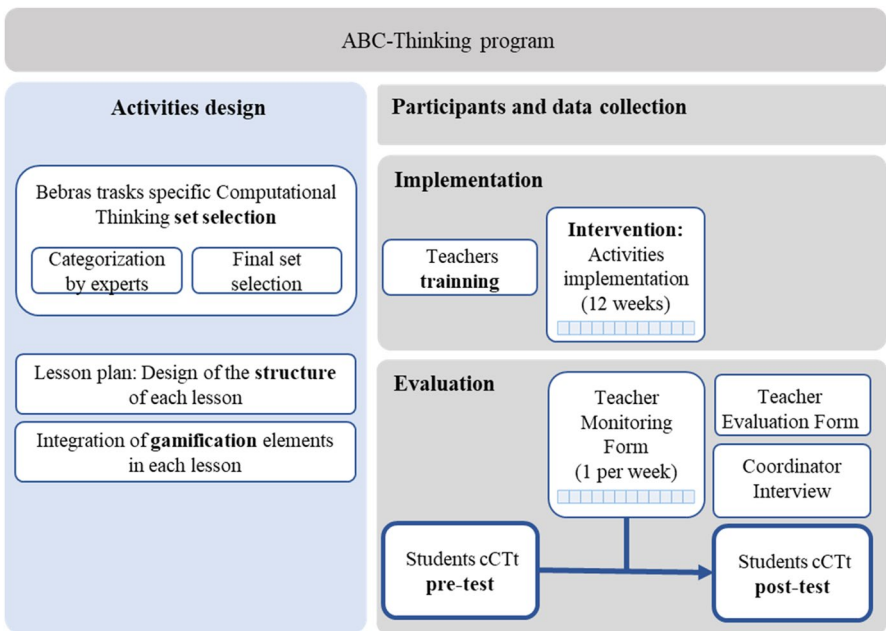


Fig. 2 ABC-Thinking program overview



## 4.1 Activities design

The ABC-Thinking program was developed for the purpose of the study. The program was provided in online (Open edX platform) and paper formats, although all participating schools opted for the online format. It consisted of 12 lessons of one hour each, which lasted 12 weeks (one lesson per week), in a format similar to that used in other studies (Chevalier et al., 2022) and suitable for curricular implementation by the participant teachers.

Each lesson was composed by: (1) A set of Bebras exercises (see Section 4.1.1) selected to cover and develop different CT skills (Dagiene & Stupuriene, 2016); (2) A lesson plan (see Section 4.1.2) to guide teachers on how the lesson should be taught, enabling them to function more effectively on a subject with which many of them were unfamiliar; and (3) Gamification elements for each activity (see Section 4.1.3) to provide alternative means of orchestrating the classroom activities in a fun and engaging way, allowing students to develop their curricular, cognitive and social competences (Manzano-León et al., 2021).

### 4.1.1 Bebras tasks CT set

The ABC-Thinking Development Program comprised 12 sets of Bebras exercises, with each set encompassing different Bebras tasks. The selection of tasks for the respective sets was carefully conducted by three expert researchers in Computational Thinking (CT), taking into consideration the specific CT skills addressed by each task. In this way, the CT skills targeted by each Bebras task was established using (Dagiene et al., 2017) criteria for classification of CT tasks. Thus, each Bebras task was classified by the researchers, independently, by attributing one or more of the 5 CT skills to the tasks using Dagiene et al.'s (2017) classification guide: (CT1) Abstraction, (CT2) Algorithmic thinking, (CT3) Decomposition, (CT4) Evaluation, and (CT5) Generalisation.

In cases where 100% agreement was not reached, the researchers met to discuss the classification of each of the tasks until agreement was reached. Coherence of the classification between similar tasks was also verified during the procedure. The resulting classification can be seen in Table 1, which describes the Bebras exercise sets used classified by CT skills. Moreover, Table 5 (Appendix 1) shows the specific Bebras tasks used in each set, together with their individual classification.

After classification, both specific and mixed sets were created. In specific sets, all exercises intended to focus on the same specific CT skill. The skills categorization proposed by (Selby & Woollard, 2013) was considered, so the specific exercises sets could focus on one of these skills: abstraction, algorithmic thinking, decomposition, evaluation, and generalisation. On the other hand, mixed sets were composed of exercises that worked on multiple CT skills. This brings an interleaved practice component into the learning process, whose beneficial effects have already been proven in other disciplines (Taylor & Rohrer, 2010).

Moreover, the exercise sets could be interactive or non-interactive. Interactive sets were composed of exercises where the students would actively manipulate physical objects. For example, the Bebras task “Push-Away Parking” can and was reproduced



**Table 1** Bebras exercise sets included in the ABC-Thinking program, classified by CT skills. Abbreviations - CT1: Abstraction, CT2: Algorithmic thinking, CT3: Decomposition, CT4: Evaluation, CT5: Generalisation

Set	Topic	Type of exercise sets	Interactive	Number of activities	Number of tasks covering each CT skill				
					CT1	CT2	CT3	CT4	CT5
1	General introduction	Mixed	No	3	1	2	1	0	1
2	Queues	Specific	No	4	0	4	0	2	1
3	Patterns	Specific	No	4	1	3	1	4	3
4	Algorithms	Specific	Yes	1	0	1	0	0	0
5	Checkpoint	Mixed	No	6	2	3	0	6	2
6	Instructions	Specific	No	3	0	3	0	0	0
7	Binary code	Specific	No	4	0	3	0	3	2
8	Checkpoint	Mixed	No	8	1	7	2	5	3
9	Algorithms	Specific	Yes	1	0	0	0	1	0
10	Conditions	Specific	No	4	1	4	0	3	0
11	Finite states	Specific	No	4	0	4	0	4	2
12	Final checkpoint	Mixed	No	8	1	5	0	6	3
Total			2	50	7	29	4	32	16

in plasticine (i.e., playdough) and solved through direct experience interacting with objects. Figure 3 shows an example of an exercise that can be reproduced physically and, consequently, can be used in an interactive way. Non-interactive sets were composed of exercises that do not require object manipulation by the students (i.e., students would just answer to the question without reproducing it physically). Figure 2 shows an example of a non-interactive exercise.

#### 4.1.2 Lesson plan for teachers

As mentioned earlier, the ABC-Thinking program consists of 12 lessons, one lesson per week. Each lesson was structured considering three distinct moments: introduction, exercise, and debriefing. The introduction intended to gain the students' attention, to provide the necessary background information and to establish the class direction. It lasted 15 minutes and most activities were modelling activities where teachers would solve an example exercise while sharing their train of thought. Sometimes, retrieval practice activities and short CT activities were also used in the introduction.

After the introduction, the students moved to the exercises, and solve the Bebras exercises sets using gamification elements. The exercises aimed to develop the students' CT skills and lasted 30 minutes.

The final moment of the lesson was a debriefing review moment that lasted 15 minutes and fostered a reflective practice. Most debriefing activities were

## Push-Away Parking

In a parking lot, cars can be parked in parking spaces or in front of these parking spaces as shown in the image below. Cars that are parked in front of the parking spaces can be pushed forward or backward carefully if they block a car that wants to leave its parking space.

For example, in the image below, car A is not blocked and can leave its parking space. Car L is blocked by car M. Car M must be pushed away before car L can leave its parking space.



### Question / Challenge

**Which car needs two other cars to be pushed away before it can leave its parking space?**

Fig. 3 Example of an exercise that can be reproduced physically

about class discussion of a specific exercise and, when the lesson was more challenging, modelling activities were used to review the exercises.

### 4.1.3 Gamification elements

In the educational domain, gamification refers to the use of game elements to increase the students' motivation and engagement in learning activities (Dichev & Dicheva, 2017). Gamification is of interest because several studies suggest that it can improve students' school motivation and academic performance (Hamari et al., 2014; Manzano-León et al., 2021; Sailer & Homner, 2019).

To reap the benefits of gamification and to help teachers motivate and engage students in learning, the lessons from the ABC-Thinking program integrated gamification elements. In this sense, gamification elements such as points, missions, feedback, levels, and time manipulation - as defined by (Schöbel et al., 2020) were used to build a set of different games. These games followed a collaborative-based design, where the students worked together to achieve a common goal, which previous studies have found to be more effective than competitive-based designs (Sailer & Homner, 2019). Table 6 (Appendix 2) details the games created following the principles described above.

## 4.2 Participants and data collection

### 4.2.1 Students

Several schools and teachers that were enrolled in the national Bebras competition were contacted and an explanation was provided about the ABC-Thinking program and what it would take to participate. As a result, a total of 131 students, 72 girls and 59 boys, participated in the ABC-Thinking program. The students belonged to 4 different primary schools in Portugal and attended the 3rd and 4th grades (ages between 8 and 10). Table 2 shows the distribution of the students' sample by gender, grade, and age.

The activities corresponding to the ABC-Thinking program were included in the curricular practice by the participating teachers. Accordingly, all students from the classes involved participated in the program and all the responses/data were collected during the class, under the teachers' supervision. None of the students had prior formal experience in CT or programming before the intervention. To carry out the study, authorization was requested from the parents of the participants, with the collaboration of the schools and their teachers. All data collected and processed were anonymized.

To assess the impact of the intervention on students, a pre-post-test experimental design was employed (see Fig. 2), using the competent Computational Thinking test (cCTt) (El-Hamamsy et al., 2022). The cCTt is a validated CT assessment for grades 3 and 4. Its design derives from the Beginners' CT test (BCTt) (Zapata-Cáceres et al., 2020) - a test developed for grades 1 to 6 - with format and content changes in order to target, more specifically, students in grades 3 and 4.

The cCTt is composed of 25-item multiple choice questions of progressive difficulty. These questions focus on the CT concepts suggested by Brennan & Resnick, 2012 (sequences, loops, conditionals, and while statements). The test also partially addresses Brennan & Resnick's (2012) CT practices (being incremental and iterative; testing and debugging; reusing and remixing; and abstracting and modularizing) that occur in the process of solving the test questions. These aspects encompass the CT skills proposed by Dagiéné et al. (2017) except for those related to CT perspectives (expressing, connecting, and questioning) (Román-González et al., 2017), which can be assessed primarily through student surveys, or in the present case

**Table 2** Sample distribution by gender, school year and age ( $N = 131$ )

			Gender		
			Boys	Girls	Total
Grade	3 (8–9-year-olds)	n	21	21	42
		% of total	50%	50%	32%
	4 (9–10-year-olds)	n	38	51	89
		% of total	43%	57%	68%
Total		n	59	72	131
		% of total	45%	55%	100%

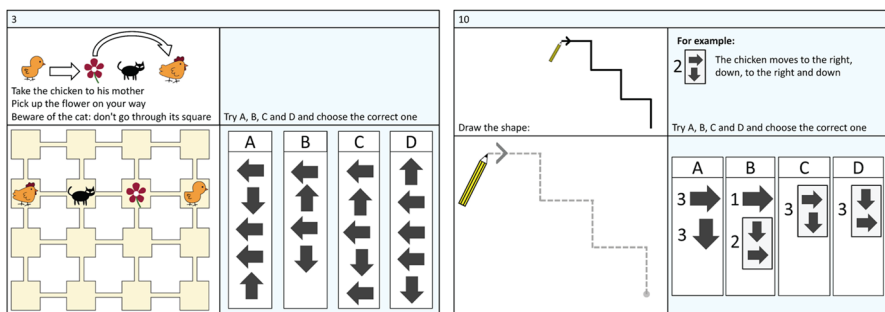
through qualitative teacher-data pertaining to their observations during the ABC-Thinking program. Both grid and canvas questions are used in the cCTt, according to the distribution shown in Table 3. Figure 4 presents two examples of the cCTt questions.

According to Item Response Theory, the cCTt appears to discriminate best for low-medium ability students: “In the final stage, the Item Response Theory analysis supported these findings and further indicated that the test was better suited at evaluating and discriminating between students with low and medium abilities” (El-Hamamsy et al., 2022). Moreover, the cCTt has an overall reliability of 0.85 given by Cronbach’s alpha, and a Confirmatory Factor Analysis that provides evidence of the construct validity of the instrument, when it comes to measuring the targeted CT-concepts (El-Hamamsy et al., 2022). As such, the cCTt was used in a pre-test post-test design in the present study. The pre-test was applied before the intervention and allowed the measurement of the students’ initial CT-competence. Once the intervention was completed, the post-test was applied to measure the students’ final CT-competence. Since all participating schools opted for the online format, an online version of the cCTt was used for both tests. This version was integrated in the same platform used to provide the ABC-Thinking program. The impact of the intervention and the results are described in Section 5.1.

Results from the pre-post-test were analysed using one-way analyses of variance with the dependent variables being i) the raw score obtained on the cCTt, ii) the

**Table 3** cCTt question concepts and types of distribution (El-Hamamsy et al., 2022)

Concept	Grid (3×3)	Grid (4×4)	Canvas	Total
Sequences	1	1	2	4
Simple loops	0	4	0	4
Complex loops	0	5	2	7
Conditional statements	1	3	0	4
While statements	1	3	0	4
Combinations	0	2	0	2
Total	3	18	4	25



**Fig. 4** Examples of cCTt questions. On the left, an example of a grid question. On the right, an example of a canvas question (El-Hamamsy et al., 2022)

normalised change (Coletta & Steinert, 2020), a symmetric version of the learning gain which is computed as follows:

$$NC = (posttest - pretest)/(100\% - pretest) \text{ if } posttest \geq pretest \text{ else } (posttest - pretest)/(pretest)$$

The independent variables considered in the analysis were i) the students' gender, ii) grade, iii) when the test was administered (pre or post in the case of the raw score), iv) the students relative ranking according to the results of the pre-test. More specifically, students were divided into 3 groups of similar size: initially low performers, medium performers and high performers using a ranking method and attributing  $61/3$  in each group. To account for the multiple tests and avoid the occurrence of false positives (Type I errors), a  $p$  value correction was applied using the Benjamini-Hochberg approach which controls for false positive rates (Thissen et al., 2002). To ensure that we have sufficient statistical power (0.8), we also consider the minimum effect size that can be reported (Cohen's  $D$ ) with respect to the sample size and number of groups. As such, we report the following in the case of significant results which meet the minimum effect size requirements: i) the test statistic, ii) the corrected  $p$  value, and iii) the effect size.

#### 4.2.2 Teachers

As mentioned before, the schools and teachers that were enrolled in the national Bebras competition were contacted to participate in this study. In this contact, schools interested in participating were asked to indicate a set of teachers interested in running the program, as well as a teacher who would coordinate the program at the school and serve as a point of contact. This way, a total of 15 teachers participated in the study, 11 who ran the CT program described in section 2.1 in their classrooms and 4 who coordinated the CT program in their respective school.

Before the intervention (see Fig. 2), all participating teachers received a two-hour online training and a set of support documentation which covered 3 topics:

1. The CT development program - how the program was structured and how to apply it.
2. The Competent CT test (cCTt) - how to apply it and the pre-test-post-test design.
3. The ABC-Thinking Platform - how to use the platform where the ABC-Thinking program was provided.

During and after the intervention (see Fig. 2), qualitative data was collected to gain insight into the teachers' perception of the CT development program and its adequacy with respect to their practice, as well as observing student performance. Accordingly, the following instruments were used: (1) Teacher Monitoring form; (2) Teacher Evaluation form; and (3) Coordinator Interview.

The (1) Teacher Monitoring Form were online forms that the teachers running the program filled out every week, after the implementation of the program. In this way, it was intended to obtain feedback on each week of the program through the collection of information regarding what the teacher observed during the application of the program. The form was the same every week, the questions were of two types:

5-point Likert scale, and open questions. The basic structure of the form is shown in Table 4.

The (2) Teacher Evaluation Form was an online form that the teachers running the program filled out at the end of the program. This form assessed general aspects of the CT development program and it was more extensive and comprehensive than the monitoring forms. Briefly, the form included questions about. Table 4 shows the basic structure of the form.

Finally, for the (3) Coordinator Interview, the coordinating teachers were interviewed at the end of the program. The interview wanted to capture insights from these teachers, who were not applying the program directly, but who were in contact with the teachers running the program and thus had a broad perspective of how the program was going. A design thinking activity - Feedback Capture Grid (Lewrick et al., 2020) - was used to structure the interview. Table 4 shows the basic aspects that were covered during the interview. It was thus possible to investigate, from the coordinator's perspective, the efficiency of the program and student performance, according to the teachers' perception.

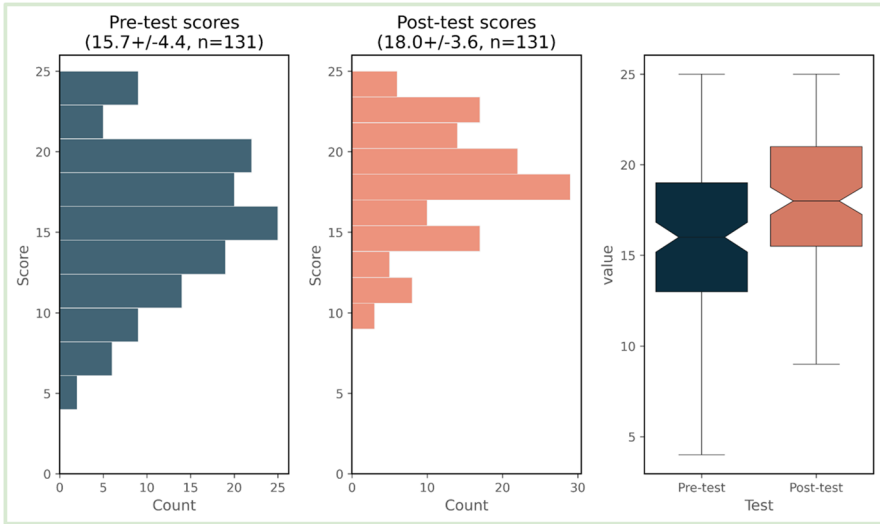
## 5 Results

### 5.1 Student learning as a result of the ABC-Thinking program

Students scored an average of  $15.7 \pm 4.4$  out of 25 on the pre-test (see Fig. 5). These scores were normally distributed (Shapiro Wilk test,  $W = 0.985$ ,  $p = 0.167$ ; skew =  $-0.0600$ ; kurtosis =  $-0.146$ ) around the mean. Following the intervention, the students scored an average of  $18.0 \pm 3.6$  on the post-test (see Fig. 5). The scores

**Table 4** Structure and basic aspects covered in Forms and Interview

Forms and Interview	Structure and basic aspects covered
Teacher Monitoring Form	How easy it was to carry out the week's activities How would you rate the students' level of engagement Mention one thing that went well Mention one thing to improve Other comments
Teacher Evaluation Form	Program duration Program activities Teachers' preparation for program application Student engagement level
Coordinator interview	Platform experience What went well What needs to be changed What new ideas have arisen from the program application What questions have arisen from the program application

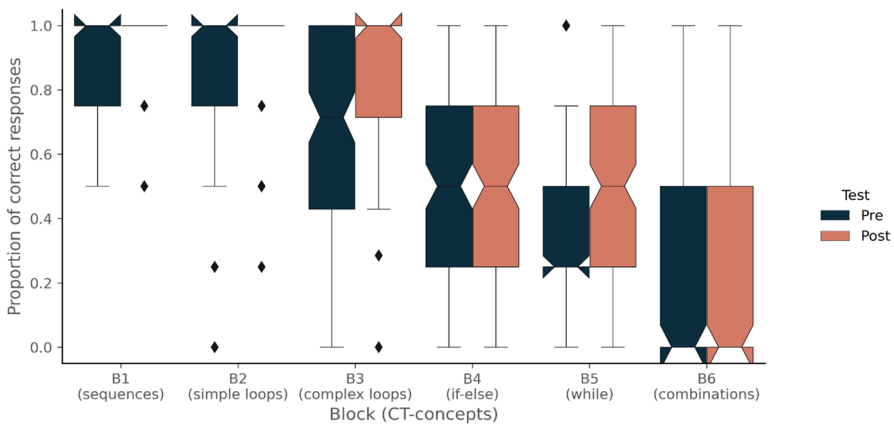


**Fig. 5** Distribution of the scores obtained for the 131 students that could be followed between the pre and post-tests

appear to be non-normally distributed and beginning to exhibit a ceiling effect (Shapiro Wilk test,  $W = 0.979, p = 0.0437$ ; skew =  $-0.225$ ; kurtosis =  $-0.567$ ).

The increase is significant (Kruskal-Willis test,  $H = 26.5, p = 3.3e-6$ ) with a medium to large effect size (Cohen’s  $D = 0.565$ ) and appears to indicate that the students improved their mastery of CT-concepts through the CT Development Program.

Looking at the distribution of scores per block (see Fig. 6), the pre-test scores appear to indicate that students had a good mastery of sequences (B1, 90% of correct responses) and simple loops (B2, 81% of correct responses). The results also



**Fig. 6** Distribution of the proportion of correct responses per block of questions in the cCT



indicated that could still progress on more advanced CT-concepts such as complex loops (B3, 68% of correct responses) conditionals (B4, 52% of correct responses), while statements (B5, 39% of correct responses) and the combination of statements (B6, 23% of correct responses) prior to the intervention. Not all sub-samples have a normal distribution (B1,  $D=6.7$ ,  $p=0.03$ ; B2,  $D=16.4$ ,  $p=0.0003$ ; B3  $D=13.3$ ,  $p=0.001$ ; B4,  $D=0.1$ ,  $p=0.9$ ; B5,  $D=2.2$ ,  $p=0.3$ ; B6,  $D=3.3$ ,  $p=0.2$ ), according to the Omnibus Test of Normality (d'Agostino, 1971), so it is necessary to use a non-parametric test. The results indicate that the intervention led to improvements, but this differed according to the CT-concepts (Kruskal-Willis test,  $H=18.9$ ,  $p=0.002$ ). Using non-parametric Dunn's test for multiple comparisons helps shed light on this. Following the intervention, the students had near perfect mastery of sequences (B1, 96% of correct responses, +6.1%,  $p=0.0717$ ,  $D=0.474$ ) and simple loops (B2, 93% of correct responses on average, +11.5%,  $p=0.0042$ ,  $D=0.551$ ). They also improved on their mastery of complex loops (B3, 83% of correct responses, +14.8%,  $p=0.0002$ ,  $D=0.568$ ) and of while statements, although marginally for the later (B5, 49 of correct responses, +9.4%,  $p=0.0511$ ,  $D=0.285$ ). Interestingly, they did not improve on conditional statements (B4, +1.5%,  $p=0.8528$ ,  $D=0.055$ ) or more advanced questions using the combination of concepts (B6, +3.3%,  $p=0.6194$ ,  $D=0.095$ ). This could be an indication that the students require more targeted interventions to continue to improve on these more advanced CT-concepts.

When dividing the sample according to students' gender, the distribution of the sub-samples is normal (pre-test girls,  $D=0.5$ ,  $p=0.8$ ; post-test girls,  $D=5.3$ ,  $p=0.07$ ; pre-test boys,  $D=0.09$ ,  $p=0.9$ ; post-test boys,  $D=3.2$ ,  $p=0.2$ ). No significant differences are observed between genders (see Fig. 7) in the pre and post test scores (one-way ANOVA between genders in the pre-test  $F(1)=0.013$ ,  $p=0.91$ ; in the post-test  $F(1)=0.75$ ,  $p=0.39$ ). However, there appears to be an interaction effect between the pre-post tests and genders (two-way ANOVA  $F(2)=12.27$ ,  $p=1.6e-5$ ). Employing Dunn's post-hoc test for multiple comparisons indicates that girls improved more significantly between the pre-test and

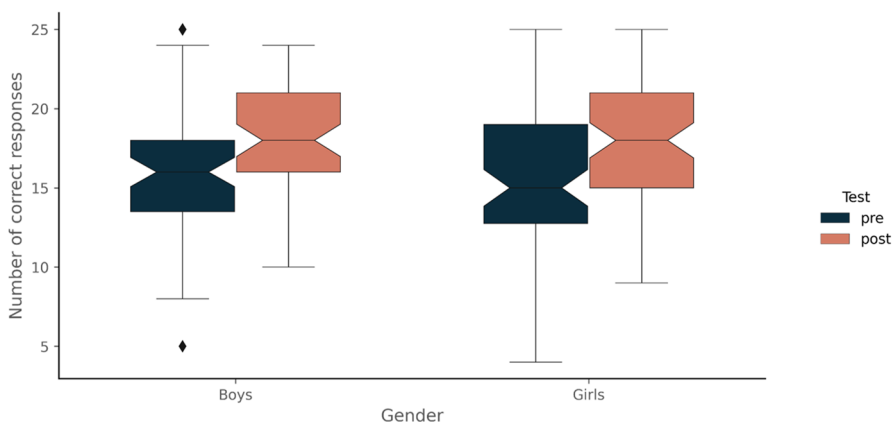


Fig. 7 Distribution of the scores obtained in the pre and post-tests according to gender

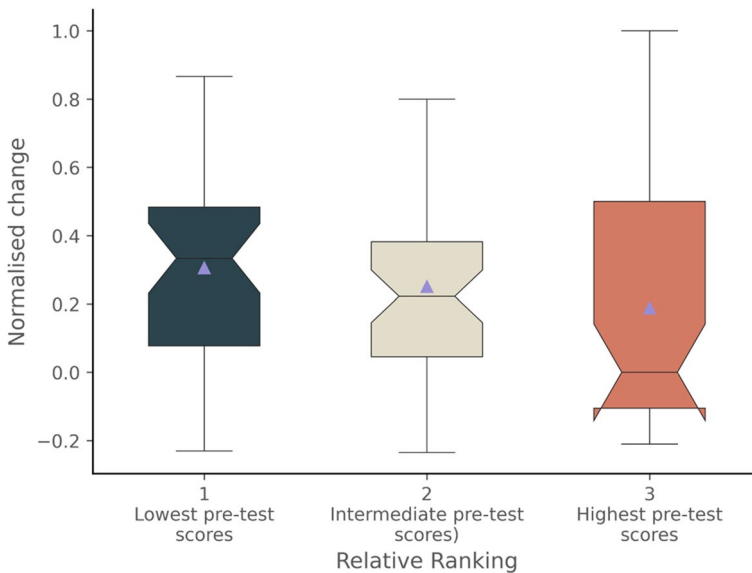
post-test ( $\Delta = 2.514$  points out of 25,  $p = 0.002$ , Cohen's  $D = 0.615$ ) than boys ( $\Delta = 2.017$  points out of 25,  $p = 0.012$ , Cohen's  $D = 0.5$ ). While the differences are not significant, it would appear that boys were initially performing slightly better than girls ( $\Delta = 0.597$ pts,  $p = 0.6367$ ,  $D = 0.134$ ) and that the gap is closing after having participated in the program ( $\Delta = 0.1$ pts,  $p = 0.9026$ ,  $D = 0.028$ ).

To further refine the analysis, we consider the normalised change to account for the progress made by each individual student. Over 75% of students progressed on CT-concepts following the intervention with an average of 24% of progress. Just 25% showed a decrease in their scores (i.e., did better on the pre-test than the post-test).

The distribution of the initial ranking of the students in the pretest is normal for each group (Group 1 (low),  $D = 0.6$ ,  $p = 0.7$ ; Group 2 (medium),  $D = 0.6$ ,  $p = 0.7$ ; Group 3 (high),  $D = 0.6$ ,  $p = 0.7$ ). Considering the distribution of the normalised change according to the students initial ranking in the pre-test (see Fig. 8), it would appear that there is a difference between the initially low, medium and high performers established through a relative rank-based approach (one-way ANOVA  $F(1) = 6.58$ ,  $p = 0.037$ ). Dunn's test for multiple comparisons indicates that students in Group 1, or the low performers, differ significantly from those in Group 3, our high performers ( $\Delta NC = 0.117$ ,  $p = 0.0071$ , Cohen's  $D = 0.374$ ).

## 5.2 Teachers' perception of the ABC-Thinking program

Several results can be drawn from the (1) Teacher Monitoring Form (see Section 4.2.2), that the teachers filled in each week. As expected, the ease with which the teachers perceived the students to carry out the activities increased slightly over the weeks with an overall average of 4.2 (Likert scale from 1 to 5), thus, hardly any



**Fig. 8** Distribution of the normalised change for the students according to their ranking on the pre-test

difficulties were observed in this aspect, with a mean of 3.8 on the Likert scale of 1-5 in terms of ease to carry out the activities at the beginning of the program. Similarly, the teachers considered that the students were enthusiastic from the beginning of the program (with an average of 4.29), and even considered that it increased over time (with an overall average of 4.5).

Based on their classroom observations, the teachers considered that the program offered numerous benefits, including a high level of student autonomy from the beginning and a willingness to collaborate with their peers. In addition, the students showed a very high motivation that the teachers believed did not decline over time, an improvement in their ability to work in groups, as well as in their “individual autonomy”, “self-esteem and confidence”, and in their “ability to solve problems easily”.

The teachers also identified weaknesses in the program, such as the students’ need for additional pencil and paper to think through problems in addition to using the computer, or even the need to print the challenges on paper directly, although this need decreased over the weeks as students became more confident in their abilities and developed specific skills. Some teachers mentioned difficulties with the usage of the web-based platform, which took students’ concentration away and limited the amount of time that could be spent on the activity, which was generally too short for some students. Another problem was the use of overly complex language in the Bebras task statements given the students’ age.

In the (2) Teachers Evaluation Form (see Section 4.2.2), they indicated that they felt prepared to teach the program, with an average score of 3.86+/1.1 (Likert scale from 1 to 5). What the teachers valued most was the wide variety of exercises proposed, which they would not have had access to if it had not been for the program. However, they also mentioned the need for more time to complete the exercises and for better resources to access the online platform (e.g., better internet connection or better devices). In general, all teachers were very or fully satisfied with the experience with an average score of 4.67+/0.6 (Likert scale from 1 to 5) and would like to repeat the experience next year.

The difficulty of the exercises was perceived by the teachers as medium or high, as well as the fact that it was not easy for the students to use the online platform to do the exercises. In summary, the teachers believed they improved their training and understanding of computational thinking and found the experience highly rewarding and very motivating for the students, as well as perceived improvements in their CT and teamwork skills and, therefore, in their CT perspectives dimension. In an open question, most of the teachers indicated they would like to see the program incorporated into the school curricula on a permanent basis.

Finally, the project coordinators, in their concluding (3) Coordinator Interview (see Section 4.2.2), highlighted the great enthusiasm of the students throughout the program, as they felt that they were playing while learning, they valued very positively the type of exercises proposed, and were looking forward to “the day of computational thinking”. In addition, the coordinators felt that the students made clear progress in their logical reasoning, and even learned to pose different hypotheses for the solution of a problem and to work in an interdisciplinary way. Although there were some difficulties at the beginning, especially with the Internet connection, the web platform was intuitive, and the children were able to use the tablets as well as

the computers. As an aspect to be improved, the coordinators proposed to modify the online platform so that it is not necessary to use pencil and paper in addition to the computer, since it is possibly an unnecessary expense, or at least, to try to make it possible to print the exercises in black and white without losing relevant information for solving the problems. In addition, it was proposed to increase the time to carry out the activities in each session, especially to reflect on the problem, since it was sometimes insufficient.

The coordinators considered that the teachers adapted very easily to the work dynamics and felt that they were prepared to implement the project, which is remarkable since the coordinators consider that the project is compatible with the new Portugal school curriculum that includes this type of competencies, and also implementing this project has helped that, in addition to the students and teachers, the students' families also become familiar with the term "computational thinking". As ideas that could be incorporated in the future, the coordinators proposed the creation of a teacher-forum to exchange experiences in an effort to create a community of practice (Coburn & Stein, 2006), an element which has been found to contribute to promoting and sustaining changes in teacher practices, as well as an initial joint presentation for all teachers. To improve instruction, the coordinators also suggested incorporating a set of introductory exercises to be offered to students who require them, thus creating a more inclusive program, and send feedback at the end of the program so teachers may gain insight into where their students stand in terms of CT. Finally, the coordinators suggested adapting and testing the ABC-Thinking program on younger children to broaden the range of validated Bebras CT curricula for use by researchers and practitioners alike.

## 6 Discussion and limitations

### 6.1 Discussion

First of all, it is important to emphasize that teachers were adequately trained, and that the ABC-Thinking program includes gamified Bebras tasks, that have been selected in a balanced and development-oriented way for the development of CT according to (Dagiené et al., 2017) categorization (Table 1).

To address the first research question (RQ1) regarding which CT skills can be developed through the Bebras tasks, an analysis of the results has been carried out after having implemented the ABC-Thinking program and collected the perspectives of the teachers involved, as well as the results of the cCTt applied as a pre- and post-test to the students. The results show a significant overall increase in the development of CT at the end of the program with a medium to large Cohens' effect size (see Section 5.1). When we consider what particular skills are developed regarding CT computational concepts (Brennan & Resnick, 2012), we can observe a major improvement in the case of simple concepts - less in the more complex ones, such as *conditionals*, *while statements* or in the combinations of concepts - (see Section 5.1), which could indicate that the ABC-Thinking program is suitable for

developing CT skills but that additional interventions are necessary for students to improve in these advanced CT concepts. In terms of gender, boys perform slightly better on the pre-test than girls, but the gap is closes after having participated in the ABC-Thinking program. Similarly, it is noteworthy that the improvement is significantly greater in the group of students who are low performers, which may indicate that this program may be especially suitable for those students who present some difficulty or low prior knowledge in the area.

Regarding RQ2, on the suitability of the ABC-Thinking program with respect to teachers practice, teachers mostly highlighted the very high motivation of the students, their ability to solve problems in a collaborative way, as well as their self-esteem and self-efficacy. The selection of Bebras tasks specifically chosen for the development of the CT was positively valued. Furthermore, they considered that they had the appropriate training to be able to teach it and that it could be included in the official school curriculum, as they found it practical and easy to incorporate into their professional practice, even if the school had limited technological resources.

Currently, there are scarce comprehensive programmes for the development and evaluation of CT for schools. The ABC-Thinking program combines (1) training for educators; (2) a selection and adaptation of activities for the development of CT, in this case selected according to (Dagienė et al., 2017) categorisation and adapted through gamification; (3) a monitoring system of the program through different forms and interviews; (4) an evaluation of the improvement in CT through a validated test, in this case the cCTt.

Once the ABC-Thinking program has been implemented and the results have been analysed, it has been demonstrated that, with a complete educational approach, the development of CT can be enhanced significantly, while students and teachers enjoy and are motivated. In addition, the assessment of the improvement in CT is a problem that teachers regularly face. In this sense, the program also provides this assessment.

## 6.2 Limitations

There are several limitations which can be raised regarding the present study. From the students' perspective, we did not have access to a control group and so the progress observed may not just result from the participation in the ABC-Thinking program and may also be impacted by related elements in the maths curriculum for instance. The cCTt mainly focuses on CT concepts, and partially on CT practices. While the instrument is well aligned with the ABC-Thinking program that is highly oriented towards algorithmic thinking and evaluation, there are other dimensions of CT that could be considered in the validation of the program (Lye & Koh, 2014). In particular, their CT practices and perspectives could be better understood if other assessment methods were included in the proposed "system of assessments". Indeed, multiple data sources, including qualitative data, would also be interesting to triangulate the findings. It is however important to note the limited number of validated instruments permitting the assessment of these dimensions at scale at the primary school level. As the students participated in the ABC-Thinking curriculum and solved tasks in pairs and groups, we also lack the individual Bebras Tasks scores which would have been interesting to

correlate with their performance on the post test. Where the gender findings are concerned, it would be interesting to replicate the analysis on a larger sample to determine whether there is indeed an existing gender-gap prior to participating in the program and if the gap is being closed after having partaken in it.

More generally, it seems that for some learners this was the first time they had come into contact with a web-based platform, as many of them had little experience with computers. This also begs a question regarding the fact that these students are considered “digital natives” (Prensky, 2001). Indeed, the new generation of students are expected to be proficient in the use of technology because they were born in the era of technology. The result is that, while there is a focus on integrating Computer Science and Computational Thinking in curricula worldwide, there is also a contradiction because students find themselves struggling with the basic use of a computer (Chevalier et al., 2022). The use of a web-based platform, while initially considered to be a limitation, may thus also be perceived as an opportunity of the program by helping kids to learn how to use the internet/computers. In sum, creators of Computer Science and CT programs must bear in mind that students may not have basic digital skills.

From the perspective of the ABC-Thinking program, it would appear that the students improved on certain CT-concepts more than others, possibly indicating that the tasks should be selected to target the specific competences being assessed, in accordance with the principles of constructive alignment. Although it may be that the students may require more targeted interventions for certain CT-concepts, more generally the ABC-Thinking tasks have a stronger focus on algorithmic thinking ( $n = 29$ ) and evaluation ( $n = 32$ ) than decomposition ( $n = 4$ ) or abstraction ( $n = 7$ ). The selection of activities could thus be reviewed in the future to propose a more balanced set of tasks. It is important to note that this relies on having a CT classification for a larger sample of Bebras Tasks. Only by doing so will researchers and practitioners be able to appropriately select tasks which cover the full range of CT-competences. Researchers and practitioners should indeed consider these elements and the multi-dimensional classification of Bebras tasks before deciding to employ them as assessments or interventions.

From the teachers’ perspective, the study involved a small cohort of voluntary teachers which may introduce a selection bias. Indeed, these teachers are likely those who were more intrinsically motivated to participate in the program. It would thus be interesting to evaluate the program with teachers who are not familiar with CT and not initially motivated to teach CT. The evaluation with teachers could also include more detailed rubrics according to established frameworks such as TPACK (Koehler & Mishra, 2009), Utility Usability Acceptability (Tricot et al., 2003), Technology Acceptance Models (King & He, 2006).

## 7 Conclusions

After having applied the ABC-Thinking program and analysed the perspectives of the teachers involved, as well as the results of the cCTt applied as a pre- and post-test to the students, it can be concluded that the program contributes significantly to

the learning of CT at the Primary Education stage. Although the results show a significant overall improvement, there is a greater improvement in the simpler CT computational concepts and a smaller improvement in the more complex ones, which could indicate that additional interventions are necessary for students to improve in these advanced CT concepts.

In terms of gender, boys perform slightly better on the pre-test than girls. This result is consistent with previous research, regarding the higher performance on Bebras tasks of boys versus girls, increasing dramatically with age (Bellettini et al., 2015; Hubwieser & Mühlhng, 2015; Kaleliođlu et al., 2015; Román-González et al., 2017). Currently, work is being done on this aspect, so that gender gaps in CT and STEM can be closed and these differences do not affect later academic choices (Rachmatullah et al., 2022). It is relevant, therefore, that the girls appeared to improve more than boys (who were performing better initially) through the ABC-Thinking program, as this could help to bridge the existing digital divide before it begins to widen with age.

Similarly, it is noteworthy that the improvement is significantly greater in the group of students who are low performers, which may indicate that this program may be especially suitable for those students who present some difficulty or low prior knowledge in the area. Given that unplugged type activities, such as Bebras tasks, for the development of CT are considered to be specially adapted for young students or students with special needs (Zapata-Ros, 2019), this observation could reinforce this argument.

Regarding the suitability of the ABC-Thinking program, teachers mostly highlighted the very high motivation of the students, as well as their self-efficacy. Furthermore, they considered that they had the appropriate training to be able to teach it and that it could be included in the official school curriculum, as they found it practical and easy to incorporate into their professional practice, even if the school had limited technological resources. The program can also act as a “disseminator” of CT, so that in the student’s environment the subject becomes more familiar.

This study also demonstrates that CT can be appropriately enhanced and developed through comprehensive programmes, such as the ABC-Thinking program, which combine a comprehensive educational approach covering several selected and categorised areas of CT, including a variety of teaching and assessment activities and strategies. Of course, the programme has limitations, but we consider it as a first step towards establishing specific and effective programmes in the school curricula that are also motivating for teachers and students.

The ABC-Thinking program was designed to have teachers teach CT in their classrooms and was evaluated using quantitative data from the cCTt and qualitative data from teacher observations, as part of a “system of assessments”. We consider that the results could be a first step to incorporate the Bebras tasks also as a CT assessment tool, as part of the “system of assessments”, although with caution in very young children or with special needs, and not as the only tool, since it contains a high reading comprehension skills load. Furthermore, it is necessary to consider that the different CT skills were not completely balanced in the ABC-Thinking program, since the Bebras tasks are not focused only on CT, indicating that special care should be taken in the selection of the set of tasks when developing a Bebras-based CT program.



## Appendix 1

**Table 5** The Bebras tasks included in the ABC-Thinking program and their CT skills classification. Abbreviations - CT1: Abstraction, CT2: Algorithmic thinking, CT3: Decomposition, CT4: Evaluation, CT5: Generalisation

Exercise set	Bebras task name	Computational Thinking skills covered by the task*				
		CT1	CT2	CT3	CT4	CT5
1	Cleaning		✓			
1	Classifier	✓				✓
1	Bridges and islands		✓	✓		
2	Candy shop		✓			
2	Socks		✓			
2	Box of balls		✓		✓	
2	Snowmen's hats		✓		✓	✓
3	Scratch art and paper	✓			✓	✓
3	Birthday cake		✓		✓	✓
3	Rangoli design		✓		✓	✓
3	Seating plan		✓	✓	✓	
4	Push away parking		✓			
5	Movie theater		✓		✓	
5	Way home		✓		✓	
5	Cloud communication				✓	✓
5	Stamps	✓			✓	✓
5	Bath at the lido	✓			✓	
5	Beavers dance contest		✓		✓	
6	Rabbit in the forest		✓			
6	Treasure hunt		✓			
6	Snow removal		✓			
7	Lockers and coding		✓		✓	✓
7	Beavercoins				✓	
7	Cryptic language		✓			
7	Binary bulbs		✓		✓	✓
8	Cleaning		✓			
8	Scratch art and paper	✓			✓	✓
8	Socks		✓			
8	Box of balls		✓		✓	
8	Birthday cake		✓		✓	✓
8	Rangoli design		✓		✓	✓
8	Bridges and islands		✓	✓		
8	Seating plan		✓	✓	✓	
9	Plates				✓	
10	Colorful route		✓			
10	Golfer bebras		✓		✓	
10	Koko's animals	✓	✓		✓	
10	Meal planning		✓		✓	

**Table 5** (continued)

Exercise set	Bebras task name	Computational Thinking skills covered by the task*				
		CT1	CT2	CT3	CT4	CT5
11	Space travelling		✓		✓	
11	Carpeting		✓		✓	✓
11	Light buttons		✓		✓	
11	Making stitches		✓		✓	✓
12	Stamps	✓			✓	✓
12	Cloud communication				✓	✓
12	Colorful route		✓			
12	Golfer bebras		✓		✓	
12	Lockers and coding		✓		✓	✓
12	Beavercoins				✓	
12	Treasure hunt		✓			
12	Space travelling		✓		✓	

## Appendix 2

**Table 6** Games used in the ABC-Thinking program

Game	Instructions	Rules	Goal
Teamwork	Split the class in pairs. Ask each pair to solve the exercises.	Each correct answer adds 1 point (easy questions), 2 points (medium questions), or 5 points (hard questions) to the class. Each wrong answer subtracts 1 point to the class. Each non answered question subtracts 2 points to the class.	The class wins if the sum of the scores of all pairs is equal to or greater than the number of pairs.
Beat the sum	Each student solves the exercises individually. Then, after 20 minutes, they join a colleague, discuss the exercises, and agree on a pair answer.	Each correct answer adds 1 point (easy questions), 2 points (medium questions), or 5 points (hard questions) to the class. Each wrong answer subtracts 1 point to the class. Each non answered question subtracts 2 points to the class.	The class wins if the sum of the scores of all pairs is greater than the number of pairs.
Double time	Split the class in pairs. Each pair has to solve four problems, but there is a time limit for each problem. When the time is up, all the pairs need to move to the following problem.	Easy questions have a time limit of 3 minutes. Medium questions have a time limit of 5 minutes. Hard questions have a time limit of 7 minutes. Each correct answer adds 1 point (easy questions), 2 points (medium questions), or 5 points (hard questions) to the class. Each wrong answer subtracts 1 point to the class. Each non answered question subtracts 2 points to the class.	The class wins if the sum of the scores of all pairs is equal to twice the number of pairs.
Easy going	Each student solves the exercises individually. Then, after 20 minutes, they join a colleague, discuss the exercises, and agree on a pair answer.	Each correct answer adds 1 point (easy questions), 2 points (medium questions), or 5 points (hard questions) to the class. Each wrong answer subtracts 1 point to the class. Each non answered question subtracts 2 points to the class.	The class wins if each pair has, at least, a total of 2 points.

Table 6 (continued)

Game	Instructions	Rules	Goal
Reach the level	Split the class in pairs. Each student of the pair solves a different problem. Once both students solve their problems, they show the answers to the teacher. If both answers are correct, the teacher gives them two new problems (one for each student). If one of the answers is wrong, the teacher provides feedback and challenges the students to try again.	<p>There are three levels (Level 1, Level 2, Level 3) and each level has two problems (Level 1 has Problem 1 and 2, Level 2 has Problem 3 and 4, Level 3 has Problem 5 and 6).</p> <p>Each student of the pair solves a different problem, according to the level where they are. For example, if they are at Level 1, Student A gets Problem 1 and Student B gets Problem 2.</p> <p>To move to the next level, both students need to solve their problem correctly.</p> <p>Once the pair finishes, the teacher verifies the answers and if both are correct, the pair gets two new problems and moves to the next level.</p> <p>There are no scores in this activity.</p>	The class wins if more than half of the pairs complete the Level 2 successfully.
Once a group, always a group	Each student solves the exercise individually. Every time the class does not reach a specific score, the students are grouped in pairs to try again.	<p>There are three levels (Level 1, Level 2, Level 3) and each level has one problem (Problem 1, Problem 2, Problem 3).</p> <p>To move to the next level, two thirds of the class need to answer the problem correctly.</p> <p>Each time this number is not reached, the students are grouped in pairs to review the answer and try again.</p> <p>There are no scores in this activity.</p>	The class wins if they complete Level 3 successfully.

**Acknowledgements** Thanks to the participants, teachers and schools involved in the learning experiences.

**Authors' contributions** MZC contributed to the paper conceptualization, methodology, writing, and submission preparation. PM contributed to the methodology, investigation, and writing. LEH contributed to the formal analysis, validation, and writing. EMB contributed to the supervision, review, and editing preparation. All authors read and approved the final manuscript.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was funded by the Calouste Gulbenkian Foundation through the Gulbenkian Knowledge Academies programme; supported by the NCCR Robotics, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 51NF40\_185543); and supported by the Madrid Regional Government, through the project e- Madrid-CM (P2018/TCS-4307) which is also co-financed by the Structural Funds (FSE and FEDER).

**Data availability** The datasets generated and/or analysed during the current study are available in the Zenodo repository, [persistent web link to datasets will be publicly available upon publication].

## Declarations

**Competing interests** The authors have no competing interests to declare that are relevant to the content of this article.

**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

- Acha, M. T., Puchades, J. M. R., & Ajamil, D. L. (2021). Pensamiento computacional en la resolución de problemas contextualizados en un cuento en Educación Infantil. *Edma 0-6: Educación matemática en la infancia*, 9(2), 73–92. <https://doi.org/10.24197/edmain.2.2020.73-92>
- Bell, T., Alexander, J., Freeman, I., & Grimley, M. (2009). Computer science unplugged: School students doing real computing without computers. *The New Zealand Journal of Applied Computing and Information Technology*, 13(1), 20–29.
- Bellettini, C., Lonati, V., Malchiodi, D., Monga, M., Morpurgo, A., & Torelli, M. (2015). How challenging are Bebras tasks? An IRT analysis based on the performance of Italian students. *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*, 27–32. <https://doi.org/10.1145/2729094.2742603>
- Brackmann, C. P., Román-González, M., Robles, G., Moreno-León, J., Casali, A., & Barone, D. (2017). Development of computational thinking skills through unplugged activities in primary school. *Proceedings of the 12th Workshop on Primary and Secondary Computing Education*, 65–72. <https://doi.org/10.1145/3137065.3137069>
- Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 Annual Meeting of the American Educational Research Association*, 1, 25.
- Caeli, E. N., & Yadav, A. (2020). Unplugged approaches to computational thinking: A historical perspective. *TechTrends*, 64(1), 29–36. <https://doi.org/10.1007/s11528-019-00410-5>

- Cartelli, A., Dagiene, V., & Futschek, G. (2010). Bebras contest and digital competence assessment: Analysis of frameworks. *International Journal of Digital Literacy and Digital Competence (IJDLDC)*, 1(1), 24–39. <https://doi.org/10.4018/jdlc.2010101902>
- Chevalier, M., Giang, C., El-Hamamsy, L., Bonnet, E., Papaspyros, V., Pellet, J., Audrin, C., Romero, M., Baumberger, B., & Mondada, F. (2022). The role of feedback and guidance as intervention methods to foster computational thinking in educational robotics learning activities for primary school. *Computers & Education*, 180, 104431. <https://doi.org/10.1016/j.compedu.2022.104431>
- Coburn, C. E., & Stein, M. K. (2006). Communities of practice theory and the role of teacher professional community in policy implementation. *New directions in education policy implementation: Confronting complexity*, 63, 65–72.
- Coletta, V. P., & Steinert, J. J. (2020). Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. *Physical Review Physics Education Research*, 16(1), 010108. <https://doi.org/10.1103/physrevphyseducres.16.010108>
- Dagiene, V., & Dolgopolas, V. (2022). Short tasks for scaffolding computational thinking by the global Bebras challenge. *Mathematics*, 10(17), 3194. <https://doi.org/10.3390/math10173194>
- Dagiene, V., & Futschek, G. (2008). Bebras international contest on informatics and computer literacy: Criteria for good tasks. *International Conference on Informatics in Secondary Schools Evolution and Perspectives*, 19–30. [https://doi.org/10.1007/978-3-540-69924-8\\_2](https://doi.org/10.1007/978-3-540-69924-8_2)
- Dagiene, V., & Stupuriene, G. (2016). Bebras - a sustainable community building model for the concept based learning of informatics and computational thinking. *Informatics in education*, 15(1), 25–44. <https://doi.org/10.15388/infedu.2016.02>
- Dagiene, V., Mannila, L., Poranen, T., Rolandsson, L., & Stupuriene, G. (2014). Reasoning on children's cognitive skills in an informatics contest: Findings and discoveries from Finland, Lithuania, and Sweden. *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*, 66–77. [https://doi.org/10.1007/978-3-319-09958-3\\_7](https://doi.org/10.1007/978-3-319-09958-3_7)
- Dagiene, V., Sentance, S., & Stupuriene, G. (2017). Developing a two-dimensional categorization system for educational tasks in informatics. *Informatica*, 28(1), 23–44. <https://doi.org/10.15388/informatica.2017.119>
- d'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2), 341–348. <https://doi.org/10.2307/2334522>
- Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education*, 14(1), 1–36. <https://doi.org/10.1186/s41239-017-0042-5>
- El-Hamamsy, L., Zapata-Cáceres, M., Barroso, E. M., Mondada, F., Zufferey, J. D., & Bruno, B. (2022). The competent computational thinking test (cCTt): Development and validation of an unplugged computational thinking test for upper primary school. *Journal of Educational Computing Research*, 60(7), 1818–1866. <https://doi.org/10.1177/07356331221081753>
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38–43. <https://doi.org/10.3102/0013189x12463051>
- Grover, S., & Pea, R. (2015). “Systems of Assessments” for deeper learning of computational thinking in K-12. *Proceedings Annual Meeting American Educational Research Association*, 2015, 15–20.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – A Literature Review of Empirical Studies on Gamification, 47th Hawaii International Conference on System Sciences, 3025–3034. <https://doi.org/10.1109/hicss.2014.377>
- Hsu, T., Chang, S., & Hung, Y. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education*, 126, 296–310. <https://doi.org/10.1016/j.compedu.2018.07.004>
- Hubwieser, P., & Mühlhng, A. (2015). Investigating the psychometric structure of Bebras contest: Towards measuring computational thinking skills, 2015 International Conference on Learning and Teaching in Computing and Engineering, 62–69. <https://doi.org/10.1109/latic.2015.19>
- Jun, S., Jeon, Y., Kim, S., Kim, D., & Jeong, I. (2018). Elementary school students' level of computational thinking through Bebras challenge 2017. *Journal of The Korean Association of Information Education*, 22(3), 345–356. <https://doi.org/10.14352/jkaie.2018.22.3.345>
- Kalas, I., & Tomcsanyi, M. (2009). Students' Attitude to Programming in Modern Informatics. *Proceedings 9th WCCE 2009, Education and Technology for a Better World*.
- Kalelioğlu, F., Gülbahar, Y., & Madran, O. (2015). A snapshot of the first implementation of Bebras international informatics contest in Turkey. *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*, 131–140. [https://doi.org/10.1007/978-3-319-25396-1\\_12](https://doi.org/10.1007/978-3-319-25396-1_12)

- Kazakoff, E. R., Sullivan, A., & Bers, M. U. (2013). The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. *Early Childhood Education Journal*, 41(4), 245–255. <https://doi.org/10.1007/s10643-012-0554-5>
- King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, 43(6), 740–755. <https://doi.org/10.1016/j.im.2006.05.003>
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary issues in technology and teacher education*, 9(1), 60–70.
- Lehtimäki, T., Monahan, R., Mooney, A., Casey, K., & Naughton, T. J. (2022). Bebras-inspired computational thinking primary school resources co-created by computer science academics and teachers. *Proceedings of the 27th ACM Conference on Innovation and Technology in Computer Science Education*, 1, 207–213. <https://doi.org/10.1145/3502718.3524804>
- Lewrick, M., Link, P., & Leifer, L. (2020). *The design thinking toolbox: A guide to mastering the most popular and valuable innovation methods*. Wiley.
- Lockwood, J., & Mooney, A. (2018). Developing a computational thinking test using Bebras problems. *Proceedings of TACKLE: the 1st Systems of Assessments for Computational Thinking Learning*.
- 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>
- Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M. A., Guerrero-Puerta, L., Aguilar-Parra, J. M., Trigueros, R., & Alias, A. (2021). Between level up and game over: A systematic literature review of gamification in education. *Sustainability*, 13(4), 2247. <https://doi.org/10.3390/su13042247>
- Prensky, M. (2001). *Digital game-based learning*. McGraw-Hill.
- Rachmatullah, A., Vandenberg, J., & Wiebe, E. (2022). Toward more generalizable CS and CT instruments: Examining the interaction of country and gender at the middle grades level. *Proceedings of the 27th ACM Conference on Innovation and Technology in Computer Science Education*, 1, 179–185. <https://doi.org/10.1145/3502718.3524790>
- Román-González, M., Pérez-González, J., & 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., 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>
- Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In: Anonymous *Computational thinking education* (pp. 79–98). Springer. [https://doi.org/10.1007/978-981-13-6528-7\\_6](https://doi.org/10.1007/978-981-13-6528-7_6)
- Sailer, M., & Homner, L. (2019). The gamification of learning: A Meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Schöbel, S. M., Janson, A., & Söllner, M. (2020). Capturing the complexity of gamification elements: A holistic approach for analysing existing and deriving novel gamification designs. *European Journal of Information Systems*, 29(6), 641–668. <https://doi.org/10.1080/0960085x.2020.1796531>
- Selby, C., & Woollard, J. (2013). *Computational thinking: the developing definition*. Southampton, UK: University of Southampton. Retrieved from <https://core.ac.uk/download/pdf/17189251.pdf>
- Shute, V. J., & Moore, G. R. (2017). Consistency and validity in game-based stealth assessment. In H. Jiao & R. W. Lissitz (Eds.), *Technology enhanced innovative assessment: Development, modeling, and scoring from an interdisciplinary perspective*, 296, 31–51
- Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142–158. <https://doi.org/10.1016/j.edurev.2017.09.003>
- Suh, S., Lee, M., & Xia, G. (2020). Coding strip: A pedagogical tool for teaching and learning programming concepts through comics, *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 1–10. <https://doi.org/10.1109/vl/hcc50065.2020.9127262>
- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24(6), 837–848. <https://doi.org/10.1002/acp.1598>
- Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 27(1), 77–83. <https://doi.org/10.3102/10769986027001077>
- Tricot, A., Plégat-Soutjis, F., Camps, J., Amiel, A., Lutz, G., & Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité: interpréter les relations entre trois dimensions de l'évaluation des EIAH, *Environnements Informatiques pour l'Apprentissage Humain*, 391–402



- Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., & Ninaus, M. (2022). A cognitive definition of computational thinking in primary education. *Computers & Education*, 179, 104425. <https://doi.org/10.1016/j.compedu.2021.104425>
- Waite, J., Maton, K., Curzon, P., & Tuttiett, L. (2019). Unplugged computing and semantic waves: Analysing crazy characters. *Proceedings of the 1st UK & Ireland Computing Education Research Conference*, 1–7. <https://doi.org/10.1145/3351287.3351291>
- Zapata-Cáceres, M., & Martín-Barroso, E. (2021). Applying game learning analytics to a voluntary video game: Intrinsic motivation, persistence, and rewards in learning to program at an early age. *IEEE Access*, 9, 123588–123602. <https://doi.org/10.1109/access.2021.3110475>
- Zapata-Cáceres, M., Martín-Barroso, E., & Román-González, M. (2020). Computational thinking test for beginners: Design and content validation. *2020 IEEE Global Engineering Education Conference (EDUCON)*. Porto, Portugal, 1905–1914. <https://doi.org/10.1109/educon45650.2020.9125368>
- Zapata-Ros, M. (2019). Computational thinking unplugged. *Education in the Knowledge Society*, 20, 1–29.

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

## Authors and Affiliations

María Zapata-Cáceres<sup>1</sup>  · Pedro Marcelino<sup>2</sup>  · Laila El-Hamamsy<sup>3,4</sup>  ·  
Estefanía Martín-Barroso<sup>1</sup> 

✉ María Zapata-Cáceres  
maria.zapata@urjc.es

Pedro Marcelino  
pedro.marcelino@treetree2.org

Laila El-Hamamsy  
laila.elhamamsy@epfl.ch

Estefanía Martín-Barroso  
estefania.martin@urjc.es

<sup>1</sup> Computer Science Department, Universidad Rey Juan Carlos, C/ Tulipán, s/n 28933, Móstoles, Madrid, Spain

<sup>2</sup> Computational Thinking Department, TreeTree2, Lisbon, Portugal

<sup>3</sup> MOBOTS Group, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

<sup>4</sup> LEARN, Center for Learning Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland