Computational Thinking in Primary School Using Educational Robots: Construction and Validation of an Assessment Tool

Ana Isabel Santos[1, 2,*], José Manuel Cascalho[3,4], Armando B. Mendes[3,4], Matthias Funk[3,4], Bárbara Amaral[3,4], Francisco Marques[3,4], and Paulo Medeiros[3]

[1]Faculty of Social Sciences and Humanities & NICA, University of the Azores, Ponta Delgada, Portugal
[2]CIE-ISPA, Lisbon, Portugal
[3]Faculty of Sciences and Technology, University of the Azores, Ponta Delgada, Portugal
[4]IS2E and Artificial Intelligence and Computer Science Laboratory, Porto, Portugal
Email: ana.is.santos@uac.pt (A.I.S.); jose.mv.cascalho@uac.pt (J.M.C.); armando.b.mendes@uac.pt (A.B.M.); gunther.ma.funk@uac.pt (M.F.); barbara.c.amaral@hotmail.com (B.A.); francisco6457marques@gmail.com (F.M.); paulo.jf.medeiros@uac.pt (P.M.)
*Corresponding author

Manuscript received August 22, 2024; revised September 10, 2024; accepted September 29, 2024; published December 11, 2024

***Abstract*—The aim of this paper is to present the construction, testing and validation of an instrument to assess the Computational Thinking skills of elementary school children, a tool that was developed as part of the PeCOT project, a research and pedagogical intervention project that seeks to introduce robotics activities in primary school classrooms, using the robot Azbot-1C, designed by the research team. The instrument whose validation we present is designed to assess the impact of the activities on the development of children's computational thinking considering four essential competences: algorithmic thinking, patterns, decomposition, and abstraction. It takes the form of an activity sheet describing problems of different levels of complexity related to each of the four CT competences, which the children must solve for about an hour, in order assess the development of these skills in the children who will participate in the educational robotics activities that will be implemented during the second year of the project. Eight children aged between 8 and 10 took part in this validation. They answered the challenges at two different moments during the school year, thus reinforcing the instrument's reliability. The results of solving each problem were scored, unforeseen situations were recorded and the analysis of the tasks presented made it possible to adjust the assessment tool in terms of its clarity, length and complexity. Ultimately, the instrument should be a valid contribution to help teachers and researchers observe and assess CT skills of children of this age, especially those currently carrying out activities related the implementation of the CT Project in the Azores.**

***Keywords*—computational thinking, assessment instrument, educational robotics**

I. Introduction

Nowadays, Computational Thinking (CT) has gained a lot of importance, both in terms of a better clarification of its concept [1–3], and in terms of its applicability, especially in an educational context [4–6]. In any case, the relevance of its analysis and approach has been widely documented, considering the positive results obtained in various fields, such as mathematics and science [7] or creativity [8], in aspects related to students' personal and social development [9] and in different educational contexts, from Early Childhood Education [10] to Secondary Education [11].

In Portugal, since 2022, CT has become part of everyday life in many schools, with its introduction in the Core Learning of Mathematics. This skill, recognized by the European Union as fundamental and to be promoted from the earliest years, leads us to the articulated development of a set of practices that contribute to solving problems, particularly those related to programming. In the Autonomous Region of the Azores, where this research is being carried out, CT began to be introduced in the first year of basic education in the 2022/2023 school year, through the implementation of unplugged activities, with a transdisciplinary nature, in a logic of curricular enrichment [12].

With this in mind, the PeCOT project—Computational Thinking with Tangible Objects - has emerged as a project that seeks to reinforce the importance of pedagogically supported work in this field, offering teachers a set of pedagogical intervention strategies that allow them to use robots in the classroom for educational purposes, and, on the other hand, to develop a tool that makes it possible to assess these CT skills and contribute to the use of tangible objects.

In this context, the aim of this document is to describe the work carried out to design and validate such a tool, which was evaluated by children participating in the second year of the project, who carried out a series of activities using a simple educational robot built by the research team. Using Azbot-1C, a programmable robot with buttons, the children participating in the project will have the opportunity to solve challenges in different curricular areas through its use in a grid scenario, while at the same time working on CT skills.

II. Literature Review

A. Computational Thinking and Its Evaluation

The concept of Computational Thinking refers to a set of skills that enable students to solve problems logically and efficiently. Including the ability to identify and formulate problems, develop algorithms to solve them, test and debug these algorithms and communicate their solutions, CT "is a fundamental skill for everyone, not just for computer scientists" [13].

Although there is no agreed definition among researchers, Allsop [14] identifies five aspects that are intrinsic to CT:

- is a cognitive process
- is regulated by metacognitive practices
- involves the application of a series of computational concepts

- includes the utilisation of learning behaviours
- aims to design solutions to problems that are susceptible to automation (p.31).

As Gane *et al*. [15] point out, CT is not just about computing skills, it goes far beyond the mere ability to program. It includes activities related to algorithms, decomposition, sequencing instructions and recognizing and repeating patterns.

In a paper published in 2014 that reviews the literature on the concept of CT, Selby and Woollard [2] conclude that there are some terms that are consensual among researchers, such as thought process, related to the ability to solve problems, abstraction, and decomposition, and that there are others that are commonly used by many of these researchers, such as algorithmic thinking, logical thinking or generalization. Following these ideas, Beecher [16] suggests that the terms logical thinking, algorithmic thinking, decomposition, generalization and pattern recognition, modelling, abstraction, and evaluation should be considered fundamental.

While its definition has not been consensual, its assessment has not been a simple task either. As Li *et al*. [17] point out, there have been many methodological strategies have been used to assess CT, such as questionnaires, portfolios, knowledge, and skills tests, among others. As the authors state, "many CT assessments are embedded in specific programming environments or training curricula or require knowledge of a specific programming language" (p.4), which makes it difficult to apply them to other contexts, considering that CT-related skills are transversal thinking skills.

Based on a review of the literature in this area, Vinu *et al*. [18] identify two broad categories that encompass the way in which CT has been evaluated in an educational context: workshops and curriculum-based interventions, increasingly using strategies that are considered less traditional, far from block programming, such as art, visualization techniques, deep learning, unplugged activities, augmented, and virtual reality environments. However, they conclude that there are no standardized instruments for assessing CT in different contexts and situations, which is why it is important to continue to develop methodological strategies for this purpose.

In our case, because the environment in which the project is applied is socially disadvantaged and because all the students have difficulties in accessing technology, we chose a strategy of assessing CT through an unplugged activity. An activity sheet was used that included a series of problem situations related to CT activities, which will be explained in the following sections. However, other ways of assessing CT will also be implemented as part of the project, taking into account the recommendations of studies such as Allsop [14] or Çoban and Korkmaz [19] to diversify the ways of assessing CT.

On the other hand, factors such as the age or developmental level of the children or young people for whom the CT promotion activities are targeted [20, 21], gender [22] or the educational context [19] may play a role in the results of the assessment of CT skills, due to the type of conditioning that these factors can represent in the involvement of children and young people in CT-related tasks. These aspects adaptation of assessment instruments, considering the specificities of the context in which the instruments will be used [23].

B. *PeCOT Project*

The PeCOT project aims to test how the use of tangible objects contributes to the development of computational thinking in children, specifically using a robot that has been specially designed to allow children to instruct it to move forward, backward, to the right, to the left and to stop, using a simplified interface. This type of interface, combined with the design of the contexts in which the robot moves, means that the children have to use the concepts associated with the CT to solve the problems posed. Throughout the project, the researchers invite the children to explicitly represent the execution of the robot's steps. Firstly, they need to understand how the robot moves (e.g. to turn left, it must first rotate 90º and only then take a step forward but now in this new direction). Secondly, once they understand how to program, the children can solve problems that will involve the main concepts associated with the CT, namely algorithms, patterns, decomposition and abstraction. The use of the robot in the classroom context, through the challenges it poses, invites children to use these concepts. In order to illustrate the strategies adopted, as well as the variety of activities proposed, for exploring these concepts, the following examples will be presented.

1) *Patterns*

The algorithm competence is worked on the construction of the steps that a robot will have to carry out to achieve a goal. Patterns can be worked on by detecting repeated behaviors and encouraging the use of a more systematic programming language using the cycle/loop concept. One way to promote understanding of these patterns is to ask the robot to perform similar tasks in sequence e.g. the robot has to make two squares, the robot performs a task, and repeats it twice, following a drawn pattern (Fig. 1).

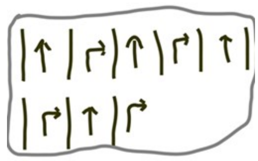

Fig. 1. Pattern to execute a square and the representation of a program executing two squares.

2) *Decomposition and abstraction*

Decomposition and abstraction are discussed in the context of more complex challenges. In the programming language, the concept of stopping has been implemented. In this case, students are asked to divide the tasks into steps and organize them in different sequences. An example of a problem involving decomposition is when the robot has to go to three places to fulfil its mission. Abstraction involves the "disassembly" of a process through a given task for its execution. An example of this is the choice of possible routes taken by the robot using partially hidden code. This strategy forces children to work out a route without being able to execute it because some of the commands are not known.

In parallel with the development of the robot and its associated tasks, the assessment instrument was designed to evaluate some of the skills considered to be fundamental. As mentioned above, this tool, in the form of a worksheet with a set of problem situations to be solved by the students, aims to assess these skills objectively and accurately in a context other than that of using the robot (which will also be assessed), considering that they are transversal to other learning situations and contexts.

In general, the PeCOT project aims to enrich the pedagogical practices of teachers in the Region by offering them diversified strategies for introducing educational robotics into their classrooms, which can complement the work carried out within the framework of the Regional Government's CT Project, which only uses unplugged activities to promote skills in this area, contribute to the development of the children who take part in it, as well as test the use of a robot that can be built in the schools themselves, in a do-it-yourself logic, and that is easy to use for teachers and children.

III. Materials and Methods

The need to adjust an instrument for assessing children's Computational Thinking skills to reality and to the still emerging practice of pedagogical work in this field, led the researchers to design an activity sheet that would allow them to understand the development of these skills in a group of children, who will work with educational robots during the second year of the PeCOT Project. These robots - Azbot-1C - will be introduced into the classroom by the researchers, who will demonstrate how they work, and will be freely explored by the children before being used for educational purposes. Throughout the second year of the project, the robots will be used by teachers and children to pedagogically explore content from different curriculum areas through presentation, analysis and problem solving.

The sample was made up of eight (8) children aged between 8 and 10 years, attending the 3rd grade of a public school on the island of S. Miguel, Azores, in an area considered to be socially and economically disadvantaged.

This activity sheet, made up of 16 problem situations designed to assess 4 CT competences (4 problems for algorithmic thinking, 4 for patterns, 4 for decomposition and 4 for abstraction), was presented by the researchers and completed by the children in the classroom. It was completed by the same children on two separate occasions, each lasting about an hour, in order to anticipate all possible difficulties and errors and to gradually adapt the questions until a final version was obtained that was more appropriate in terms of clarity, execution time and level of difficulty. Each item on the worksheet was read aloud out by one of the researchers as the children solved the problems. If they had any doubts, the statements were read again, and some doubts were clarified. Each child answered his or her activity sheet individually. During the process of solving the activity sheet, the researchers kept a descriptive record of the pupils' difficulties and of any situations that required adjustment. Examples of the 16 problems presented on the sheet will be given in the Results and Discussion section, along with how the children solved them and the difficulties they encountered.

This data and the children's answers were analyzed quantitatively and qualitatively. The children's answers to each of the 16 problems were given 1 point if they were completely correct, 0.5 points if they were partially correct and 0 points if they were incorrect or not answered at all. For problems that required two answers, points were awarded equally for both answers. The results were analyzed statistically using a frequency analysis and the Wilcoxon signed-rank test. At the same time, the descriptive notes made by the researchers were analyzed to understand which aspects of the application of the task should be reconsidered and improved.

All the data collection work was preceded by a request for authorization from the Ethics Committee of the University of the Azores, which gave a positive opinion (Ethics Committee, no. 49/2022), as well as the signing of consent forms by the children and the informed consent of the parents and guardians and the two teachers who collaborated in the research.

IV. Result and Discussion

A. The Construction of the CT Assessment Tool

After reviewing the literature on CT and its assessment, four competences were initially considered when constructing the assessment tool – algorithmic thinking, patterns, decomposition and abstraction, following the logic of PISA 2021, which states that "Computational thinking skills include pattern recognition, designing and using abstraction, pattern decomposition, determining which (if any) computing tools could be employed in analyzing or solving a problem, and defining algorithms as part of a detailed solution" [24] (p.5).

Following the proposal by Csizmadia *et al*. [25], these competences can be clearly and simply defined as: a) algorithmic thinking "is a way of getting to a solution through a clear definition of the steps"; b) patterns "is associated with identifying patterns, similarities and connections, and exploiting those features"; c) decomposition "is a way of thinking about artefacts in terms of their component parts"; and d) abstraction "is the process of making an artefact more understandable through reducing the unnecessary detail" (pp. 7-8).

Four problems were designed for each of these CT skills, based on the Bebras problems used in The Bebras competition, in particular "Bebras Unplugged: Advanced card set" [26] and the Mitarbeit Biber der Informatik [27], as well as the work of Sun *et al*. [28]. The problems used in The Bebras competition were chosen because they had already been validated for children in the same age groups as the children in the sample, they could be carried out without having to buy any equipment and they clearly identified the CT skills that could be assessed in each of them.

Based on the tasks identified as being most appropriate for the age of children for whom the assessment form was intended, tasks were translated and adapted to represent different levels of difficulty, in order to have an instrument that would allow us to assess the skills of children at different developmental levels. This was achieved by including items for children under 8 and over 10 in the form.

Each correct answer scored 1 point, so in total each child could achieve 16 points. In tasks where the children had to give more than one answer, the points were distributed equally between the various expected answers.

B. Analysis of the children's answers and reformulation of the assessment tool

The first four tasks assess the children's algorithmic thinking.

In the first task, the children had to identify which of the sequences presented corresponded to the route the doll had to take to get to school (Fig. 2).

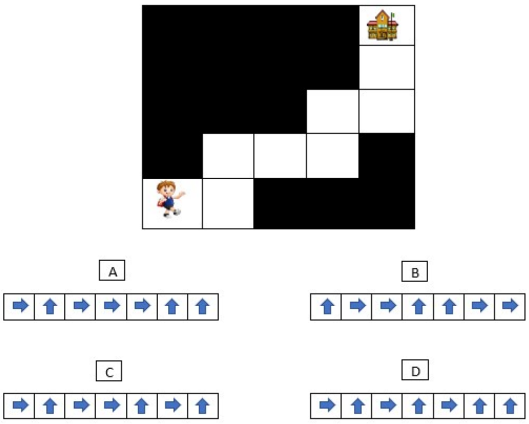

Fig. 2. Algorithm—task 1.

Two aspects of this task were found to influence the children's responses. Firstly, the fact that the possible answers did not include the rotation arrows (left and right), so that the movements shown did not correspond to the real movement of the doll. Secondly, in the answer that was considered correct, the doll didn't enter the school square but stood in front of the door. These difficulties may have had a direct influence on the way the children responded to Task 2, which was very similar in nature. In this second task, the children had to define two possible routes for the doll to get to school, stopping at the fruit shop to buy an apple, and they also had to identify the shortest route. As they followed the logic of exercise 1, many of the children didn't include the rotation arrows and many didn't enter the grocery store and school squares. The strategies used by the children were to draw the arrows in the squares, as shown in Fig. 3.

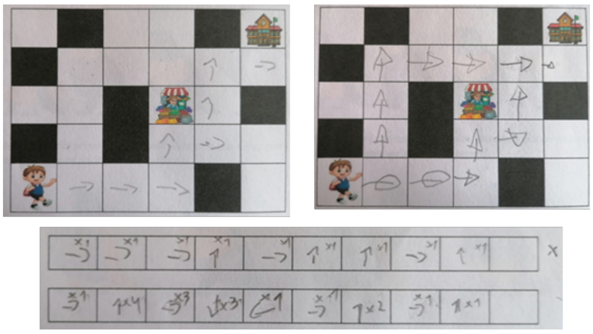

Fig. 3. Algorithm—task 2.

There were some attempts to write according to the code, as explored by the teachers, which indicated the number of times each arrow was used, but the children still didn't include the rotation arrows. In this sense, there are again issues related to the way the questions were asked - the doll must explicitly "enter" the grocery store and the school and the questions in the statement must be placed on two separate lines, as the children only answered one of the questions asked. It is also important to consider the number of squares needed to write the code and to give a name to each set of squares, for example, " route 1" and " route 2" or "A" and "B," to make it easier to answer when more than one answer is required. It could also be improved by replacing the picture of the child with an picture of a robot, as they don't move in the same way, which could make the task easier.

The third and fourth tasks involved sorting a set of Fig. 4 and Fig. 5. In the case of task 3, which is not considered to be a very difficult task, many children confused two of the Figs 3–4, which is why it is considered more appropriate to use colored pictures when it comes to identifying small details. As for task 4, despite some initial difficulties in identifying the order in which the ice cream balls had been placed in the cone (only 3/8 children were able to identify the correct answer), the vast majority were able to complete the task without difficulty on the second attempt (7/8 children). Even so, we considered changing the scenario of the task by placing another object that was more familiar to the children, such as Lego pieces.

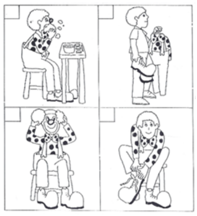

Fig. 4. Algorithm—task 3.

Fig. 5. Algorithm—task 4.

The next four tasks were designed to assess how well the children follow patterns.

The first task asked the children to find out the colors of the last parrot (Fig. 6) and most of the children had no difficulty in solving it.

Fig. 6. Patterns—task 5.

Some of them chose to color the kite as they discovered the solution to the task, so it seems relevant and of an acceptable level of difficulty for children of this age.

The second task related to patterns asked the children to complete the sequence and to indicate in the last rectangle the repetitive pattern that was followed (Fig. 7).

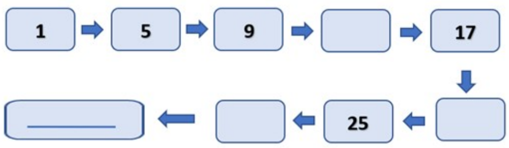

Fig. 7. Patterns—task 6.

Although it wasn't a difficult task, the vast majority of children wrote "33" in the last rectangle instead of the number for the repeating pattern (4), leading the researchers to consider splitting the statement into two parts: one for filling in the blanks and another for the answer to the repeating pattern, with a different graphical presentation to the blue rectangles (e.g. "Answer: _________").

Task 7 asked: What do these pieces have in common? (Fig. 8).

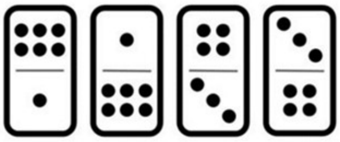

Fig. 8. Patterns—task 7.

In this task, few students answered that the sum of the dots in all pieces was 7. Some answered that the first two pairs and the last two were the same, others that they had the same numbers, others that the pieces were upside down and still others that they had the "little balls" in common. To avoid this dispersion of answers, some of which were unrelated to the aim of the question, it was decided to eliminate the two middle pieces, leaving only the 1st and 4th pieces.

The fourth task related to patterns and asked the children to find out, from a sequence of pictures, which ones were missing at the bottom. This was a very difficult task for the children, and only one managed to solve it on the first attempt. To find the solution, some of the children tried to draw the expected pictures on the squares, but as can be seen in Fig. 9, they found it very difficult to reproduce the figure exactly.

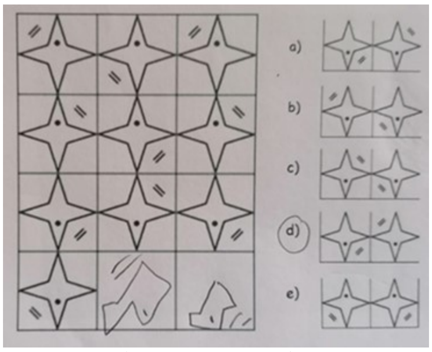

Fig. 9. Patterns—task 8.

In this case, the change would be to use a geometrically simpler figure and a pattern with fewer squares, making it easier to identify and reproduce the pattern.

The next four tasks were related to decomposition. The first asked the children to trace a path within a numbered square, following a set of rules (Fig. 10).

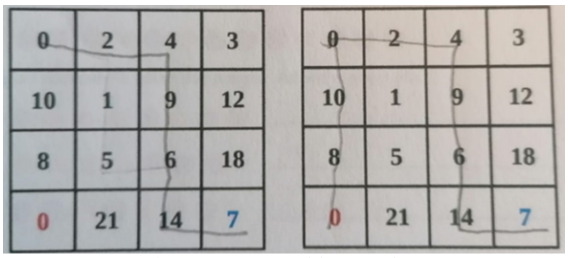

Fig. 10. Decomposition—task 9.

To solve the problem, many students passed the same house twice, some used diagonal lines or ended up drawing the same path for both answers. Whilst there were no major difficulties with the content of this task, the issues relating to the way in which the instructions are given should be reconsidered, highlighting some instructions, such as "you can't pass the same house twice" or including the indication that the path should only be drawn horizontal and vertical movement and not diagonally.

In the second decomposition task, the children had to solve a problem in which they had to distribute a group of beavers in two lifts while respecting the weight limit. None of the students solved this problem correctly. The solution required considering several conditions and making combinations, which the students were unable to do. In the different attempts (Fig. 11), it was observed that the students drew lines to distribute the beavers in the elevators, recorded the weights of the beavers in each elevator, wrote down the number of beavers in each elevator, distributed the pictures of the beavers with a vertical line, excluded those who didn't go in the elevator with an "x," but without success.

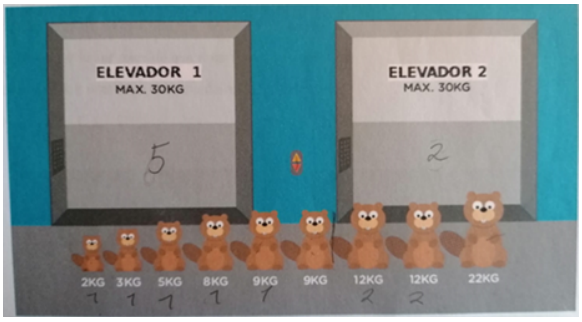

Fig. 11. Decomposition—task 10.

In the next exercise, the children had to write a set of words by deciphering a given code. Most students understood the nature of the task, and most got at least one name right. To make the task easier, the students used the following strategies: drawing vertical lines or circles to delimit the code corresponding to each letter, writing down the letter corresponding to each set of symbols, using lines to separate the letters they found and writing them down, as can be seen in Fig. 12 (A, B).

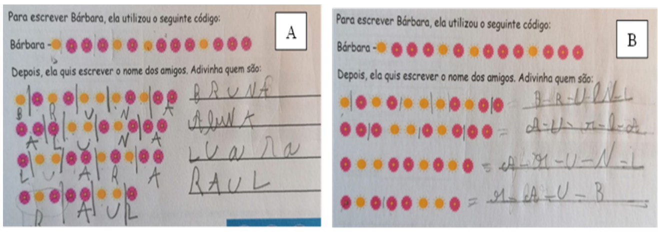

Fig. 12. Decomposition—task 11.

The last decomposition task asked students to find out which picture (Fig. 13) corresponded to the set of clues given in the question.

Being a medium-difficulty task, the main difficulties were related to issues of visualizing the colors of the flowers.

The fourth and final part of the worksheet was designed to assess abstraction.

In the first exercise, the children were asked to indicate the number of rings they had hit on the post and to score them according to the position of the throw (Fig. 14).

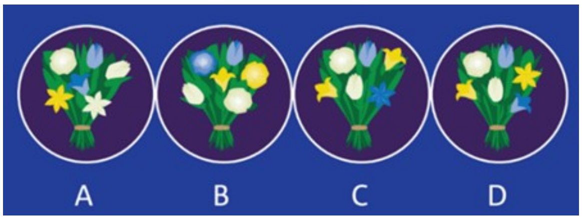

Fig. 13. Decomposition—task 12.


Fig. 14. Abstraction—task 13.

In this task, many children forgot to add up the points and most only hit one of the rings, so it was important to split the questions into two lines, the first relating to the order in which the throws were made, and the second asking for the total score. In terms of complexity, this was not an easy task for the children, with only two managed it in the first moment and five in the second.

The second abstraction task asked the children to identify the order in which the wallpaper shown in the illustration had been placed (Fig. 15).

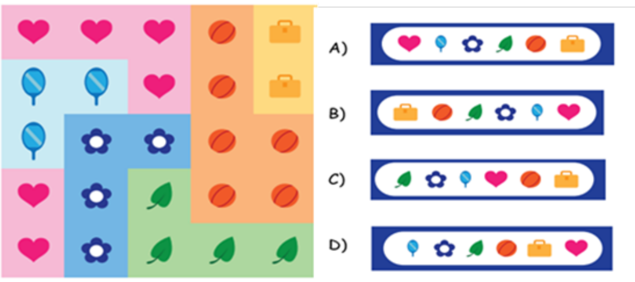

Fig. 15. Abstraction—task 14.

This task did not present any difficulties for the children and was answered correctly by 6/8 of the children in the first moment and by all of them in the second moment.

The third abstraction task asked the students to solve a problem related to the distance the beavers should travel to meet each other with the least possible effort (Fig. 16).

Fig. 16. Abstraction—task 15.

As it was a medium-difficulty task, more than half of the students managed to solve it both times (5/8 children). To solve it, the students decided to draw a line along the route. As this exercise is reminiscent of a subway line, and as the students are probably not familiar with this context, the possibility of using another image, for example a game board, should be considered to bring the task closer to the students' context and to simplify the wording, making it easier to understand.

The last task on the evaluation sheet was for the children to match the beavers with their plates according to a set of clues (Fig. 17). This task did not present any difficulties for the children, and was answered by all of them, despite a few occasional mistakes.

Fig. 17. Abstraction—task 16.

C. *Quantitative results*

In quantitative terms, Table 1 reflects the total score obtained in each of the 16 tasks proposed in the evaluation form, organized according to the CT skills to be evaluated.

Table 1. Results by CT competencies

CT competencies	Questions	1st moment	2nd moment	$\bar{x}$
Algorithmic	1	5	5	
	2	0	0	
	3	2	1	
	4	3	7	
	Total	**10**	**13**	**$\bar{x}$ = 11.5**
Patterns	5	3	4	
	6	5	2.5	
	7	7	8	
	8	1	0	
	Total	**16**	**14.5**	**$\bar{x}$ = 15.25**
Decomposition	9	1	5	
	10	1	1	
	11	3.5	4.75	
	12	3	5	
	Total	**8.5**	**15.8**	**$\bar{x}$ = 12.15**
Abstraction	13	1	2.5	
	14	6	8	
	15	5	3	
	16	5.25	5.75	
	Total	**17.3**	**19.3**	**$\bar{x}$ = 18.3**

The results show that, although there was a slight improvement between the first and second moments of the worksheet, which was to be expected considering the activities that were carried out by the teachers, the abstraction tasks were the ones that showed the best results overall (18.3 mean points), followed by patterns (15.25 mean points), decomposition (12.15 mean points) and, finally, algorithmic skills (11.5 mean points). Considering the two moments of application of the task, the results for decomposition were clearly better in the second moment, going from an initial average of 8.5 points to a final average of 15.8 points, followed by algorithmic and abstraction skills.

Finally, considering that the maximum score that could be obtained was 16 points, Table 2 shows the score obtained by each child in the two moments of completion of the form and shows that in most cases the difference was slight, with an average of 6.46 points in the first moment and 7.81 points in the second moment.

Table 2. Results for each child

Children	1st moment	2nd moment
a	9	11
b	6.25	9
c	8.25	8.25
d	5	5.75
e	6	6.75
f	6.25	7.25
g	6.5	10.5
h	4.5	4
$\bar{x}$	**6.46**	**7.81**

To check for significant differences between the answers before and after, the Wilcoxon signed-rank test was used, a non-parametric test for comparing paired means, corresponding to the same student answering the questionnaire before and after the CT tasks, implemented in the scipy.stats library of Python 3.11.4. The t-test of paired means (or ANOVA tests) is not applicable because the small number of responses does not allow them to follow a normal distribution. In the case of null differences, the zsplit strategy is used, which involves counting the zeros simultaneously for the number of negative and positive differences. The results are shown in Table 3.

The column "Question" refers to the question number, the statistics to the value of the statistic used in the test and the p-value to the probability of significance of the test, signif. * -> significant at 5%, ** -> significant at 1% and *** -> significant at 0.1%.

Table 3. Wilcoxon signed-rank test

Question	Statistics	p-value	signif.
1	14.0	1.000000	-
2	14.0	1.000000	-
3	14.0	1.000000	-
4	5.0	0.118571	-
5	10.5	0.527089	-
6	3.0	0.056438	-
7	10.5	0.527089	-
8	10.5	0.527089	-
9	2.5	0.046875	*
10	14.0	1.000000	-
11	8.5	0.375000	-
12	8.5	0.340174	-
13	8.5	0.340174	-
14	7.5	0.253298	-
15	10.5	0.527089	-
16	14.0	1.000000	-
Total	0.5	0.022254	*

(signif. Column: * -> significant at 5%, ** -> significant at 1% and *** -> significant at 0.1%.)

From the previous results it is possible to verify that both question 9 and the total scores have significant differences at 5%, which is an indicator, especially in the total scores, of some development in the CT skills of the children involved. Therefore, we conclude that this instrument can be relevant for building the reliability in the assessment of the CT skills to be measured.

D. Discussion

Having analyzed the different tasks performed by the children, in relation to the four proposed CT skills – algorithmic thinking, patterns, decomposition and abstraction, some particular aspects deserve our attention when it comes to the construction and use of an instrument for the assessment of younger children, a situation that is not as simple as some studies suggest [17, 18].

First, the length of the assessment instrument. From the first moment of its application, it was clear that it took too long for the children to complete it, due to the number of tasks required, the need for the questions to be read out by the researchers, given the children's lack of reading fluency, as well as the difficulty in solving some of the tasks proposed. Therefore, it was decided to remove one question from each of the four groups: question 1, because it conditioned the subsequent answer; question 8, due to the difficulty of identifying and repeating the pattern; question 10, because of the difficulty of the task and the need to mobilize skills that the children did not have sufficiently developed; and question 13, because of the inherent difficulty of the task itself, combined with the difficulty of reading the picture.

In parallel with the elimination of these tasks, in the remaining tasks, children's familiarity with the content, and the dispersion of answers, are elements to be considered in the reformulation of the remaining items. In terms of form, it is still necessary to clarify some of the questions and instructions, as described above, and to check the way in which answers can be given. It is also necessary to consider aspects related to the quality and type of images to be used. This situation had already been analyzed in a first version of the instrument, which was reformulated by the researchers even before experimenting with children, as it was considered relevant, pertinent, and conditioning the performance of the task.

Another condition for the performance of the task is related to the children's familiarity with the situations presented to them. This is the case with the task relating to ice cream and the organization of flavors in the order in which they are placed (task 4). The fact that there is no way of visualizing this situation in the immediate context of the children's actions can lead to less understanding of the task and its implementation.

The need to answer more than one question in some tasks also proved to be a difficulty for the children, some of whom ended up answering only one part, making it necessary to create specific spaces for the different possible answers that the students had to give.

In addition to the length of the form and the identified questions, the ability of students to understand the content implicit in the tasks to be completed was also assessed, an aspect that does not seem to represent a problem, except for task 10 (decomposition), where only one child managed to reach the solution.

Despite the need to consider all these factors that may affect the way children perform the task, in general there were no major difficulties in mobilizing CT skills. Considering that the instrument was designed to include different levels of difficulty in each group of questions - algorithmic thinking, abstraction, patterns and decomposition - all the children were able to solve several tasks in each of these groups. This, combined with the fact that the proposed tasks have been previously validated on children of the same age in situations of a similar nature [26–28], gives us some degree of confidence in the assessment of this type of competence.

The identification, analysis and restructuring of all these aspects, relating to the nature and content of the tasks proposed in this assessment instrument, are essential for the construction of a tool that is aims to be correct, efficient, simple and usable, as advocated by Beecher [16].

V. Conclusion

In general terms, throughout the construction and validation of the CT evaluation instrument, some aspects were identified that need to be reconsidered and improved for its use in the second phase of the project, in line with the results of research in this area which recommend progressively lead to gradually adapt the assessment tool to be used.

Aspects related to the design of the tasks and the instrument, such as their length, their level of difficulty, the clarity of the problems formulation and the questions to be answered by the children, the images to be used or the multiplication of the number of answers for the same problem situation, as well as aspects related to children's development and learning, such as their familiarity with the content or their reading fluency, must be taken into account when constructing an instrument with the purpose of assessing CT skills in children attending the first levels of schooling.

In accordance with the results presented here, in spite of the need to make some adjustments to the assessment instrument, it can be said that the students managed to mobilize their CP skills in the proposed unplugged activities. It is therefore hoped that they can also mobilized in the second phase of the project through the use of the Azbot-1C robot in pedagogically designed activities to address different contents of the curricular areas.

As a recommendation, it is suggested that, in the future, alternative and complementary instruments can be used that allow a deeper analysis of the data collected, as suggested in studies of the same nature [15, 19], especially considering the nature of this study and its small sample.

Finally, it is hoped that this instrument can be another valid contribution to help teachers and researchers observe and evaluate the computational thinking skills of children of this age, judging by the statistical results obtained between the two moments of its application. Furthermore, in line with what Li *et al*. [17] and Zhang *et al*. [23], It is hoped that this tool will allow us to respond to the specificities of the educational context where the project will be implemented, establishing itself as an instrument designed to adapt to the characteristics of this particular setting.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The conceptualization of the research was developed by Ana Isabel Santos, José Manuel Cascalho and Matthias Funk; the methodology was designed by Ana Isabel Santos, José Manuel Cascalho and Matthias Funk; the implementation of the research and data collection was carried out by Bárbara Amaral, Ana Isabel Santos, and José Manuel Cascalho; the formal analysis was done by Bárbara Amaral, Francisco Marques, Ana Isabel Santos, José Manuel Cascalho and Armando Mendes; obtaining the necessary resources was ensured by Matthias Funk and José Manuel Cascalho; the writing of the original draft of the paper was accomplished by Ana Isabel Santos; review and editing was done by José Manuel Cascalho, Armando Mendes, and Paulo Medeiros; all authors contributed to the writing of the article, read and approved the final version of the manuscript.

Funding

This research was funded by DRCT - Direção Regional de Ciência e Tecnologia da Região Autónoma dos Açores (Regional Directorate for Science and Technology of the Autonomous Region of the Azores), grant number M1.1.C/C.S./058/2021/01.

References

[1] N. O. Ezeamuzie and J. S. C. Leung, "Computational thinking through an empirical lens: A systematic review of literature," *Journal of Educational Computing Research*, vol. 60, no. 2, pp. 481–511, 2022. DOI: 10.1177/07356331211033158

[2] C. C. Selbi and J. Woollard, "Computational thinking: The developing definition," *SIGCSE 2014*, pp. 5-8, March 2014.

[3] F. Xu and S. Zhang, "Understanding the source of confusion with computational thinking: A systematic review of definitions," in *Proc. 2021 IEEE Integrated STEM Education Conference (ISEC)*, DOI: 10.1109/ISEC52395.2021.9764144

[4] A. M. L. Carreira, "Computational thinking at the service of cross curricular integration in the elementary school," *RE@D - Revista de Educação a Distância e Elearning*, vol. 5, no.1, 2022. https://doi.org/10.34627/redvol5iss1e202202

[5] R. E. Fantinati and S. S. Rosa, "Computational Thinking: Skills, Strategies and challenges in Basic Education," *Informática na Educação: Teoria & Prática*, vol. 24, no. 1, pp. 129–141, 2021. https://doi.org/10.22456/1982-1654.110751

[6] J. Dobgenski, A. F. G. Silva and F. D. Mazzarino, "Knowledge about computational thinking: Noticing from the reflective teaching practice," *JIEEM*, vol. 15, no. 1, pp. 27–35, 2022. DOI:10.17921/2176-5634.2022v15n1p27-35

[7] D. Alyahya and A. Alotaibi, "Computational thinking skills and its impact on TIMSS achievement: An instructional design approach," *Issues and Trends in Educational Technology*, vol. 7, no. 1, 2019. https://doi.org/10.2458/azu_itet_v7i1_alyahya

[8] C.-Y. Chang, Z. Du, H.-C. Kuo and C.-C. Chang, "Investigating the impact of design thinking-based STEAM PBL on students' creativity and computational thinking," *IEEE Transactions on Education*, vol. 66, no. 6, pp. 673–681, 2023. DOI: 10.1109/TE.2023.3297221

[9] L. E. Hamamsy, B. Bruno, C. Audrin, M. Chevalier, S. Avry, J. D. Zuferey, and F. Mondada, "How are primary school computer science curricular reforms contributing to equity? Impact on student learning, perception of the discipline, and gender gaps," *International Journal of STEM Education*, vol. 10, no. 60, 2023. https://doi.org/10.1186/s40594-023-00438-3

[10] A. Misirli and V. Komis, "Computational thinking in early childhood education: The impact of programming a tangible robot on developing debugging knowledge," *Early Childhood Research Quarterly*, vol. 65, pp. 139–158, 2023. https://doi.org/10.1016/j.ecresq.2023.05.014

[11] D. S. Rana, S. C. Dimri, P. Malik and S. A. Dhondiyal, "Impact of computational thinking in engineering and K12 education," in *Proc. the 4th International Conference on Inventive Research in Computing Applications*, ICIRCA 2022. DOI: 10.1109/ICIRCA54612.2022.9985593

[12] Regional Secretariat for Education and Cultural Affairs, "Referencial Pedagógico - Pensamento Computacional nos 1.º e 2.º Ciclos do Ensino Básico. Direção Regional da Educação e da Administração

Educativa" (Pedagogical Framework - Computational Thinking in the 1st and 2nd Cycles of Basic Education. Regional Directorate of Education and Educational Administration), 2022.

[13] J. M. Wing, "Computational thinking," *Communications of the ACM*, vol. 49, no. 3, pp. 33–35, 2006.

[14] Y. Allsop, "Assessing computational thinking process using a multiple evaluation approach," *International Journal of Child-Computer Interaction*, vol. 9, pp. 30–55, 2019. https://doi.org/10.1016/j.ijcci.2018.10.004

[15] B. D. Gane, M. Israel, N. Elagha, W. Yan, F. Luo, and J.W. Pellegrino, "Design and validation of learning trajectory-based assessments for computational thinking in upper elementary grades," *Computer Science Education*, vol. 31, no. 2, pp. 141–168, 2021. https://doi.org/10.1080/08993408.2021.1874221

[16] K. Beecher, *Computational Thinking. A Beginner's Guide to Problem Solving and Programming*, BCS Learning & Development, Swindon, UK, 2017.

[17] Y. Li, S. Xu and J. Liu, "Development and validation of computational thinking assessment of chinese elementary school students," *Journal of Pacific Rim Psychology*, vol. 15, pp. 1–22, 2021. https://doi.org/10.1177/18344909211010240

[18] V. V. V. Varghese and V. G. Renumol, "Assessment methods and interventions to develop computational thinking—A literature review," in *Proc. 2021 International Conference on Innovative Trends in Information Technology (ICITIIT)*. DOI: 10.1109/ICITIIT51526.2021.9399606

[19] E. Çoban and Ö. Korkmaz, "An alternative approach for measuring computational thinking: Performance-based platform," *Thinking Skills and Creativity*, vol. 42, 2021. https://doi.org/10.1016/j.tsc.2021.100929

[20] M. U. Bers, "Coding, playgrounds and literacy in early childhood education: The development of KIBO robotics and ScratchJr," presented at 2018 IEEE Global Engineering Education Conference (EDUCON), 2100–2108, Santa Cruz de Tenerife, Canary Islands, Spain.

[21] E. Relkin, L. Ruiter and M. U. Bers, "TechCheck: Development and validation of an unplugged assessment of computational thinking in early childhood education," *Journal of Science Education and Technology*, vol. 29, pp. 482–498, 2020. https://doi.org/10.1007/s10956-020-09831-x

[22] L. Z. Sauer, C. E. R. D. Reis, G. Dall'Acua, I. G. D. Lima, O. Giovannini, and V. Villas-Boas, "Work-in-progress: Encouraging girls in science, engineering and information technology," in *Proc. 2020 IEEE Global Engineering Education Conference (EDUCON)*, Porto, Portugal, pp. 28–32. doi: 10.1109/EDUCON45650.2020.9125310

[23] S. Zhang, G. K. W. Wong and G. Pan, "Computational thinking test for lower primary students: Design principles, content validation, and pilot testing," in *Proc. 2021 IEEE International Conference on Engineering, Technology & Education (TALE)*. DOI: 10.1109/TALE52509.2021.9678852

[24] OECD. (2022). *PISA 2021 Mathematics Framework*. [Online]. Available: https://www.oecd.org/pisa/publications/pisa-2021-assessment-and-analytical-framework.htm

[25] A. Csizmadia, P. Curzon, M. Dorling, S. Humphreys, T. Ng, C. Selby, and J. Woollard, *Computational Thinking. A Guide for Teachers*, 2015.

[26] B. Unplugged. *Advanced Card Set*. [Online]. Available: https://www.amt.edu.au/bebras-unplugged

[27] E. Maier-Gabriel, "Mitarbeit Biber der Informatik 2014," *Osterreichische Computer Gesellschaft (OCG)*, gemeinn utziger Verein. Biber der Informatik, [Cooperation Biber der Informatik 2014,' Osterreichische Computer Gesellschaft (OCG), non-profit organisation. Beaver of Computer Science], Austria, 2014.

[28] L. Sun, L. Hu and D. Zhou, "Improving 7th-graders' computational thinking skills through unplugged programming activities: A study on the influence of multiple factors," *Thinking Skills and Creativity*, vol. 42, 100926, 2021. https://doi.org/10.1016/j.tsc.2021.100926