Computational Thinking Test for Beginners: Design and Content Validation

María Zapata-Cáceres
Computer Science Department
Universidad Rey Juan Carlos
Madrid, Spain
maria.zapata@urjc.es

Estefanía Martín-Barroso
Computer Science Department
Universidad Rey Juan Carlos
Madrid, Spain
estefania.martin@urjc.es

Marcos Román-González
Faculty of Education
UNED
Madrid, Spain
mroman@edu.uned.es

***Abstract*— Computational Thinking (CT) is a fundamental skill that is not only confined to computer scientists' activities but can be widely applied in daily life and is required in order to adapt to the future and, therefore, should be taught at early ages. Within this framework, assessing CT is an indispensable part to consider in order to introduce CT in the school curricula. Nevertheless, efforts involving the formal assessment of computational thinking has primarily focused on middle school grades and above; and are mostly based on the analysis of projects in specific programming environments. A Beginners Computational Thinking Test (BCTt), aimed at early ages, and based on the Computational Thinking Test [1], has been designed including several improvements; submitted to a content validation process through expert´s judgement procedure; and administered to Primary School students. The BCTt design is considered adequate by experts and results show a high reliability for the assessment of CT in Primary School, particularly in first educational stages.**

Keywords—Beginners Computational Thinking Test, Computational Thinking, Computer Science Education, Primary Education, Assessment, Programming, Early Childhood Education

I. Introduction

Computational Thinking (CT) is a fundamental skill that can be widely applied in daily life and is required in order to adapt to 21st century society [2]. CT was first defined as a human problem-solving process that uses decomposition and requires thinking at multiple levels of abstraction; it is not only the center of problem solving, but also develops and identifies the problem [3]. Subsequently, many other definitions had arisen, and this has provoked broad debate. It can be defined as the conceptual foundation required to solve problems effectively and efficiently with solutions that can be used in different contexts [4]. Likewise, it can be defined as the thinking skills that precede coding and programming, and are applied in understanding a problem and formulating a solution like a computer scientist [5].

Although there are potential risks related to this lack of consensus about CT definition [6], CT is considered an essential skill that new generations of students must acquire and, therefore, should be taught at schools [7-9]. In addition, there is evidence that programming exposes students to CT and, therefore, to problem-solving using computer science concepts such as abstraction and decomposition [10].

Consequently, there are no agreed-upon models or frameworks for developing CT in the classroom [4], e.g. Wing includes five cognitive processes in CT: problem reformulation, recursion, problem decomposition, abstraction and systematic testing [11]; CT could also be divided into five facets: abstraction, generalization, algorithm, modularity, and decomposition [12]; or categorized into the following skills: abstraction, decomposition, algorithms, debugging, iteration and generalization [4]. Brennan and Resnick propose a three-dimensional (3D) framework for CT [13]. This framework, which has attracted many researchers' attention and been cited frequently in literature in recent years [14], categorizes CT into three areas: (a) computational concepts (concepts that programmers use, such as sequences or loops), (b) computational practices (problem-solving practices that occurs in the process of programming, such as iteration or abstraction), and (c) computational perspectives (the perspectives that designers form about the world around them and about themselves, such as expressing or connecting).

Along with learning CT, assessing CT is an indispensable part to consider in order to introduce CT in the curriculum, as student evaluation for pedagogical purposes is essential [15]. Unfortunately, while there are multiple researchers that describes experiences in integrating computational thinking into the K-12 curriculum, efforts involving the formal assessment of computational thinking has primarily focused on middle school grades and above [16]. Moreover, the assessment instruments proposed by recent research are mostly based on the analysis of projects performed by students in specific programming environments.

In this way, there are some attempts to measure and assess CT in young students such as Fairy assessment in Alice [17], which measures CT aspects in an specific programming learning environment (Alice); or Computational Thinking Pattern Quiz instrument [18], to assess whether computational thinking patterns can be recognized in a non-programming context analyzing CT Patterns during the creation of a videogame with AgentSheets environment. Moreover, the Test for Measuring Basic Programming Abilities [19] and Commutative Assessment [20], are both validated instruments under a psychometric approach but aimed to middle/high school students.

Similarly, Franklin et al. propose a model for integrating CT assessment into the design of a Scratch-based curriculum [21] and a small pilot test with middle school students show positive results; Denner, Werner and Ortiz developed a coding scheme to identify the extent to which programs written by middle schools girls corresponded with computer science programming concepts [22]; Moreno León and Robles presented Dr.Scratch that analyses Scratch projects an can be used as a tool for the formative assessment of Scratch projects [23]. Seiter and Foreman introduce the Progression of Early Computational Thinking (PECT) model, which is a framework for understanding and assessing CT in Primary School (grades 1 to 6), analyzing coding design patterns in student programming projects [16].

Furthermore, Román et al. developed the Computational Thinking Test (CTt), which stands out as a self-contained instrument, independent of any programming environment, for the assessment of CT [1], which is designed under a psychometric approach and provides evidence about its reliability and content [24], criterion [25], and predictive validity [26]; it is consistent with [19] and [20]; and aligned with the international standards for psychological and educational testing [27]. In terms of Brennan 3D framework, the Román et al. CTt focuses on computational concepts, partially on computational practices and ignores computational perspectives [1].

Even though the CTt is aimed to students between 10 to 16 years old, it has been a consolidated and firm basis for the design of a Primary School targeted test: Beginners Computational Thinking Test (BCTt) that has been developed, validated and administered to Primary School students in this study, as most previous studies were limited to CT assessment on middle school grades and above. As the BCTt target population is younger than that on CTt, the test must be adapted both in form and content. Moreover, the BCTt design includes several innovations which intended to be substantial improvements. In this paper we present the guidelines that have been followed for the design of the BCTt as a stand-alone assessment instrument, independent of any programming environment; its content validation process, and some preliminary statistical analysis that show the promising consistency of the test to assess CT in Primary School.

II. METHOD

An initial test version (BCTt v.1) was designed and then submitted to a content validation process through expert´s judgment procedure. Next, attending to the results, suggestions and problems encountered, the test was improved, obtaining a second and more robust version (BCTt v.2). Finally, the test was administered to 299 Primary School students from schools in Spain to perform a statistical item analysis.

A. *Beginners Computational Thinking Test v.1*

As the BCTt target population (5 to 12 years old) is younger than that on the CTt, it must be adapted both in form and content. Moreover, the BCTt includes several innovations which intended to be substantial improvements. Pilot tests were carried out on small subsamples (n=3 to 5 subjects, 5 to 10 years old) throughout initial design.

This initial version of the BCTt is 25 items long, with an estimated time of 40 minutes, which seemed to be adequate in pilot tests. Items are designed with the least possible text, and symbols are intended to be self-explanatory in order to increase the readability of the test at early ages.

Considering that the target population has lower reading, writing, and overall skills, decisions taken are aimed at making the test easier and accessible for young people. BCTt graphic aspect is clear and intuitive and, to ease association, the symbols used are intended to connect emotionally with the students, since emotions take central stage among the factors that influence the success of the learning process [28]. In this way, the main challenge posed is to carry a chicken along to its mother (the hen).

BCTt v.1 is multiple choice type, with three response alternatives for each item, which are set out in two different graphic layouts: canvas or maze type. The canvas type is a "follow the dotted line" design that children of these educational stages are used to, from their every day school work. The maze layout consists of a square matrix where the student must figure a path in order to reach a target or solve a problem, passing from one square to another in a particular order. In this case, visual transitions were added between squares (Fig. 1). This is intended to be a substantial improvement in maze layouts as our hypothesis is that difficulties with this type of layouts at early ages are related with disorientation and hesitations about whether the current and target square, at each step, should be part of the path sequence or ignored. Besides, adding transitions turn the maze in a state diagram, a main item in algorithms and coding which has proved to improve the capability to understand problems [29-31].

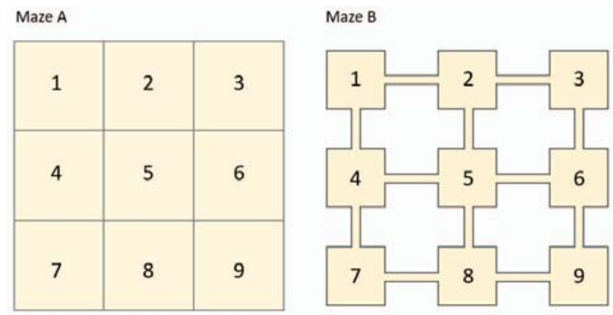

Fig. 1. Maze A: no transitions; Maze B: transitions are added between squares turning the maze into a state diagram

BCTt v.1 response alternatives are laid out as sequences of thick arrows, numbers and colors, depending on the computational concept involved in each question. Each answer has a top-bottom vertical layout, and not horizontal from left to right as in the CTt. This decision was taken considering that code sequences reading direction is top-bottom. Besides, top-bottom layout revealed to be an adequate arrangement in pilot tests, although problems related to canvas item layout were encountered: if the dotted line was to be drawn from bottom to top, students tended to read the answers from bottom to top (Fig. 2). This problem was solved avoiding these drawing directions (Fig. 3).

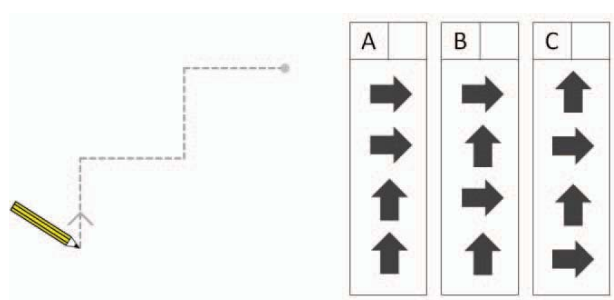

Fig. 2. Canvas type item. Drawing direction creates confusion about the answers reading direction.

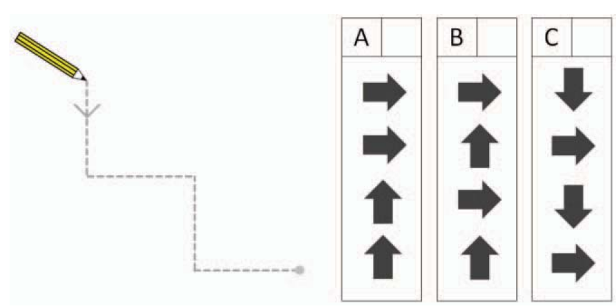

Fig. 3. Corrected canvas type item. Drawing direction is the same as the answers reading direction.

BCTt v.1 contains Brennan's 3D framework basic computational concepts, ordered in increasing difficulty, according to the target educational stages (Table 1): sequences (6 items), simple loops (5 items), nested loops (7 items) and conditionals (7 items). In each maze item, the required task is to carry a chicken along to its mother (the hen) through the maze that could be small or large format, allowing challenges of different complexity. There could be obstacles to avoid (a cat) or objects to collect along the way (pick-ups), such as another chicken (Figs. 4 to 7).

TABLE I. 3D FRAMEWORK COMPUTATIONAL CONCEPTS CONSIDERED IN EACH BCTT V.1 ITEM

Item	Interface	Size	Obstacles	Pick-ups	Computational concept					
						Loops		*Conditionals*		
					1.Sequences	*2.Simple*	*3.Nested*	*4.If-Then*	*5.If-Then-Else*	*6.While*
1	Maze	Small			x					
2	Canvas		-	-	x					
3	Maze	Small	x		x					
4	Maze	Small	x	x	x					
5	Maze	Large	x	x	x					
6	Canvas		-	-	x					
7	Maze	Small				x				
8	Maze	Small				x				
9	Maze	Small	x			x				
10	Maze	Large				x				
11	Maze	Large		x		x				
12	Maze	Large					x			
13	Canvas		-	-			x			
14	Maze	Large	x				x			
15	Maze	Large		x			x			
16	Maze	Large		x			x			
17	Canvas		-	-			x			
18	Maze	Large	x	x			x			
19	Maze	Small						x		
20	Maze	Large						x		
21	Maze	Large							x	
22	Maze	Large							x	
23	Maze	Small								x
24	Maze	Large								x
25	Maze	Large								x

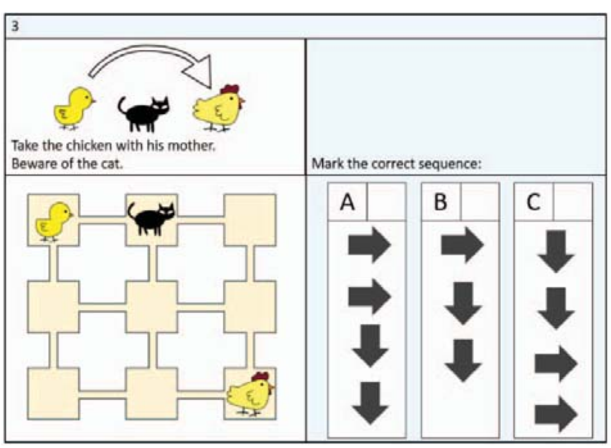

Fig. 4. BCTt v.1 item example (item number 3).

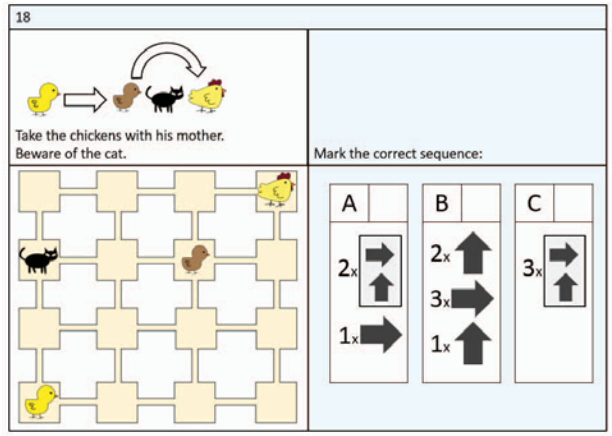

Fig. 5. BCTt v.1 item example (item number 18).

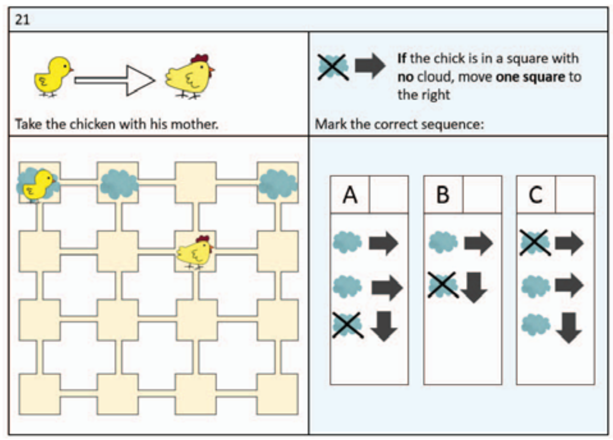

Fig. 6. BCTt v.1 item example (item number 21).

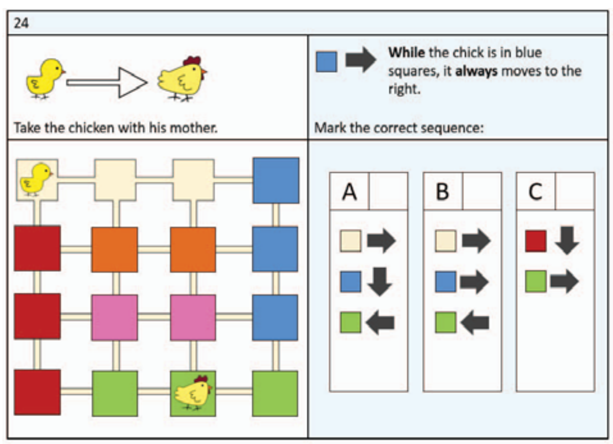

Fig. 7. BCTt v.1 item example (item number 24).

B. *Expert's judgement procedure (BCTt v.1)*

A content validation process of BCTt v.1 was completed through expert judgment procedure, where 45 experts of different profiles (Table II) provided their validation of the instrument, estimating the difficulty level and relevance to measure CT of each item; and contributing with other considerations such as test length and graphic interface adequacy or improvements applicability. Data was collected by a 66 item long on-line form (http://bit.ly/38sEc8B) resumed in Table III.

TABLE II. EXPERT'S PROFILES

Professional group / groups (multiple response allowed)	Number of experts
Computer Science Professional	21
Computer Science Teacher	14
Primary School Teacher	9
Preschool Teacher	1
University Teacher	9
No answer	4
Age	
Less than 30	3
From 31 to 50	37
More than 51	1
No answer	4
Gender	
Woman	13
Man	28
No answer	4
Expertise level in computer science teaching methodologies	
Very low	8
Low	2
Average	9
High	13
Very High	9
No answer	4

TABLE III. EXPERT JUDGEMENTS FORM DESCRIPTION

Experts' judgement Form Items #	BCTt topic addressed	Valued issue by the judges	Form items type
1 to 12	Sequences	Experts' answers to the corresponding items from the BCTt and their perceived difficulty level	Multiple choice + Likert scale
13	Sequences	Relevance to measure CT	Likert scale
14 to 23	Simple loops	Experts' answers to the corresponding items from the BCTt and their perceived difficulty level	Multiple choice + Likert scale
24	Simple loops	Relevance to measure CT	Likert scale
25 to 38	Nested loops	Experts' answers to the corresponding items from the BCTt and their perceived difficulty level	Multiple choice + Likert scale
39	Nested loops	Relevance to measure CT	Likert scale
40 to 43	If-then	Experts' answers to the corresponding items from the BCTt and their perceived difficulty level	Multiple choice + Likert scale
44	If-then	Relevance to measure CT	Likert scale
45 to 48	If-then-else	Experts' answers to the corresponding items from the BCTt and their perceived difficulty level	Multiple choice + Likert scale
49	If-then-else	Relevance to measure CT	Likert scale
50 to 55	While	Experts' answers to the corresponding items from the BCTt and their perceived difficulty level	Multiple choice + Likert scale
56	While	Relevance to measure CT	Likert scale
57 to 60	Personal data	Expert profile data	Text
61	Transitions	Preference between maze with and without transitions	Dichotomous
62	Transitions	Item 61 answer justification	Text
63	Test length	Valuation on the length of the BCTt (v.1)	Likert scale
64	Test adequacy	Valuation on the content of the BCTt (v.1): CT in Primary School	Likert scale
65	Test Interface	Valuation on graphic design and UX aspects of the BCTt (v.1)	Likert scale
66	Test Overall	Final comments and suggestions	Text

C. BCTt administration: participants and procedure (v.2)

According to the content validation process results and experts' suggestions (see section III.A), the BCTt was modified into a refined final version: BCTt v.2 (see section III.B) and administered to Primary School students to perform an item statistical analysis and to assess its design adequacy.

The participants in this study were a sample of n=299 Primary School students (5 to 12 years old) from three Spanish schools. In each school, the research focused on one educational stage as shown in Table IV. The sampling procedure is intentional and, depending on the reasons that led to sample the different subjects, these can be divided as shown in Table V.

BCTt v.2, with added transitions between squares in maze layouts, was administered to A1, B1, C1, D1, E1, and F1 subsamples. BCTt variation, with no transitions between maze squares, was administered to B2, D2 and F2 subsamples. Moreover, BCTt was re-administered to D1 subsample subjects 5 weeks later.

The research was performed under the same conditions in each school as an action protocol was followed. The tests were administered concurrently to every subject. In order to ensure that students skills or previous experience in the use of computer devices do not interfere with the results, the tests were printed and filled by the students individually in paper form. Moreover, tests were printed in greyscale, so that they were accessible to students with color blindness (see section III.B). Before taking the test, an explanatory example of an item from each of the 6 different computational concepts was performed orally in front of the students.

TABLE IV. PRIMARY SCHOOL EDUCATIONAL STAGE CONSIDERED IN EACH SCHOOL

School	Educational stage	Grades	Students ages
Colegio Público Carlos Ruiz	1st	1st and 2nd	5 - 8
Colegio Los Escolapios	2nd	3rd and 4th	7 -10
CEIP León Felipe	3rd	5th and 6th	9 -12

TABLE V. NUMBER OF STUDENTS (N) IN EACH SUBSAMPLE

Educational stage	Grade	Identifier	BCTt	BCTt variation
1st	1	A	A1: n=52	
	2	B	B1: n=18	B2: n=18
2nd	4	C	C1: n=54	
	4	D	D1: n=28	D2: n=28
3rd	5	E	E1: n=51	
	6	F	F1: n=25	F2: n=25

III. Results and Discussion

A. Expert's content validation (BCTt v.1)

Regarding BCTt v.1 length, 24.4% of the experts consulted estimated that the test contains too many questions; a 68.3% considered the test length adequate. Just a 7.3% would add more questions.

It was concluded that BCTt v.1 has a growing perceived difficulty along its items (Mean=2.8; Std. Deviation=0.83) (Fig. 8), which is consistent with the scores obtained by the experts since scores decrease throughout the test: splitting BCTt items on computational concepts sets and counting how many experts answered correctly the items of each set (Mean=37.6; Std. Deviation=2.57) (Fig. 9). Relevance for measuring CT grows similarly along the test items (Mean=3.96; Std. Deviation=0.19) (Fig. 10) and each computational concept have a medium or high perceived relevance (Likert scale from 1 to 5: 5 maximum relevance): sequences were perceived as the least relevant computational concept (3.66) and nested loops the most relevant computational concept (4.14).

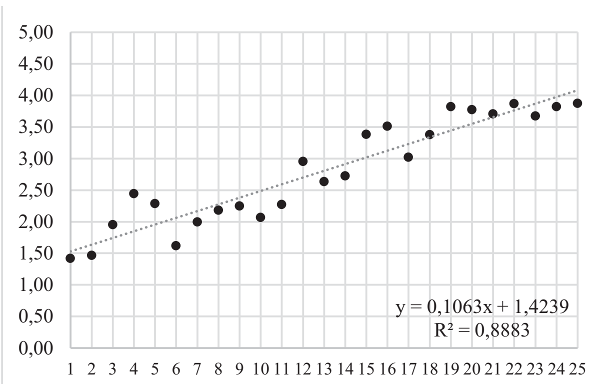

Fig. 8. BCTt item difficulty perceived by experts (ordinate axis: Likert scale from 1 to 5), per BCTt item (abscissa axis).

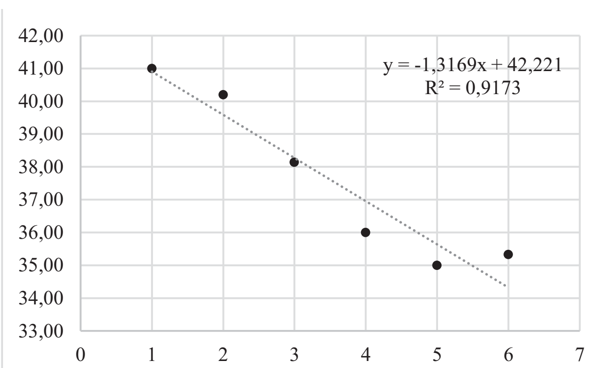

Fig. 9. Ordinate axis: BCTt score obtained (e.g. 41 means that 41 of 45 experts answer correctly); per computational concept (Abscissa axis: 1. Sequences, 2. Simple loop, 3. Nested loop, 4. If-then, 5. If-then-else, 6. While).

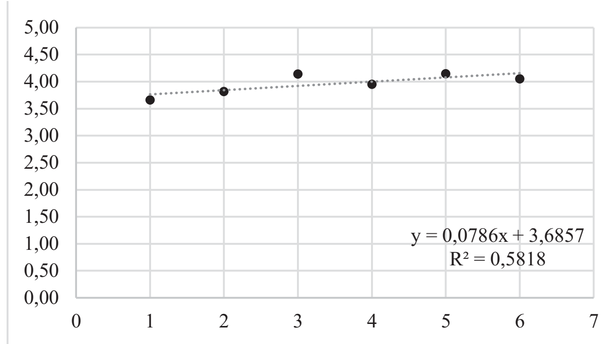

Fig. 10. BCTt computational concept relevance to measure CT, perceived by experts (ordinate axis: Likert scale from 1 to 5) by computational concept (Abscissa axis: 1. Sequences, 2. Simple loop, 3. Nested loop, 4. If-then, 5. If-then-else, 6. While)

Moreover, under the question: "What is the BCTt global level of adequacy to evaluate CT in Primary School students" (Likert scale from 1 to 5), 73.1% considered a good or very good adequacy (34.1%: very good, 39%: good, 22%: intermediate, 4.9%: bad, 0%: very bad).

With regard to the question concerning interface and graphic style adequacy to Primary School Students (Likert scale from 1 to 5), 75.6% considered a good or very good adequacy (39%: very good, 36.6%: good, 22%: intermediate, 2.4%: bad, 0%: very bad).

Furthermore, 83% of the experts estimated that the addition of transitions to the maze layout is positive and considered it a clear improvement to facilitate the understanding of the problems posed. Some answers collected were: "transitions are easily associated to arrows in the answers", "transitions incorporate edges to the mazes, as a state diagram, with a better understanding of the problem", "the allowed paths are clear, excluding diagonal movements", "by including transitions, a distinction is clearly made between the movement and the place of arrival. In the design without transitions, doubts are generated about when a character reaches another (either when it reaches the previous square or when it reaches the other character square?)".

Finally, experts made many suggestions and comments that were carefully considered to improve BCTt to a more robust second version. Some relevant recurring comments were:

- "An explanatory oral example of every different type of item is needed".
- "Regarding the number of response alternatives, I suggest 4 alternatives instead of 3, as only 3 response alternatives per item can negatively influence the total reliability of the test".
- "Children could try to jump the cat: an express indication of not touching the cat is needed".
- "It is not clear if two chicks can move together after meeting".
- "Last questions (conditionals) need more explanation".
- "If-else and if-then-else items do not correspond exactly to the computational concept".

Six experts of Primary School teacher's professional group (9 experts) consider the BCTt v.1 too difficult for Primary School students.

B. Final BCTt version (BCTt v.2)

According to the validation process results and experts' suggestions, the BCTt v.1 was modified, both in form and content, into a refined final version (BCTt v.2). The following features, among other, were modified or added:

- Before taking the test, an explanatory example of an item from each of the 6 different computational concepts must be performed orally in front of the students.
- Each item contains 4 alternative responses instead of 3 to reduce the probability of responding correctly at random (e.g. Fig. 11).
- In items that contain a cat to avoid, it is specified that the square occupied by the cat is not crossed (e.g. Fig. 11).
- In items that include another chicken, it could be ambiguous whether the two chickens should continue together after they meet or not, so in BCTt v.2, the other chicken is replaced by a flower to collect (e.g. Fig. 12).
- The examples of meaning contained in the last items are clearer (e.g. Fig. 13).
- If-else and if-then-else items are reformulated for a better correspondence with the computational concept (e.g. Fig. 13).
- To ensure that students with color blindness can read the symbols on each item, a specific shape is associated to each different color (e.g. triangle and blue). Moreover, this improvement allows the test to be printed in black and white format (e.g. Fig. 14).

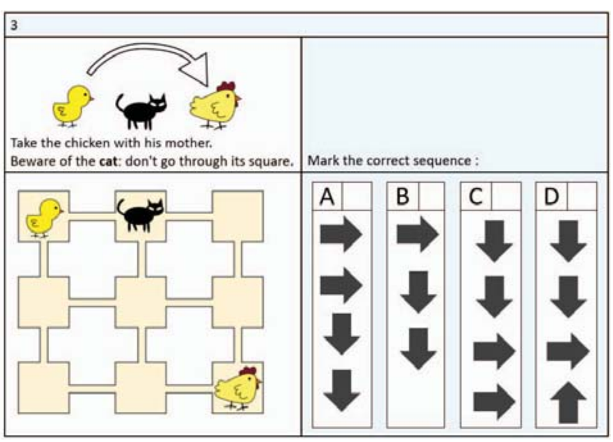

Fig. 11. BCTt v.2 item example (item number 3).

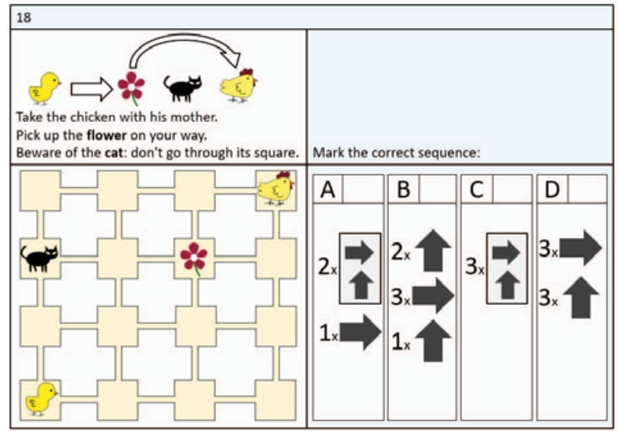

Fig. 12. BCTt v.2 item example (item number 18)

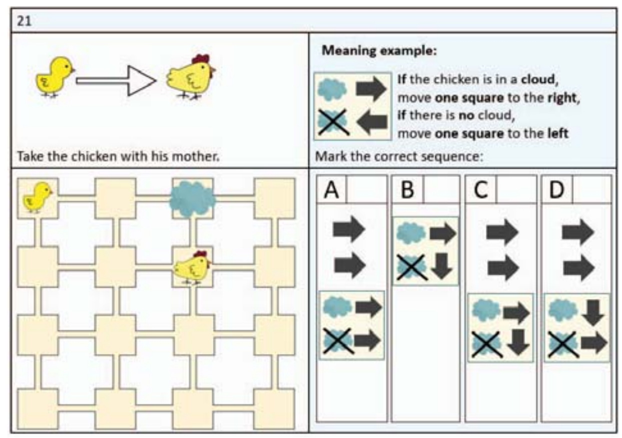

Fig. 13. BCTt v.2 item example (item number 21)

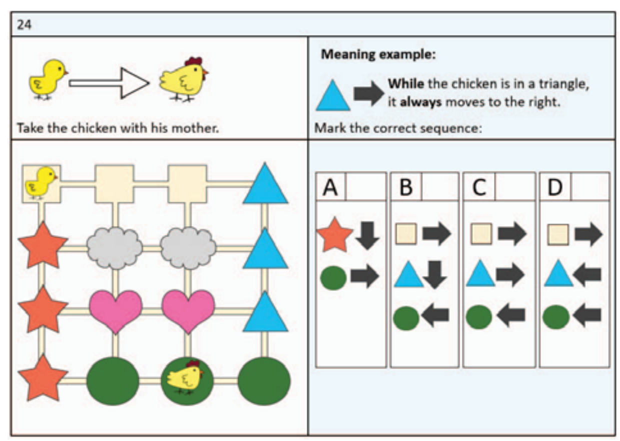

Fig. 14. BCTt v.2 item example (item number 24)

C. BCTt administration results: statistical analysis

BCTt v.2 (refined final version) was administered to Primary School students to empirically analyze design adequacy, perform an item statistical analysis, and test reliability.

1) Transitions

Considering BCTt score as the sum of correct answers along the 25 items of each student's test, to evaluate transitions relevance and effectiveness, the BCTt scores were compared to the BCTt without transitions scores, between same grade subsamples (B1 and B2; D1 and D2, F1 and F2), with Student's t-test, assuming equal variances (Levene's Test). As can be seen in Table VI, there aren't significant differences in 4th and 6th grades test scores, yet there is a significant difference in test scores (p=0.005 < 0.01) between 2nd grade subsamples. It can be concluded that only younger students benefit from the addition of transitions in maze layouts.

TABLE VI. SUBSAMPLES STATISTICS AND STUDENT´S T-TEST COMPARING BCTT WITH AND WITHOUT TRANSITIONS

Grade	Sub-sample	BCTt version	N	Mean	Std. Dev.	t-test for Equality of Means	
						t	*Sig.*
2	B1	with transitions	18	16.778	2.487	3.042	0.005
	B2	without transitions	18	14.278	2.445		
4	D1	with transitions	28	21.357	2.438	0.122	0.904
	D2	without transitions	28	21.286	1.922		
6	F1	with transitions	25	21.720	2.622	0.499	0.620
	F2	without transitions	25	21.280	3.542		

2) Item analysis

Considering BCTt score as the sum of correct answers along the 25 items of the test of each student and seeking a balance between number of subjects and educational stages, a statistical analysis of BCTt score results was performed on a subsample of each grade: A1, B1, C1, E1 and F1 subsamples (N=200). Results along grades are shown in Table VII. Entire sample preliminary analysis of the results reveals an overall high mean (19.915), and scores along subsamples suggest that the test might be aimed at Primary School first educational stages, as no significant difference is shown between 4th and 5th grades means (Student's t-test: t=0.485, p=0.63 > 0.05) nor between 5th and 6th grades means (Student's t-test: t=0.193; p=0.85 > 0.05).

Nevertheless, a second deeper analysis was made, this time, splitting BCTt items on computational concepts, and counting how many students answered correctly the items of each set (i.e., if the subsample is n=23, 21 points score in an item means that 21 of the 23 students answered it correctly). This score per item, related to each computational concept, shows interesting results, since items scores related to Nested Loops and Conditionals concepts reveal low success along every grade as is shown in Fig. 15, but Sequences and Simple loops show very high success in 5th and 6th grades, which leads us to conclude that the initial test items could be too easy for 3rd Primary School stage students, so it could be necessary to add to the test more difficult items for this stage.

TABLE VII. BCTT SCORE STATISTICS BY GRADE

Sample		Entire sample	A1	B1	C1	E1	F1
Grade		1-6	1	2	4	5	6
N		200	52	18	54	51	25
Mean		19.92	16.52	16.78	21.57	21.84	21.72
Median		20.00	16.00	18.00	23.00	23.00	22.00
Std. Deviation		3.79	3.31	2.49	3.044	2.61	2.62
Variance		14.36	10.96	6.183	9.268	6.815	6.88
Minimum		8.00	8.00	11.00	14.00	13.00	15.00
Maximum		25.00	24.00	20.00	25.00	25.00	25.00
Percentiles	25	17.00	14.00	15.75	19.00	20.00	19.50
	50	20.00	16.00	18.00	23.00	23.00	22.00
	75	23.00	19.00	18.00	24.00	24.00	24.00

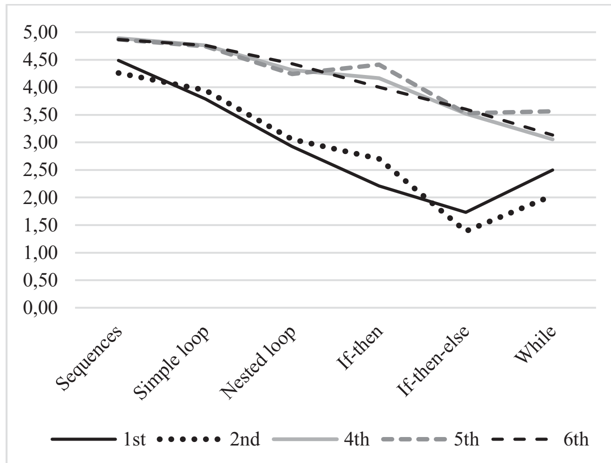

Fig. 15. Abscissa axis: computational concept by grade. Ordinate axes: BCTt item score, normalized from 0 to 5: 5 maximum score).

This score per item or difficulty index, confirms empirically the progressive difficulty anticipated by the qualitative analysis of the experts (average difficulty index = 0.81) considering the entire sample (Fig. 16). Likewise, results isolating the youngest students: 1st educational stage (subsamples A1 and B1, N=70), show increasing difficulty along the elements (N=25 items; Minimum=0.27; Maximum=0.96; average difficulty index =0.70; average total score = 16.59) (Fig. 17).

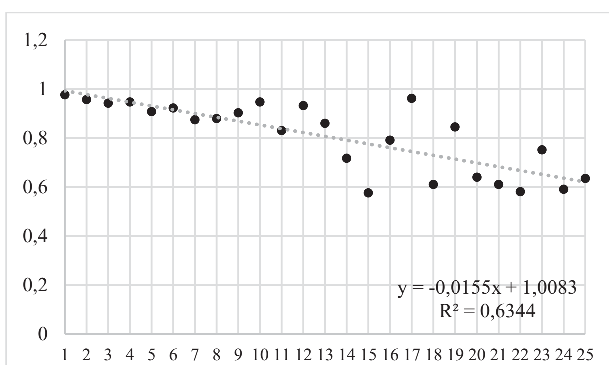

Fig. 16. Item difficulty index (ordinate axis) for each BCTt item (abscissas axis).

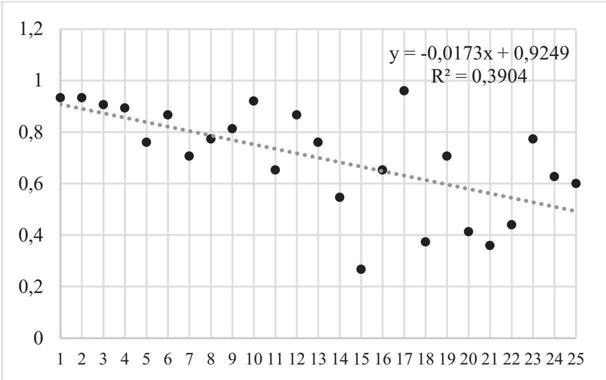

Fig. 17. Item difficulty index (ordinate axis) for each BCTt item (abscissas axis), first educational stage.

The histogram showing the distribution of the BCTt score along 1st and 2nd grades subsamples (Fig. 18), fits the normal curve and it is fairly symmetric, which suggests that the BCTt is balanced in terms of difficultly of its items for Primary School 1st educational stage.

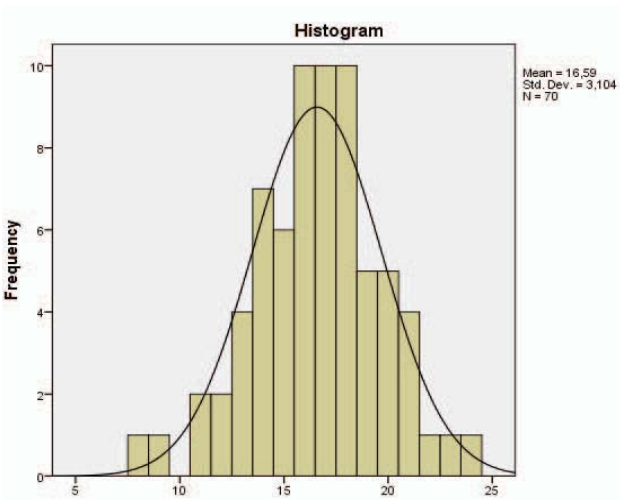

Fig. 18. Histogram of the BCTt score (1st educational stage)

3) Reliability

In order to evaluate the internal consistency associated with BCTt scores, considering all grades (N=200), a reliability analysis has been made. Cronbach's Alfa is α=0.824 (Table VIII), that can be considered as a very good reliability [32]. Reliability results by grade shows a lower Cronbach's Alpha the higher the grade is (Table IX), which leads us to conclude that BCTt is mainly aimed at first Primary School stages (grades from 1 to 4) where it shows higher reliability.

TABLE VIII. BCTT RELIABILITY STATISTICS ENTIRE SAMPLE

Sample		Reliability Statistics		Item Statistics			
N	*N of Items*	*Cronbach's Alpha*	*Cr. 's Alpha Based on Stand. Items*	*Mean*	*Min.*	*Max.*	*Variance*
200	25	0.824	0.829	0.807	0.576	0.976	0.021

TABLE IX. BCTT RELIABILITY STATISTICS BY GRADE

Subsamples				Reliability Statistics		Item Statistics	
Ed. stage	*Grade*	*Id.*	*n*	*Cronbach's Alpha*	*Cr. 's Alpha Based on Stand. Items*	*Mean*	*Variance*
1st	1	A1	52	0.833	0.838	0.742	0.041
1st	2	B1	18	0.793	0.801	0.630	0.042
2nd	4	C1	54	0.771	0.735	0.837	0.022
3rd	5	E1	51	0.660	0.683	0.863	0.012
3rd	6	F1	25	0.657	0.648	0.844	0.015

A second reliability analysis was made, this time performing task and re-task method with D1 sample (N=28). BCTt was administered to D1 subjects (2nd educational stage, 4th grade) and, 5 weeks later, was re-administered in the same conditions to the same subjects. As BCTt scores are not normally distributed in D1 subsamples (Shapiro-Wilk test of normality: test Sig.=0.03; re-test Sig.=0.01), non-parametric Spearman's test was used, showing a very strong, positive and significant correlation (r_s=0.93; p<0.01). Therefore, an excellent reliability as stability was found for the BCTt in this subsample.

IV. CONCLUSIONS

The BCTt design is considered adequate by experts for the assessment of CT in Primary School students, both in form and content, and contains relevant improvements such as transitions between squares. In addition, BCTt items seem to be ordered in increasing difficulty and relevancy. These considerations were confirmed in the statistical analysis from the administration of the test to Primary School students.

Transitions between squares are shown as a relevant improvement in maze layouts for younger students (1st educational stage), resulting in very significant higher scores compared to students with no transitions in BCTt. However, 4th and 6th grade students do not benefit from the inclusion of transitions nor were negatively affected by them. This leads us to conclude that transitions are an improvement that can be included in this type of problems since it is a significant scaffold for younger students without affecting negatively older ones.

Entire sample overall mean and BCTt scores along subsamples suggest that the test might be aimed at Primary School first educational stages, as high means are shown in 2nd and 3rd educational stages subsamples, and no significant difference is shown between total score means in older students. Thus, BCTt test seems to be aimed at 1st to 4th grades and especially to 1st and 2nd grades, as it is balanced in terms of difficulty of its items. However, there are significant differences between grades in all educational stages' subsamples score means, from question 12 and over, which leads us to conclude that the test could be aimed at all Primary School stages, but the first part of the test might include items that are too easy for older students, contrary to what was expected from the expert's content validation comments, therefore, including more items with a higher difficulty for 3rd educational stage students could be a valuable improvement in a future BCTt version.

The BCTt proved a high reliability throughout the entire sample with Cronbach's Alfa=0.824, however, a higher coefficient was obtained in younger students than in the older ones. As expected, the BCTt is more reliable in 1st and 2nd grades than in higher grades, since the difficulty level of the test fits better with the lower ones. Thus, the BCTt, as the CTt [1] is a self-contained instrument that has revealed to be reliable for the assessment of CT in Primary School and can be administered as pre-test and post-test in researches that requires it, however, as it focuses on 3D framework computational concepts, partially on computational practices and ignores computational perspectives, it is recommended to use in parallel with other assessment tools to cover its limitations [33].

From this research it can be concluded that BCTt can be used in Primary School students, particularly in first grades (5 to 10 years old), since older students (9 to 12 years old) scores results revealed that the BCTt was too easy for students of the highest grades, although it can be used focusing only on the more complex test items. Therefore, BCTt can be considered a reliable extension of the Román et al. CTt for younger students, since CTt is aimed to 10 to 16 years old students.

Further research concerns the administration of the test to 3-4 years old students, as upper age limit has been stablished, but there are concerns about lower BCTt age limit. Moreover, additional research on 3rd grade students may be necessary to exactly determine the BCTt scope. In addition, it could be enlightening to replicate the study in other countries and populations.

ACKNOWLEDGMENT

Thanks to the participants, teachers and schools involved in the learning experiences. This work has been co-funded by the Madrid Regional Government, through the project e-Madrid-CM (P2018/TCS-4307). The e-Madrid-CM project is also co-financed by the Structural Funds (FSE and FEDER).

REFERENCES

[1] M. Román-González, J. Pérez-González and C. Jiménez-Fernández, "Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test," Computers in Human Behavior, vol. 72, pp. 678-691, 2017.

[2] T. Hsu, S. Chang and Y. Hung, "How to learn and how to teach computational thinking: Suggestions based on a review of the literature," Computers & Education, vol. 126, pp. 296-310, 2018.

[3] J.M. Wing, "Computational Thinking Test," CACM Viewpoint, pp. 33-35, March. 2006.

[4] V.J. Shute, C. Sun and J. Asbell-Clarke, "Demystifying computational thinking," Educational Research Review, vol. 22, pp. 142-158, Nov. 2017.

[5] S. Grover, "The 5th 'C' of 21st Century Skills? Try Computational Thinking (Not Coding)," Feb 25, 2018.

[6] J.A. Velazquez-Iturbide, "Towards an Analysis of Computational Thinking," pp. 1-6, Sep 2018.

[7] H. Kim, H. Choi, J. Han and H. So, "Enhancing teachers' ICT capacity for the 21st century learning environment: Three cases of teacher education in Korea," Australasian Journal of Educational Technology, vol. 28, 2012.

[8] K. Tang, T. Chou and C. Tsai, "A Content Analysis of Computational Thinking Research: An International Publication Trends and Research Typology," The Asia-Pacific Education Researcher, pp. 1-11, Mar 21, 2019.

[9] S. Grover and A. Korhonen, "Unlocking the Potential of Learning Analytics in Computing Education," ACM Transactions on Computing Education (TOCE), vol. 17, pp. 1-4, Aug 29, 2017.

[10] S.Y. Lye and J.H.L. Koh, "Review on teaching and learning of computational thinking through programming: What is next for K-12?" Computers in Human Behavior, vol. 41, pp. 51-61, 2014.

[11] J.M. Wing, "Computational thinking and thinking about computing," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 366, pp. 3717-3725, Oct 28, 2008.

[12] S. Atmatzidou and S. Demetriadis, "Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences," Robotics and Autonomous Systems, vol. 75, pp. 661-670, Jan. 2016.

[13] K. Brennan, M. Resnick and MIT Media Lab, "New frameworks for studying and assessing the development of computational thinking," American Educational Research Association Meeting, Vancouver, BC, Canada, 2012.

[14] B. Zhong, Q. Wang, J. Chen and Y. Li, "An Exploration of Three-Dimensional Integrated Assessment for Computational Thinking," Journal of Educational Computing Research, vol. 53, pp. 562-590, Jan. 2016.

[15] NRC, "Report of a workshop on the pedagogical aspects of computational thinking," 2011.

[16] L. Seiter and B. Foreman, "Modeling the Learning Progressions of Computational Thinking of Primary Grade Students," in Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research, pp. 59-66, 2013.

[17] L. Werner, J. Denner, S. Campe and D. Kawamoto, "The fairy performance assessment," pp. 215-220, 2012.

[18] A. Basawapatna, K.H. Koh, A. Repenning, D. Webb and K. Marshall, "Recognizing computational thinking patterns," pp. 245-250, Mar 9, 2011.

[19] A. Mühling, A. Ruf and P. Hubwieser, "Design and First Results of a Psychometric Test for Measuring Basic Programming Abilities," pp. 2-10, Nov 9, 2015.

[20] D. Weintrop and U. Wilensky, "Using Commutative Assessments to Compare Conceptual Understanding in Blocks-based and Text-based Programs," pp. 101-110, Aug 9, 2015.

[21] D. Franklin, P. Conrad, B. Boe, K. Nilsen, C. Hill, M. Len, G. Dreschler, G. Aldana, P. Almeida-Tanaka, B. Kiefer, C. Laird, F. Lopez, C. Pham, J. Suarez and R. Waite, "Assessment of computer science learning in a scratch-based outreach program," pp. 371-376, Mar 6, 2013.

[22] J. Denner, L. Werner and E. Ortiz, "Computer games created by middle school girls: Can they be used to measure understanding of computer science concepts?" Computers & Education, vol. 58, pp. 240-249, 2012.

[23] J. Moreno-León and G. Robles, Dr. Scratch: a Web Tool to Automatically Evaluate Scratch Projects, 2015.

[24] M. Román González, "Computational Thinking Test: Design Guidelines and Content Validation," in EDULEARN15 At: Barcelona, 2015.

[25] M. Román-González, J. Pérez-González, J. Moreno-León and G. Robles, "Extending the nomological network of computational thinking with non-cognitive factors," Computers in Human Behavior, vol. 80, pp. 441-459, 2018.

[26] M. Román-González, J. Pérez-González, J. Moreno-León and G. Robles, "Can computational talent be detected? Predictive validity of the Computational Thinking Test," International Journal of Child-Computer Interaction, 2018.

[27] AERA, APA and NCME, Standards for educational and psychological testing, Washington: American Educational Research Association, 2014.

[28] D. Căprioară, "Emotions and the Learning of School Mathematics," Bulletin of the Transilvania University of Braşov, Series VII: Social Sciences and Law, vol. 10, pp. 9-18, 2017.

[29] C. Chen and P. Herbst, "The interplay among gestures, discourse, and diagrams in students' geometrical reasoning," Educational Studies in Mathematics, vol. 83, pp. 285-307, 2013.

[30] H.Y. Durak and M. Saritepeci, "Analysis of the relation between computational thinking skills and various variables with the structural equation model," Computers & Education, vol. 116, pp. 191-202, 2018.

[31] T. Watanabe, "Visual Reasoning Tools in Action: Double Number Lines, Area Models, and Other Diagrams Power Up Students' Ability to Solve and Make Sense of Various Problems," Mathematics Teaching in the Middle School, vol. 21, pp. 152, 2015.
[32] C.E. Lance, M.M. Butts and L.C. Michels, "The sources of four commonly reported cutoff criteria: What did they really say?" Organ.Res.Methods, vol. 9, pp. 202-220, 2006.
[33] M. Román-González, J. Moreno-León and G. Robles, "Combining Assessment Tools for a Comprehensive Evaluation of Computational Thinking Interventions," in Computational Thinking Education, S. Kong and H. Abelson, Singapore: Springer Singapore, 2019, pp. 79-98.