



Article

# Design and Validation of a Computational Thinking Test for Children in the First Grades of Elementary Education

Jorge Hernán Aristizábal Zapata <sup>1</sup>, Julián Esteban Gutiérrez Posada <sup>2</sup> and Pascual D. Diago <sup>3,\*</sup>

<sup>1</sup> Department of Mathematics, Faculty of Education, Quindío University, Armenia 630001, Quindío, Colombia; jhaz@uniquindio.edu.co

<sup>2</sup> Department of Computer and Systems Engineering, Faculty of Engineering, Quindío University, Armenia 630001, Quindío, Colombia; jugutier@uniquindio.edu.co

<sup>3</sup> Department de Didàctica de la Matemàtica, Facultat de Magisteri, Universitat de València, 46022 València, Spain

\* Correspondence: pascual.diago@uv.es

**Abstract:** Computational thinking (CT) has garnered significant interest in both computer science and education sciences as it delineates a set of skills that emerge during the problem-solving process. Consequently, numerous assessment instruments aimed at measuring CT have been developed in the recent years. However, a scarce part of the existing CT measurement instruments has been dedicated to early school ages, and few have undergone rigorous validation or reliability testing. Therefore, this work introduces a new instrument for measuring CT in the early grades of elementary education: the Computational Thinking Test for Children (CTTC). To this end, in this work, we provide the design and validation of the CTTC, which is constructed around spatial, sequential, and logical thinking and encompasses abstraction, decomposition, pattern recognition, and coding items organized in five question blocks. The validation and standardization process employs the Kuder–Richardson statistic (KR-20) and expert judgment using V-Aiken for consistency. Additionally, item difficulty indices were utilized to gauge the difficulty level of each question in the CTTC. The study concludes that the CTTC demonstrates consistency and suitability for children in the first cycle of primary education (encompassing the first to third grades).

**Keywords:** computational thinking; problem solving; elementary education; psychometric test



**Citation:** Aristizábal Zapata, J.H.; Gutiérrez Posada, J.E.; Diago, P.D. Design and Validation of a Computational Thinking Test for Children in the First Grades of Elementary Education. *Multimodal Technol. Interact.* **2024**, *8*, 39. <https://doi.org/10.3390/mti8050039>

Academic Editor: Jerry Alan Fails

Received: 10 April 2024

Revised: 2 May 2024

Accepted: 7 May 2024

Published: 9 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction and Background

Computational thinking (CT) traces its roots back to Papert’s work [1], where programming in LOGO involved a robotic turtle and provided students with the opportunity to explore, construct, and devise their own problem-solving strategies using commands. Subsequently, Jeannette Wing introduced the term CT [2] and emphasized its foundational role for all individuals, as it entails problem-solving skills that are not limited to computer scientists. CT has surged in popularity in recent years due to its problem-solving efficacy across various domains [3–5] as well as its associations with logic, creativity [6], and imagination [7]. Regarding educational policies, numerous countries have integrated CT-related activities into their curricula [8,9], echoing earlier calls [10] for CT’s integration as a core competency in education.

Furthermore, introducing CT in classrooms can foster innovation and creativity among children, equipping future generations with the cognitive tools, skills, and strategies necessary to navigate diverse problems and changing environments. Research indicates that CT can be introduced in early years of basic education using developmentally appropriate tools [11–16], thereby significantly enhancing skill development from an early age [3,17,18]. Moreover, according to Voogt et al., CT is “considered a universal competence, which should be added to every child’s analytical ability as a vital ingredient of their school learning” [19] (p. 715). Despite the extensive research conducted on CT in

recent years, which has included various types of assessments such as pre-tests, post-tests, rubrics, observations, and semi-structured interviews, among other methods, Fields et al. note that “one of the main challenges in assessing CT is evaluating students’ depth of understanding” [20] (p. 226).

Thus, the interest in measuring CT and the skills associated with it has grown in recent years. As established in one of the former works on the importance of CT by educational levels, the assessment of CT plays a critical role in K–12 classrooms [21]. Indeed, the mastery and application of CT skills in real-life situations have heightened the interest in measuring the abilities related to CT at all educational levels [22]. Consequently, research in the field of education has produced numerous instruments aimed at this purpose [23]. The work of Tsarava [24] provides a comprehensive overview of the assessment tools designed in recent years. As can be seen, various assessment tests have been designed that focus on different research areas related to education (specific educational programming environments, specific STEM topics, etc.). However, one of the most relevant approaches to assess CT has been the development of psychometric-like tests designed to assess CT independent of specific programming environments. However, as we describe in the following section, most CT measurement instruments are targeted towards students at the end of primary education or in subsequent stages. Furthermore, few of these instruments have undergone validation tests or statistical reliability assessments, as we will describe in the systematic review.

The present work outlines the process of designing, constructing and validating a new instrument: the Computational Thinking Test for Children (CTTC). The CTTC aims to serve as a diagnostic tool to gauge the level of CT among children in the early grades of primary education through the evaluation of the ability to logically solve problems and without the need for prior programming knowledge. Therefore, this paper focuses on the following research question:

**RQ** Is the CTTC a valid and reliable test from a psychometric approach for measuring CT in the first grades of elementary education?

## 2. Background and Systematic Literature Review

The study of CT has been approached from various perspectives. Numerous systematic reviews and meta-analyses on CT provide insight into published works that focused on the relationship between CT and subject matter domains, associated cognitive constructs, the technological tools used, or the potential effects on programming skills, among other topics [23,25,26]. In particular, some of these studies have focused on measuring CT in various ways [8,23]: (i) measuring core elements of programming [27–30]; (ii) measuring problem-solving abilities [8,28,31,32]; and (iii) addressing reasoning skills, spatial ability, and problem-solving skills [8].

Nevertheless, only some of these works aim to provide psychometric tests to measure CT [8,31,33,34]. Delving into the reliability and validity of CT assessments, the work of Tang et al. reports that among 96 studies analyzed, only 45% reported reliability evidence, and 18% provided validity evidence [23]. The authors conclude that while some CT assessments offer reliability and validity information, the majority lack sufficient evidence in these areas. This absence of evidence poses challenges for confidently utilizing these assessments in classroom settings to evaluate students’ CT learning. In particular, focusing on the early levels of primary education, few studies provide reliable and validated evidence at these ages [24,29].

The construction of the CTTC began with a systematic literature review of questionnaires assessing CT and skill development in children. The review identified only eight studies at the international level wherein CT is evaluated using psychometrically validated instruments (see Table 1). Seven of these studies focused on primary and middle school students ranging from 3rd to 10th grade, with one targeting pre-university students [8,27–29,31,32,35]. Only one study specifically addressed children in the first and second grades of primary school [30]. This scarcity of research on early childhood CT assess-

ment may stem from a lack of understanding regarding the development of CT in young children, as discussed by [36]. Consequently, research in this area, particularly concerning younger children, remains limited [37,38]. From this starting point, our research question [RQ] is revealed to be particularly pertinent as, from the systematic review conducted thus far, few instruments have been identified that have undergone psycho-technical validation.

**Table 1.** Description of psycho-metrically validated articles.

Test name	Population	Items	Measuring	Approach
Computational-thinking-based science learning (CTSiM, 2014) [27]	25 students from 6th grade	4	Abstractions, algorithms, conditionals, loops, and variables	Model accuracy metric
Computational Thinking Test (CTt, 2017) [8]	1251 students from 5th to 10th grade	40	Spatial ability, reasoning ability, and problem-solving ability	Descriptive statistics comparing with other tests: Primary Mental Abilities (PAM) and Solving Problem Test (RP30)
Assessing elementary students' computational thinking (2017) [31]	121 students from 5th grade	23	Syntax for formulating problems and solutions, data algorithms representing efficient and effective solutions	Cronbach's alpha, reliability coefficient, two-tailed, two sample t-test, and Rasch testlet model
Exploring the relationship between computational thinking skills and academic performance (2017) [28]	104 pre-university science students	29	Algorithmic thinking, cooperativity, creativity, critical thinking, and problem-solving	Partial least squares approach
Computational Thinking Self-efficacy Scale (2019) [35]	319 students from 5th to 7th grade	18	Computational thinking self-efficacy, reasoning, abstraction, decomposition, and generalization	Chi-square, Cronbach's alpha reliability coefficient, and expert opinion
TechCheck (2020) [30]	612 students from 1st to 2nd grade	15	Algorithm modularity, control structures, symbolic representation, hardware/software, and debugging	Classical Test Theory (CTT), Item Response Theory (IRT), and evaluator opinion
Computational thinking as a problem-solving strategy (2021) [32]	66 students from 10th grade	15	CT through solving problems	Root mean square error of approximation, comparative fit index (CFI), and Tucker-Lewis index (TLI)
The competent Computational Thinking test (cCTt, 2023) [29]	2666 students from 3rd to 6th grade	25	Blocks, sequences, simple loops, complex loops, conditional statements, while statements, and combinations	Classical Test Theory, Item Response Theory (IRT)

### 3. Materials and Methods

The construction of the CTTC proceeded through several stages: Initially, the previously described systematic review was conducted to justify the development process. Subsequently, a preliminary version of the questionnaire was crafted. Expert judges then assisted with validating the initial design through a rigorous evaluation process; this was followed by a pilot test to refine the instrument. Upon making necessary adjustments based on feedback, the finalized version was administered to the target population to assess the validity of the instrument. The following sections provide detailed descriptions of the outlined phases.

#### 3.1. Fundamentals and Initial Design of the CTTC

The systematic review of the validated tests described above led to the identification of three key thinking skills underpinning the construction and validation of the CTTC:

1. Spatial thinking enables children to visually represent information, comprehend and manipulate spatial relationships, and understand object positions, orientations, and directions [39]. Moreover, as noted by Kwon et al., "spatial abilities are used for

daily tasks including reading maps, driving cars, drawing objects from different perspectives, or even folding clothes" [16] (p. 4).

2. Sequential thinking involves the comprehension, ordering, and recall of sequences of events and aids children with understanding numerical order and enhances their problem-solving abilities [40].
3. Logical thinking fosters problem solving through abstraction and pattern recognition. According to Oljayevna and Shavkatovna [41], the development of logical thinking entails observing and comparing objects, identifying similarities and differences, discerning essential features, drawing conclusions from observations or facts, and presenting ideas logically and coherently.

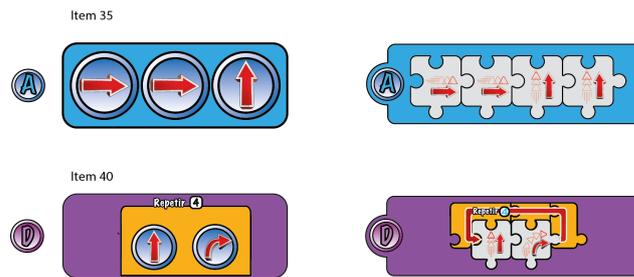
The development of the initial version of the CTTC ensued subsequent to the prior systematic review of existing supporting instruments and tests implemented within this domain. As described above, the CT concepts underpinning the construction and validation of the CTTC encompass: (1) spatial thinking, which enables children to visually represent information and comprehend and manipulate spatial relationships as well as grasp the relative positioning of objects, directions, and orientations [39]; (2) sequential thinking, which involves the capacity to understand, organize, and recall sequences of events; and (3) logical thinking, which facilitates problem solving through intricate mental processes [42,43] spanning from abstraction to pattern recognition. In light of this context, two specific tests were utilized as reference points, namely:

- The Brebas Test (<http://www.bebas.org>, accessed on 1 April 2024), which endeavors to present intriguing tasks aimed at inspiring students to delve deeper into technology-related concepts;
- The Computer Olympiad (<https://olympiad.org.za/>, accessed on 1 April 2024), which seeks to familiarize students and educators with computational thinking and computer science through enjoyable and interactive tasks. These tasks enable students to explore their aptitude for computational thinking without necessitating prior knowledge.

Based on the considerations above, the CTTC items were designed with attention to a visually appealing layout and an age-appropriate level of difficulty, ensuring they were neither excessively challenging to cause frustration nor overly simplistic to result in demotivation among children. The initial CTTC comprises a total of 40 multiple-choice items, each with a single response. The final version of the CTTC can be consulted at <https://go.uv.es/pasdadia/cttc> (accessed on 2 May 2024); Figures 1 and 2 show two examples of the designed items.

One point is assigned for each correct answer, resulting in a maximum score of 40. The selection of this number of questions was driven by the aim to cover all five defined dimensions. The test is structured into five dimensions focusing on problem-solving aspects: laterality (8 items), location (12 items), spatial rotation (7 items), sequence of instructions (7 items), and cycles (6 items). These dimensions are crucial for spatial cognition, learning, and scientific inquiry [44]. Additionally, the relationship between spatial skills and mathematical problem solving is emphasized [45], highlighting the importance of spatial thinking, sequence planning, and pattern recognition. Children often encounter challenges in solving spatial displacement problems, such as planning algorithms for movement using floor robots or programmable vehicles. Specifically, they may struggle with positioning themselves relative to the vehicle: a concept related to Piaget's egocentric stage [46]. It is essential for children to envision themselves as drivers of the vehicle to execute appropriate movements and turns, which is a skill they may not initially possess when tackling such problems [47].

It is noteworthy that the CTTC's items address abstraction, decomposition, pattern recognition, and coding [48,49], as introducing these concepts at an early age is crucial for fostering advanced reasoning skills. Following the development of the test questions, the validation process commenced.



**Figure 1.** Icons for student responses in two of the CTTC items. Each response item has been assigned a color to prevent children from confusing items when responding to the questionnaire.



**Figure 2.** An example of a CTTC question type from the test categories with its respective answer sheet.

### 3.2. Expert Assessment

For the validation and standardization process of the CTTC instrument, a panel of multidisciplinary experts at the master and doctoral levels was assembled following the criteria established by [50]. These criteria included expertise in the subject matter, academic recognition, availability of time for test review, and motivation to participate impartially. The selected experts hailed from diverse fields such as mathematics, education, engineering, and psychology. Each expert possessed significant experience in theoretical, practical, and evaluative domains. They were tasked with conducting a comprehensive evaluation of the questionnaire as well as each individual item.

To ensure content validity of the instrument, a panel of eight experts was selected in line with the recommendation by Grant and Davis that establishes that the number of judges should range from two to twenty [51]. Each expert independently rated every item on the questionnaire by assigning scores ranging from one to five (where one indicates strong disagreement and five signifies strong agreement) based on his/her assessment of the item’s illustration, design, wording, appropriateness for the questionnaire, and alignment with the intended age group. Each expert received an anonymous rating sheet to ensure precise evaluation of the questionnaire.

After all items were evaluated by the judges, interpretation of their responses was conducted. The V-Aiken method [52] was employed to ascertain the relevance of each item to the overall construct. Content validity analysis led to the modification of certain items; this involved clarification of language to prevent confusion and enhancement of questionnaire illustrations based on judges’ recommendations.

### 3.3. Pilot Test of the CTTC

Finally, a pilot test was conducted to assess the comprehension of the CTTC questionnaire items by the children. Subsequently, the adjusted pilot test, based on expert feedback, was administered to a group of six students representing the targeted grade levels: two children from first grade, two from second grade, and two from third grade. This allowed for the evaluation of children’s comprehension of the questions and their

feedback on the test design. During the administration of the pilot test, no difficulties were encountered by the participating children in understanding the items.

### 3.4. Construction of the Final Version of the CTTC

In constructing the final version of the CTTC, experts' suggestions were considered. Visual design aimed to be appealing to children by incorporating vibrant colors and using age-appropriate pictorial representations that conveyed correct actions. Icons representing familiar actions were utilized to avoid leading responses, as highlighted by [53]. Effective design facilitates interpretation and understanding, while poor design may introduce false cues, hindering comprehension. To promote reasoning skills, the initial basic design was enhanced to a more structured format. In response to judges' feedback, icons were restructured from simple illustrations of circles with arrows to puzzles in order to leverage children's familiarity with this game format. Additionally, arrows were modified to simulate movement (see Figure 1).

Considering the age of the children for whom the test is intended and their potential unfamiliarity with numbers greater than ten [54], we opted to represent each question number with a drawing of an animal that children are likely to be familiar with. These animal drawings are also present on the answer sheet, facilitating children's association between the questions and their corresponding numbers (see Figure 2). To enhance comprehension, the texts of the questions were kept concise.

### 3.5. Validation and Reliability of the CTTC

To ascertain the validity and reliability of this instrument, an intentional non-probabilistic sample of six groups, categorized by grade level from first to third grade, was drawn from the Quindío state of Colombia. The sample consisted of 118 students who participated voluntarily, comprising 56.3% females and 43.8% males (65 girls and 53 boys). The ages of the participants ranged from five to eight years old. Specifically, the sample included 33 students from first grade, 38 from second grade, and 47 from the third grade of primary school (see Table 2).

**Table 2.** Demographic information of the participants.

		Frequency	Percent	Cumulative Percent
Gender	Male	53	44.915	44.915
	Female	65	55.085	100.000
Total		118	100.000	
Grade	1st Grade	33	27.996	27.996
	2nd Grade	38	32.203	61.169
	3rd Grade	47	39.831	100.000
Total		118	100.000	

The CTTC was administered in person to evaluate the reliability of the instrument. The test was administered anonymously to the specified students in various towns. Prior to the administration of the CTTC, parental or legal guardian consent was obtained for each participant. Given the children's age, the number of questions, and their attention span variability, the CTTC was divided into five blocks. Following each block, a break was provided, during which the next block of questions was introduced along with a concrete example. This approach aimed to familiarize the children with the upcoming questions. Subsequently, the questionnaires were collected and analyzed separately for each grade to observe the performance on each question across grades.

### 3.6. Difficulty of the CTTC Items

After administering the final version of the CTTC, the difficulty level of each question comprising the test was determined. To accomplish this, the item difficulty index [55] was

utilized, which employs a five-level scale to classify the expected difficulty of questions in a questionnaire. The scale includes categories such as “easy questions”, “moderately easy questions”, “medium difficulty questions”, “moderately hard questions”, and “difficult questions”, along with their respective percentages.

#### 4. Results and Discussion

In this section, the findings regarding the validity of the CTTC are presented.

##### 4.1. Validation of Content

The expert judges made significant contributions to enhancing the CTTC, particularly regarding (1) the graphics and illustrations used in the questions, recognizing that drawing is a form of communication for children and engages various cognitive capacities [56], and (2) the restructuring of question wording to enhance item comprehension.

The content validity of the instrument was assessed through expert judgment. Agreement among the expert judges was evaluated using Aiken’s V and revealed that no question was deemed inappropriate, as the validity coefficient for each question exceeded or equaled 0.7. However, question 29 had a coefficient value of 0.698, which was very close to 0.7 (see Table 3). Consequently, all test items were retained, as Aiken’s V indicated that all questions in the questionnaire were suitable for assessing the intended categories. Table 4 shows the statistical summary of the Aiken’s V values obtained in Table 3.

**Table 3.** Values of Aiken’s V and 95% confidence intervals for the 40 CTTC items.

Item	Mean	sd	V-Aiken	95% CI	
				Lower Limit	Upper Limit
1	4.458	0.212	0.865	0.707	0.944
2	4.375	0.270	0.844	0.682	0.931
3	4.042	0.257	0.760	0.590	0.875
4	4.000	0.306	0.750	0.579	0.867
5	4.208	0.212	0.802	0.635	0.904
6	3.958	0.212	0.740	0.568	0.860
7	4.167	0.156	0.792	0.624	0.897
8	4.333	0.358	0.833	0.670	0.925
9	4.083	0.358	0.771	0.601	0.882
10	4.375	0.270	0.844	0.682	0.931
11	4.292	0.412	0.823	0.659	0.918
12	4.292	0.059	0.823	0.659	0.918
13	4.042	0.059	0.760	0.590	0.875
14	3.875	0.354	0.719	0.546	0.844
15	4.250	0.270	0.813	0.647	0.911
16	4.042	0.312	0.760	0.590	0.875
17	4.208	0.059	0.802	0.635	0.904
18	3.958	0.059	0.740	0.568	0.860
19	4.125	0.270	0.781	0.612	0.890
20	4.125	0.354	0.781	0.612	0.890
21	4.333	0.295	0.833	0.670	0.925
22	4.208	0.059	0.802	0.635	0.904
23	4.208	0.312	0.802	0.635	0.904
24	4.208	0.425	0.802	0.635	0.904
25	3.958	0.156	0.740	0.568	0.860
26	4.292	0.295	0.823	0.659	0.918
27	4.167	0.358	0.792	0.624	0.897
28	4.000	0.177	0.750	0.579	0.867
29	3.792	0.156	0.698	0.525	0.829

**Table 3.** *Cont.*

Item	Mean	sd	V-Aiken	95% CI	
				Lower Limit	Upper Limit
30	4.083	0.156	0.771	0.601	0.882
31	4.042	0.236	0.760	0.590	0.875
32	4.042	0.412	0.760	0.590	0.875
33	4.208	0.412	0.802	0.635	0.904
34	4.167	0.257	0.792	0.624	0.897
35	4.125	0.306	0.781	0.612	0.890
36	4.333	0.257	0.833	0.670	0.925
37	4.250	0.204	0.813	0.647	0.911
38	3.958	0.460	0.740	0.568	0.860
39	4.333	0.358	0.833	0.670	0.925
40	4.125	0.204	0.781	0.612	0.890

**Table 4.** Statistical summary of all Aiken's V values obtained in Table 3.

N	Min.	Max.	Mean	sd	Variance
40	1	33	4.151	0.49	0.24

#### 4.2. Internal Validity

For the reliability analysis of the CTTC, the KR-20 Alpha coefficient was utilized, as it is equivalent to Cronbach's Alpha [57]. Regarding internal consistency, given that the questionnaire items were multiple-choice with a single correct answer, a dichotomous scale was applied, where only the correct responses were scored. Children were deemed to have failed an item if they either marked more than one response for a specific question or left it unanswered. To assess internal consistency, the Kuder–Richardson reliability coefficient (KR-20) [58] was computed at a 95% confidence interval. The data were analyzed using JASP software, which yielded an estimate of 0.838 for the KR-20 Alpha coefficient. This value indicates high internal consistency, as the generally accepted threshold for reliability measures in cognitive tests is 0.8, while for ability tests, a cut-off point of 0.7 is more appropriate, as outlined by [59]. With all 40 questions considered, the calculated KR-20 Alpha was 0.838 (see Table 5), indicating adequate internal consistency for the CTTC.

**Table 5.** Frequentest scale reliability statistics for KR-20.

Estimate	$\alpha$ KR-20	Mean	sd
Point estimate	0.838	24.559	6.398
95% CI lower bound	0.792	23.405	5.673
95% CI upper bound	0.875	25.714	7.337

#### Construct Validity

During the administration of the questionnaire, no difficulties were reported in comprehension among the participating children. This outcome may be attributed to the incorporation of feedback from expert judges and the conduct of a pilot test prior to the final administration of the CTTC.

Regarding the difficulty index by item, notable disparities were observed when comparing the expected percentages to the actual values for the CTTC. Specifically, the percentage of moderately easy questions exceeded the expected value of 20% by 5%, reaching an actual value of 25%. Similarly, the percentage of medium-difficulty questions decreased by 12.5% compared to the expected value of 50%, resulting in an actual value of 37.5%. Conversely, the percentage of difficult questions increased by 7.5% relative to the expected value of 5%, reaching an actual value of 12.5%. The percentages for easy questions and

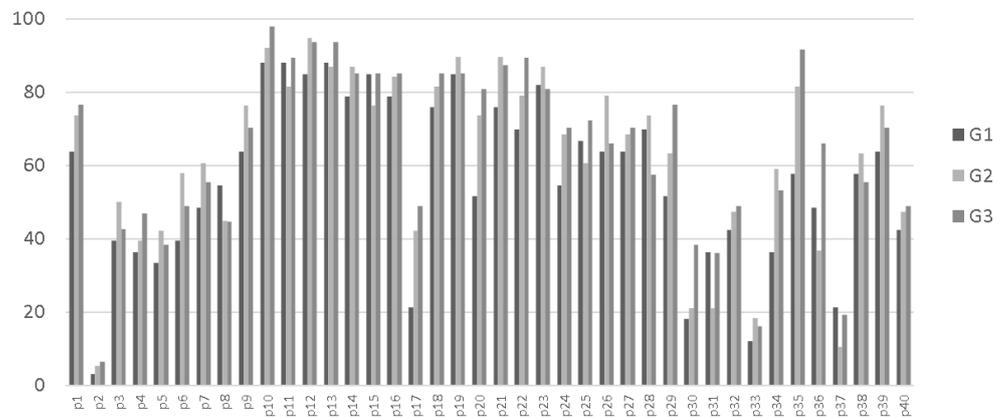
moderately hard questions aligned closely with the theoretically expected values of 5% and 20%, respectively (see Table 6).

**Table 6.** Percentage distribution of the difficulty of the items in the CTTC.

Classification by Item	Difficulty Index	Expected Percentage Distribution	Real Percentage Distribution
Easy questions	0.91–1	5%	5%
Moderately easy questions	0.81–0.90	20%	25%
Medium-difficulty questions	0.51–0.80	50%	37.5%
Moderately hard questions	0.40–0.50	20%	20%
Difficult questions	0–0.39	20%	12.5%

The difficulty levels proposed by [55] as well as the theoretically expected values differ from the observed results regarding the difficulty levels of items in the CTTC. While some items aligned closely with the expected values upon administering the CTTC, others exhibited varying degrees of difficulty compared to what was anticipated.

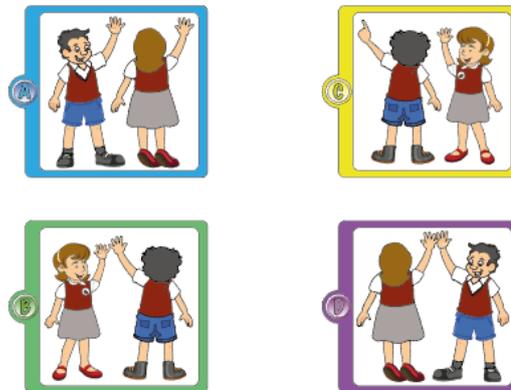
To evaluate the extent of children’s comprehension of the items and to analyze the construct validity of the developed instrument, the test was administered to 118 students in the first (G1), second (G2), and third (G3) grades of elementary education. It was observed that the test exhibited similar patterns across all three groups. Each group achieved a higher number of correct answers proportionate to its respective grade level (see Figure 3). This trend can be attributed to the students’ varying levels of cognitive maturation and the novelty of encountering this type of test for the first time across all groups.



**Figure 3.** Results for the items by grade.

The notably low performance across all three grades for question number two is striking (see Figure 3). However, it received a favorable evaluation in the expert assessment: achieving a V-Aiken value of 0.844 (see Table 3). This low score could potentially stem from the children’s challenges in recognizing the right and left sides or in mentally rotating and accurately locating their position with a raised hand (see Figure 4).

Which pair of children has their right hand raised?



**Figure 4.** Second item of the CTTC, which posed greater difficulty for children due to confusion over the position of individuals in the image.

## 5. Conclusions

The use of the CTTC proposes a crucial difference from other validated tests found so far, as some aim to measure core elements of programming [27–30], while [8,28,31,32] emphasize problem solving, with only [8] addressing reasoning skills, spatial ability, and problem-solving skills. This aligns with [60], which suggests that there are two categories for evaluation and teaching: one focused on programming concepts and the other focused on finding solutions through thinking skills such as logical organization and data analysis. In this regard, the CTTC aims to integrate both categories by identifying children’s prior knowledge by inquiring about spatial concepts, sequence recognition, and patterns and ending with key programming elements. Throughout the test, children must engage in a process of abstraction to identify the correct solution, whether it involves classifying or mentally rotating question elements to determine relevant information. Similarly, they must decompose the problem for resolution and progressively recognize regularities and patterns to generalize and identify possible algorithms.

In conclusion, the CTTC emerges as a concise and clear instrument that exhibits reliability across its diverse items. Its validity is underscored by the coherent content, which is implied by a moderate relationship among study questions and consistent responses from participants. The five key constructs incorporated in the CTTC—laterality, location, spatial rotation, sequence of instructions, and cycles—comprehensively capture the essence of computational thinking. Test performance variability reflects children’s aptitude at abstracting, decomposing, recognizing patterns, and coding.

The validation process of the CTTC has yielded valuable insights into its content validity, reliability, and difficulty index by item. The assessment of content validity through expert judgment using Aiken’s  $V$  coefficient demonstrated high agreement among expert judges regarding the appropriateness of the test items, indicating their suitability for assessing the intended categories. The reliability analysis, assessed through the KR-20 Alpha coefficient, showcased high internal consistency for the CTTC. Furthermore, the assessment of the difficulty index by item revealed noteworthy disparities between expected and actual percentages for certain question types within the CTTC. While moderately easy questions surpassed expected percentages, medium-difficulty questions exhibited a decrease, and difficult questions showed an increase compared to theoretical expectations. Nevertheless, easy and moderately hard questions closely aligned with anticipated percentages. Overall, these findings underscore the CTTC’s efficacy as a reliable and valid instrument for assessing computational thinking in children. Further research and refinement may address

discrepancies in the difficulty index and enhance the instrument's precision and utility in evaluating computational thinking skills.

It is worth mentioning that the validity of the CTTC may be affected by contextual factors that could influence the students' outcomes and may, consequently, affect the obtained results. For example, the performance of third grade students appears to surpass that of the subsequent grades, as students enhance their CT abilities as they mature, even in the absence of explicit CT instruction [61], aligning with the notion that CT is a universal skill [19]. In this regard, it would be highly interesting to propose a longitudinal study with the same students to observe and measure the acquisition of CT across various educational stages. Moreover, the outcomes of the students in the sample may vary from that of students in different regions and countries. Therefore, it would be valuable to gather data from another cohort of students spanning the same grades to replicate the study. As future work, we include these two proposals described to continue investigating the acquisition of CT in early school ages.

Another limitation worth mentioning is the study of potential gender differences in the acquisition of computational thinking. While our manuscript delineates the demographic distribution, specifically indicating 56.3% female and 43.8% male participants, we concur that a thorough analysis of potential gender-specific differences in the results is warranted. Therefore, as a future avenue of research, we intend to delve into a comparative analysis between genders to elucidate any discernible variations in computational thinking acquisition. This endeavor will involve a meticulous examination of the data to identify potential patterns or disparities accompanied by an exploration of underlying factors contributing to any observed differences or lack thereof. By addressing this gap in our current study, we aim to contribute to the broader discourse on gender-inclusive approaches to fostering computational thinking skills.

The utility of the CTTC extends beyond the confines of formal education, as it serves as a valuable tool for assessing computational thinking in individuals across various educational backgrounds. Given that CT fosters the development of essential skills such as problem solving, creativity, and analytical thinking, the CTTC holds promise as an effective assessment instrument. Moreover, it can be deployed to evaluate different intervention strategies aimed at enhancing computational thinking skills. Aligned with existing psychometric tools measuring CT, the design and evaluation of the CTTC demonstrate consistency by targeting essential computational thinking skills such as abstraction, decomposition, pattern recognition, and coding. However, it is worth noting that some children may encounter difficulties distinguishing between left and right, which may impact their performance on certain test items. Nonetheless, experts have deemed the items relevant for assessing computational thinking, further affirming the validity of the CTTC as a robust assessment tool in the first grades of elementary education.

**Author Contributions:** Conceptualization, J.H.A.Z. and P.D.D.; methodology, J.H.A.Z. and J.E.G.P.; validation, J.H.A.Z. and J.E.G.P.; resources, J.H.A.Z.; data curation, J.H.A.Z. and J.E.G.P.; writing—original draft preparation, J.H.A.Z., J.E.G.P. and P.D.D.; writing—review and editing, J.H.A.Z., J.E.G.P. and P.D.D.; supervision, P.D.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Conselleria d'Innovació, Universitats, Ciència i Societat Digital de la Generalitat Valenciana through grant number GV/2019/146.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and the Code of Good Practices in Research from Quindío University and Universitat de València.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript (sorted by the order of appearance in the text):

CT	computational thinking
CTTC	Computational Thinking Test for Children
KR-20	Kuder–Richardson Alpha coefficient
V-Aiken	Aiken’s content validity coefficients
K–12	educational levels from kindergarten to 12th grade
STEM	science, technology, engineering, and mathematics
RQ	research question
CTt	Computational Thinking Test
PAM	Primary Mental Abilities
CTT	Classical Test Theory
IRT	Item Response Theory
CFI	Comparative Fit Index
TLI	Tucker–Lewis Index
cCTt	competent Computational Thinking test

## References

- Papert, S. *Children, Computers, and Powerful Ideas*; Harvester: Birmingham, UK, 1980.
- Wing, J.M. Computational thinking. *Commun. ACM* **2006**, *49*, 33–35. [[CrossRef](#)]
- Grover, S.; Pea, R. Computational thinking: A competency whose time has come. In *Computer Science Education: Perspectives on Teaching and Learning in School*; Bloomsbury Publishing: London, UK, 2018; Volume 19; pp. 19–38.
- Angulo, J.A.P. El pensamiento computacional en la vida cotidiana. *Rev. Sci.* **2019**, *4*, 293–306. [[CrossRef](#)]
- Tekdal, M. Trends and development in research on computational thinking. *Educ. Inf. Technol.* **2021**, *26*, 6499–6529. [[CrossRef](#)]
- Fuentes Pérez, A.D.; Valladares, G.M. Desarrollo y Evaluación del Pensamiento Computacional: Una Propuesta Metodológica y una Herramienta de Apoyo. In Proceedings of the IV Congreso Internacional Sobre Aprendizaje, Innovación y Competitividad (CINAIC 2017), Zaragoza, Spain, 4–6 October 2017; pp. 577–582.
- Tsortanidou, X.; Daradoumis, T.; Barberá, E. Connecting moments of creativity, computational thinking, collaboration and new media literacy skills. *Inf. Learn. Sci.* **2019**, *120*, 704–722. [[CrossRef](#)]
- Román-González, M.; Pérez-González, J.C.; Jiménez-Fernández, C. Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Comput. Hum. Behav.* **2017**, *72*, 678–691. [[CrossRef](#)]
- Piatti, A.; Adorni, G.; El-Hamamsy, L.; Negrini, L.; Assaf, D.; Gambardella, L.; Mondada, F. The CT-cube: A framework for the design and the assessment of computational thinking activities. *Comput. Hum. Behav. Rep.* **2022**, *5*, 100166. [[CrossRef](#)]
- Zapata-Ros, M. Pensamiento computacional: Una nueva alfabetización digital. *Rev. Educ. Distancia (RED)* **2015**, *46*, 1–47. [[CrossRef](#)]
- Bers, M.U. *Blocks, Robots and Computers: Learning about Technology in Early Childhood*; Teacher’s College Press: New York, NY, USA, 2008.
- Kazakoff, E.R.; Sullivan, A.; Bers, M.U. The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. *Early Child. Educ. J.* **2013**, *41*, 245–255. [[CrossRef](#)]
- Duncan, C.; Bell, T. A pilot computer science and programming course for primary school students. In Proceedings of the Workshop in Primary and Secondary Computing Education, London, UK, 9–11 November 2015; pp. 39–48.
- Álvarez-Herrero, J. Computational thinking in early childhood education, beyond floor robots. *Educ. Knowl. Soc.* **2020**, *21*, 1–11.
- Terroba, M.; Ribera, J.M.; Lapresa, D.; Anguera, M.T. Education intervention using a ground robot with programmed directional controls: Observational analysis of the development of computational thinking in early childhood education. *Rev. Psicodidáctica* **2021**, *26*, 143–151. [[CrossRef](#)]
- Kwon, K.; Jeon, M.; Zhou, C.; Kim, K.; Brush, T.A. Embodied learning for computational thinking in early primary education. *J. Res. Technol. Educ.* **2022**, 1–21. [[CrossRef](#)]
- Wang, X.C.; Choi, Y.; Benson, K.; Eggleston, C.; Weber, D. Teacher’s role in fostering preschoolers’ computational thinking: An exploratory case study. *Early Educ. Dev.* **2021**, *32*, 26–48. [[CrossRef](#)]
- Gerosa, A.; Koleszar, V.; Tejera, G.; Gómez-Sena, L.; Carboni, A. Cognitive abilities and computational thinking at age 5: Evidence for associations to sequencing and symbolic number comparison. *Comput. Educ. Open* **2021**, *2*, 100043. [[CrossRef](#)]
- Voogt, J.; Fisser, P.; Good, J.; Mishra, P.; Yadav, A. Computational thinking in compulsory education: Towards an agenda for research and practice. *Educ. Inf. Technol.* **2015**, *20*, 715–728. [[CrossRef](#)]
- Fields, D.; Lui, D.; Kafai, Y.; Jayathirtha, G.; Walker, J.; Shaw, M. Communicating about computational thinking: Understanding affordances of portfolios for assessing high school students’ computational thinking and participation practices. *Comput. Sci. Educ.* **2021**, *31*, 224–258. [[CrossRef](#)]
- Grover, S.; Pea, R. Computational Thinking in K-12: A Review of the State of the Field. *Educ. Res.* **2013**, *42*, 38–43. [[CrossRef](#)]

22. Kalelioglu, F.; Gulbahar, Y.; Kukul, V. A Framework for Computational Thinking Based on a Systematic Research Review. *Balt. J. Mod. Comput.* **2016**, *4*, 583–596.
23. Tang, X.; Yin, Y.; Lin, Q.; Hadad, R.; Zhai, X. Assessing computational thinking: A systematic review of empirical studies. *Comput. Educ.* **2020**, *148*, 103798. [[CrossRef](#)]
24. Tsarava, K.; Moeller, K.; Román-González, M.; Golle, J.; Leifheit, L.; Butz, M.V.; Ninaus, M. A cognitive definition of computational thinking in primary education. *Comput. Educ.* **2022**, *179*, 104425. [[CrossRef](#)]
25. Zhang, Y.; Luo, R.; Zhu, Y.; Yin, Y. Educational Robots Improve K-12 Students' Computational Thinking and STEM Attitudes: Systematic Review. *J. Educ. Comput. Res.* **2021**, *59*, 1450–1481. [[CrossRef](#)]
26. Liao, Y.K.C.; Bright, G.W. Effects of Computer Programming on Cognitive Outcomes: A Meta-Analysis. *J. Educ. Comput. Res.* **1991**, *7*, 251–268. [[CrossRef](#)]
27. Basu, S.; Kinnebrew, J.S.; Biswas, G. Assessing student performance in a computational-thinking based science learning environment. In Proceedings of the Intelligent Tutoring Systems: 12th International Conference, ITS 2014, Honolulu, HI, USA, 5–9 June 2014; Proceedings 12; Springer: Berlin/Heidelberg, Germany, 2014; pp. 476–481.
28. Doleck, T.; Bazalais, P.; Lemay, D.J.; Saxena, A.; Basnet, R.B. Algorithmic thinking, cooperativity, creativity, critical thinking, and problem solving: Exploring the relationship between computational thinking skills and academic performance. *J. Comput. Educ.* **2017**, *4*, 355–369. [[CrossRef](#)]
29. El-Hamamsy, L.; Zapata-Cáceres, M.; Martín-Barroso, E.; Mondada, F.; Zufferey, J.D.; Bruno, B.; Román-González, M. The competent Computational Thinking test (cCTt): A valid, reliable and gender-fair test for longitudinal CT studies in grades 3–6. *arXiv* **2023**, arXiv:2305.19526.
30. Relkin, E.; de Ruitter, L.; Bers, M.U. TechCheck: Development and validation of an unplugged assessment of computational thinking in early childhood education. *J. Sci. Educ. Technol.* **2020**, *29*, 482–498. [[CrossRef](#)]
31. Chen, G.; Shen, J.; Barth-Cohen, L.; Jiang, S.; Huang, X.; Eltoukhy, M. Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Comput. Educ.* **2017**, *109*, 162–175. [[CrossRef](#)]
32. Ortega-Ruipérez, B.; Asensio, M. Evaluar el pensamiento computacional mediante resolución de problemas: Validación de un instrumento de evaluación. *Rev. Iberoam. Evaluación Educ.* **2021**, *14*, 153–171. [[CrossRef](#)]
33. Mühling, A.; Ruf, A.; Hubwieser, P. Design and first results of a psychometric test for measuring basic programming abilities. In Proceedings of the WIPSCÉ'15: Proceedings of the Workshop in Primary and Secondary Computing Education, London, UK, 9–11 November 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 2–10. [[CrossRef](#)]
34. Weintrop, D.; Wilensky, U. Using commutative assessments to compare conceptual understanding in blocks-based and text-based programs. In Proceedings of the ICER'15: Proceedings of the eleventh annual International Conference on International Computing Education Research, Omaha, NE, USA, 9–13 August 2015; Association for Computing Machinery, Inc.: New York, NY, USA, 2015; pp. 101–110. [[CrossRef](#)]
35. Kukul, V.; Karatas, S. Computational thinking self-efficacy scale: Development, validity and reliability. *Inform. Educ.* **2019**, *18*, 151–164. [[CrossRef](#)]
36. Yang, W.; Gao, H.; Jiang, Y.; Li, H. Beyond computing: Computational thinking is associated with sequencing ability and self-regulation among Chinese young children. *Early Child. Res. Q.* **2023**, *64*, 324–330. [[CrossRef](#)]
37. del Olmo-Muñoz, J.; Cózar-Gutiérrez, R.; González-Calero, J.A. Computational thinking through unplugged activities in early years of Primary Education. *Comput. Educ.* **2020**, *150*, 103832. [[CrossRef](#)]
38. Harper, F.K.; Caudle, L.A.; Flowers, C.E., Jr.; Rainwater, T.; Quinn, M.F.; Partnership, T.C. Centering teacher and parent voice to realize culturally relevant computational thinking in early childhood. *Early Child. Res. Q.* **2023**, *64*, 381–393. [[CrossRef](#)]
39. Uttal, D.H.; Cohen, C.A. Spatial thinking and STEM education: When, why, and how? In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 2012; Volume 57; pp. 147–181.
40. Delgoshaei, Y.; Delavari, N. Applying multiple-intelligence approach to education and analyzing its impact on cognitive development of pre-school children. *Procedia-Soc. Behav. Sci.* **2012**, *32*, 361–366. [[CrossRef](#)]
41. Oljayevna, O.; Shavkatovna, S. The Development of Logical Thinking of Primary School Students in Mathematics. *Eur. J. Res. Reflect. Educ. Sci.* **2020**, *8*, 235–239.
42. DeLoache, J.S.; Miller, K.F.; Pierroutsakos, S.L. *Reasoning and Problem Solving*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1998; pp. 801–850.
43. Muzaky, A.F.; Sunarno, W.; Harjana. Evaluating students logical thinking ability: TPACK model as a physics learning strategy to improve students logical thinking ability. *J. Phys. Conf. Ser.* **2020**, *1511*, 012027. [[CrossRef](#)]
44. Hawes, Z.; Moss, J.; Caswell, B.; Naqvi, S.; MacKinnon, S. Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention. *Cogn. Instr.* **2017**, *35*, 236–264. [[CrossRef](#)]
45. Ehrlich, S.B.; Levine, S.C.; Goldin-Meadow, S. The importance of gesture in children's spatial reasoning. *Dev. Psychol.* **2006**, *42*, 1259. [[CrossRef](#)] [[PubMed](#)]
46. Kesselring, T.; Müller, U. The concept of egocentrism in the context of Piaget's theory. *New Ideas Psychol.* **2011**, *29*, 327–345. [[CrossRef](#)]
47. Aristizábal Zapata, J.H.; Gutiérrez Posada, J.E. Collaborative Spatial Problem-Solving Strategies Presented by First Graders by Interacting with Tangible User Interface. In Proceedings of the HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, 24–29 July 2021; Proceedings, Part III 23; Springer: Berlin/Heidelberg, Germany, 2021; pp. 64–71.

48. Wing, J.M. Computational thinking and thinking about computing. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2008**, *366*, 3717–3725.
49. Ángel-Díaz, C.M.; Segredo, E.; Arnay, R.; León, C. Simulador de robótica educativa para la promoción del pensamiento computacional. *Rev. Educ. Distancia (RED)* **2020**, *20*, 63. [[CrossRef](#)]
50. Skjong, R.; Wentworth, B. Expert Judgment and Risk Perception, 2001. In Proceedings of the Eleventh (2001) International Offshore and Polar Engineering Conference, Stavanger, Norway, 17–22 June 2001; pp. 537–544.
51. Grant, J.S.; Davis, L.L. Selection and use of content experts for instrument development. *Res. Nurs. Health* **1997**, *20*, 269–274. [[CrossRef](#)]
52. Penfield, R.D.; Giacobbi, P.R., Jr. Applying a score confidence interval to Aiken's item content-relevance index. *Meas. Phys. Educ. Exerc. Sci.* **2004**, *8*, 213–225. [[CrossRef](#)]
53. Norman, D.A. *La Psicología de los Objetos Cotidianos*; Editorial Nerea: Donostia, Spain, 1998; Volume 6.
54. Castaño-García, J. Una aproximación al proceso de comprensión de los numerales por parte de los niños: Relaciones entre representaciones mentales y representaciones semióticas. *Univ. Psychol.* **2008**, *7*, 895–908.
55. Escudero, E.B.; Reyna, N.L.; Morales, M.R. Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *REDIE Rev. Electrónica Investig. Educ.* **2000**, *2*, 1.
56. Castro, A. El dibujo en la escuela. *Rev. Digit. Innov. Exp. Educ.* **2010**. Available online: [https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero\\_26/ANA\\_BELEN\\_MAESTRE\\_CASTRO\\_01.pdf](https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero_26/ANA_BELEN_MAESTRE_CASTRO_01.pdf) (accessed on 6 April 2024).
57. Blacker, D. Endicott, J. Psychometric properties: Concepts of reliability and validity. In *Handbook of Psychiatric Measures*, 2nd ed.; APA: Washington, DC, USA, 2002; pp. 7–14.
58. Muñoz Fernández, J. *Teoría Clásica de Los Tests*; Pirámide: Rome, Italy, 2003.
59. Field, A.; Miles, J.; Field, Z. *Discovering Statistics Using R*; SAGE Publications Ltd.: Southend Oaks, CA, USA, 2012.
60. Chang, L.C.; Lin, W.C. Improving Computational Thinking and Teamwork by Applying Balanced Scorecard for Sustainable Development. *Sustainability* **2022**, *14*, 11723. [[CrossRef](#)]
61. Cerdán, F. El paradigma agrícola. Personal communication, 2008.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.