Assessing Computational Thinking: The Relation of Different Assessment Instruments and Learning Tools

Vaida Masiulionytė-Dagienė[✉] and Tatjana Jevsikova[✉]

Vilnius University Institute of Data Science and Digital Technologies, Akademijos street 4, 08412 Vilnius, Lithuania
{vaida.masiulionyte-dagiene,tatjana.jevsikova}@mif.vu.lt

Abstract. The relevance of computational thinking as a skill for today's learners is no longer in question, but every skill needs an assessment system. In this study, we analyze two validated instruments for assessing computational thinking - the CTt (Computational Thinking Test) and the CTS (Computational Thinking Scale). The study involved 49 students in grades 8 and 9 (age 14–16). Prior to the study, students in both grades were taught computational thinking differently. One group learned computational thinking by completing tasks and creating projects in Scratch, the other group learned by completing tasks in "Minecraft: Education Edition". The students were asked to take the CTt and CTS tests. The nature of these tests is different, one is computational thinking diagnostic tool, the other is a psychometric self-assessment test consisting of core abilities (subconstructs) important for computational thinking. The aim of this study was to determine how these tests related to each other and whether students' gender and the different tools chosen to teach computational thinking had an impact on the level of computational thinking knowledge and abilities acquired based on the tests. The results have shown that the scores of the two tests correlated with each other only for male students' subgroup. For a whole group CTt scores correlated only with CTS algorithmic thinking subconstruct. The results have also shown that teaching tools do have an impact on the acquisition of different computational thinking concepts skills: students taught with different tools had different test results. This study provides useful implications on computational thinking teaching improvement and its assessment better understanding.

Keywords: Computational thinking · Assessment · Computational thinking assessment instruments · CTt · CTS · Gender differences · Learning tools

1 Introduction and Background

Computational thinking education is an important component in the process of digitalization of society and the economy, as discussed in a recent study organized by the European Commission [2]. The European Commission encourages this focus by making computational thinking education a priority in order to improve digital skills and

competences as part of the digital transformation. This study highlights that the transfer of computational thinking into educational subjects is a new field, which poses many challenges that need to be assessed in educational practice.

Evaluation and measurement of results are important in educational practice when implementing different teaching methods and tools. In the absence of a single unified definition of computational thinking, researchers working in this area apply different definitions [19, 25, 26]. As the result, there is also no single tool to assess computational thinking. For the assessment of computational thinking, various methods and tools have been developed today, such as Dr. Scratch, Bebras tasks, Zoombinis, CTt, CTS, etc.

Román-González categorized the tools for assessing computational thinking into the following groups: diagnostic tools, summative tools, formative-iterative tools, data-mining tools, skill transfer tools, perceptions-attitudes scales, vocabulary assessment [20].

Dr. Scratch as formative-iterative tool automatically analyzes Scratch programming projects and also can be used to develop computational thinking [27, 28]. Dr. Scratch analyzes code based on these computational thinking concepts: abstraction and problem decomposition, logical thinking, synchronization, parallelism, algorithmic notions of flow control, user interactivity, data representation [16].

Bebras tasks are used as another computational thinking assessment tool, classified as skill transfer tool [4, 5, 14, 20]. It is mentioned that the Bebras challenge tasks refer to analytics or analytical thinking concept [14], but there is not yet a representative set of Bebras tasks that has been validated as an assessment instrument for computational thinking.

Zoombinis [1] is an award-winning educational game from the nineties that has been rebuilt for modern platforms. It is not only used for learning computational thinking, but also in recent years for computational thinking assessment as a data-mining assessment tool. In Zoombinis, all the players actions are logged and then analyzed for the purpose of learning or the assessment. Concepts that are assessed in Zoombinis: problem decomposition, pattern recognition, abstraction, algorithm design [21].

More approaches to computational thinking assessment appear in recent research, e.g., CT-cube, a framework for the design, realization, analysis, and assessment of computational thinking activities [17], data driven approaches based on students' artefacts [6].

CTt (computational thinking test) was developed and validated by Román-González et al. [19]. CTt consists of 28 questions, divided in 7 groups: basic directions and sequences; repeat times; repeat until; simple conditional; complex conditional; while conditional; simple functions [19]. The test focuses on middle school children (mainly for 12–14 years old, but it can be used from 5th to 10th grade). Computational concepts used in the test are aligned with the CSTA (Computer Science Teachers Association) Computer Science Standards for the 7th and 8th grade [3]. Guggemos [9] mentions the main computational thinking concepts that CTt covers: abstraction, decomposition, algorithms, and debugging. The CTt has some advantages, like the ability to be conducted in large groups in pre-test scenarios, allowing for early detection of students with high abilities (or special needs) for programming tasks; and the ability to collect quantitative

information before the evaluations of the effectiveness of curricula designed to foster computational thinking [19].

Another validated tool for assessing computational thinking independently of programming is the CTS (Computational Thinking Scales) test [13]. This test identifies the following components of computational thinking based on ISTE (International Society for Technology in Education): creativity, algorithmic thinking, critical thinking, problem solving and collaboration skills [12]. The specificity of this test is that it is a self-assessment instrument. In this respect, it is completely different from the CTt test, which is knowledge assessment test.

Since computational thinking is broader than programming as defined in The European Commission's Staff Working Document accompanying the Digital Education Action Plan 2021–2027 (DEAP) [7]: "Computational thinking, programming and coding are often used interchangeably in education settings, but they are distinct activities", therefore it is important to pay attention to computational thinking assessment tools that assess beyond programming or algorithmic skills. Commonly used validated tests for assessing computational thinking that are independent of programming languages are the CTt and CTS [10, 15, 18, 24]. Other tools mentioned above are associated with specific platforms for programming or gaming (e.g. Dr. Scratch, Zoombinis). The Bebras tasks are tool independent, but not yet validated as a set of tasks for the assessment of computational thinking. For this reason, the tool-independent tests mentioned above (CTt and CTS) were chosen for this study. Both CTt and CTS are presented to students in a form of tests (a set of questions/statements with a set of answer options), while allowing to assess computational thinking from different points of view.

Due to the complexity of computational thinking, it is also relevant to assess computational thinking in a complex way, using more than one tool or method [22, 28]. According to the beforementioned classification [20], the CTt test falls into the diagnostic tools group and the CTS test into the perceptions-attitudes scales group [20]. Guggemos et al. [10] also mention that a system of different assessments is essential because various computational thinking profiles can be identified using multifunctional methods. Guggemos et al. [10], in their research for the assessment of computational thinking, analyze these two tests: CTt and CTS as well. However, researchers do not consider the computational thinking teaching tools used, nor do they differentiate between students' computational thinking test scores according to their gender.

The *aim* of this study is to investigate the relationship between CTt and CTS tests in relation to students' gender and the tools used to develop computational thinking.

We pose the following *research questions*:

RQ1: Is there a relationship between students' CTt and CTS (including its subconstructs) scores?
RQ2: How students' CTt and CTS tests' results are associated with the learning tool used to develop CT and differ in gender groups?

The paper is structured as follows. First, we present learning methods and tools, describe respondents, instruments and data analysis methods. Next, we present the results of the study according to the research questions. Finally, we discuss our findings, describe limitations and provide directions for future research.

2   Methods

2.1   Learning Tools and Methodology

Students were taught computational thinking using Scratch and "Minecraft: Education Edition". The teaching tools were not a part of the study, the computational thinking knowledge was acquired during the regular computer science lessons using the above-mentioned tools. Each class was familiar with both tools, but the main tool for grade 8 was the Scratch platform, while grade 9 learned in the "Minecraft: Education Edition" environment. On the Scratch platform, the students had to do "open" tasks using computer science concepts, such as cycles, conditions: they wrote programs that draw different shapes, developed a project with self-created characters, environment and implemented a created scenario of interaction between the characters. On the "Minecraft: Education Edition" platform, students learnt from the pre-designed lessons based on CSTA and ISTE guidelines [3, 12]. They first completed the tasks from the block programming fundamentals lessons, based on the 5 lessons provided, and then were introduced to the basics of Python programming (also completing the tasks from the 5 lessons). One lesson in "Minecraft: Education Edition" required 1 or 2 academic lessons to complete all the activities. On average, both grades had 12–14 lessons using these tools. As all tasks were completed individually and there were no team tasks during this learning period, the concept of cooperativity was removed from the CTS test in this study.

2.2   Respondents

In total, 49 students (51% female and 49% male), studying in school grades 8 and 9 (aged 14–16), took part in the survey. There were 24 students of 8th grade (51%), learning computational thinking with Scratch as dominating tool, and 25 students of 9th grade (49%), learning with "Minecraft: Education edition" as a primary tool.

All respondents were informed of the purpose of the study and gave their free will consent to participate in the study.

2.3   Instruments

In this study, besides the questions on basic demographical information, the two following instruments were used.

CTt. A validated instrument, consisting of 28 questions [19]. CTt test is claimed to be unidimensional [10] although addressing 7 cognitive operations (4 items for each cognitive operation arranged in increasing difficulty direction): basic directions and sequences, loops repeat times, loops repeat until, if simple conditional statement, if/else complex conditional statement, while conditional, and simple functions. For each question, four answer options are suggested with only one correct. Each item is rated as 1 (correct) or 0 (incorrect).

CTS. A validated computational thinking assessment scale, originally consisting of creativity, algorithmic thinking, critical thinking, problem solving and cooperativity subconstructs, rated on a 5-point Likert scale [13]. In our study, we included all subconstructs of this scale except for cooperativity, as mentioned before.

2.4 Data Analysis

For the analysis of the collected data, quantitative methods were used. Data normality for a whole sample has been checked with Kolmogorov-Smirnov and Shapiro-Wilk tests. CTt scores were not normally distributed. Due to this reason as well as analysis involving relatively small subgroup analysis, we used distribution-free non-parametric measures:

- To compare differences between two independent samples, the Mann–Whitney U test was used, and η^2 was used as an effect size measure.
- To test the monotonous relationships between the pairs of variables, Spearman's rank correlations were used.

We computed scores of the tests and their parts as a sum of the item scores.

The reliability of CTS psychometric scale subconstructs was examined using Cronbach's Alpha. After evaluating subscale reliability, item 4 from problem solving subconstruct was dropped to improve subscale reliability. Cronbach's Alphas for scale subconstructs were satisfying ($\geq$0.7): 0.701 for creativity (8 items), 0.765 for algorithmic thinking (6 items), 0.725 for critical thinking (5 items), 0,703 for problem solving (5 items).

The significance level was set to $\alpha = 0.05$.

For the statistical analysis, IBM SPSS Statistics 28 software package and MS Excel were used.

3 Results

3.1 An Association of Students' CTt and CTS Results

In a whole group of students, the CTt scores ranged from 9 to 27 with mean scores of 21.2, while CTS scores ranged from 57 to 107 with mean value of 77.4. The descriptive statistics for the results of both tests in general and according to tests' subscales, are presented in Table 1.

Table 1. Descriptive statistics for test scores (N = 49).

Cognitive operation/construct	Score range	Min.	Max.	Mean	Std. deviation
CTt	0–28	9	27	21.2	4.5
Sequences	0–4	2	4	3.7	0.5
Loops (times)	0–4	0	4	3.5	0.8
Loops (until)	0–4	1	4	3.2	0.9
If (simple)	0–4	0	4	2.7	1.1
If (complex)	0–4	0	4	2.6	1.1
While	0–4	0	4	2.5	1.3
Functions	0–4	0	4	3.0	1.2
CTS	30–150	57	107	77.4	10.0
Creativity	8–40	19	39	28.2	4.0
Algorithmic thinking	6–30	10	24	17.4	3.5
Critical thinking	5–25	5	22	15.0	3.2
Problem solving	5–25	12	23	16.8	2.6

Table 2. Spearman's rank correlations between CTS and its subconstructs and CTt scores.

	CTS	Creativity	Algorithmic thinking	Critical thinking	Problem solving
ρ	0.174	0.000	0.307*	–0.034	0.212
p	0.232	1.000	0.032	0.817	0.143

* Significant at 0.05 level (2-tailed)

Spearman's rank correlations for 49 students have been calculated between CTt scores and CTS general scores as well as its subconstructs (Table 2).

Significant (at 0.05 level) relationship was found between CTt and algorithmic thinking scores (ρ = 0.307, p = 0.032). However, there was no significant association between CTt scores and CTS general results (ρ = 0.174, p = 0.232). Further analysis on differences between students' groups is presented in the following section.

3.2 CT Assessment Scores in Learning Tool and Gender Groups

In order to observe the relationships between CTt scores and CTS including its subconstruct scores between groups studying with different learning tool (Minecraft: Education Edition, Scratch) and between male and female students' subgroups, Spearman's rank correlations have been computed (Table 3). We use Minecraft as a shortened name form of "Minecraft: Education Edition".

Table 3. Spearman's rank correlations between CTS and its subconstructs and CTt scores for Minecraft and Scratch learning groups and gender.

Group	Measure	CTS	Creativity	Algorithmic thinking	Critical thinking	Problem solving
Minecraft	ρ	0.371	0.216	0.414*	0.206	0.259
(n = 25)	p	0.068	0.299	0.040	0.323	0.211
Scratch	ρ	0.285	0.103	0.301	0.128	0.307
(n = 24)	p	0.178	0.631	0.154	0.550	0.144
Males	ρ	0.445*	0.301	0.576**	0.119	0.276
(n = 24)	p	0.029	0.153	0.003	0.581	0.191
Females	ρ	–0.033	–0.188	0.137	–0.138	0.159
(n = 25)	p	0.877	0.367	0.514	0.510	0.447

* Significant at 0.05 level (2-tailed); ** significant at 0.01 level (2-tailed)

We found a significant monotonous relationship between CTt and CTS scores in a subgroup of male students ($\rho = 0.445$, p = 0.029). However, there were no correlations between scores in other subgroups (female students or subgroups based on learning tool). The strongest significant monotonous relationship was found between CTS's algorithmic thinking and CTt scores in a group of boys ($\rho = 0.576$, p = 0.003). In a group of students learning with Minecraft as a primary tool, this relationship was also significant, but weaker ($\rho = 0.414$, p = 0.040).

Graphically, the differences in CTt scores between groups studied are presented in Fig. 1.
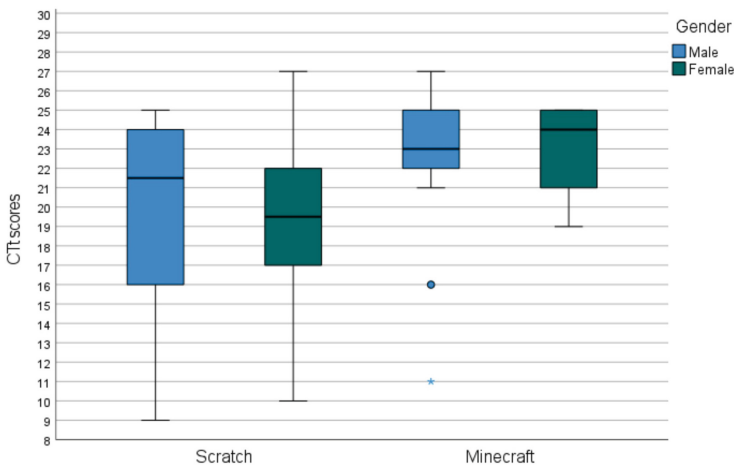
Fig. 1. CTt scores for Scratch and Minecraft groups' male and female students.

The mean ranks of scores for different test constructs and subconstructs in groups studied and results of the Mann-Whitney U tests are presented in Table 4 (light grey shading for significance at 0.05 level, dark grey for significance at 0.01 level).

Table 4. Differences between groups (Scratch, Minecraft, male and female): Mann-Whitney U tests' results.

Cognitive opera-tion/Construct	Mean rank		Z	p	Mean rank		Z	p
	Scratch	Minecraft			Males	Females		
CTt total scores	20.44	29.38	-2.20	0.028	25.67	24.36	-0.32	0.748
Sequences	24.40	25.58	-0.40	0.690	26.35	23.70	-0.89	0.371
Loops Times	24.23	25.74	-0.43	0.665	24.38	25.60	-0.35	0.725
Loops Until	21.38	28.48	-1.87	0.061	25.15	24.86	-0.08	0.940
If simple	19.67	30.12	-2.68	0.007	25.25	24.76	-0.13	0.900
If complex	23.21	26.72	-0.91	0.363	21.17	28.68	-1.95	0.052
While	20.52	29.30	-2.22	0.027	26.35	23.70	-0.67	0.502
Functions	24.27	25.70	-0.38	0.705	26.65	23.42	-0.86	0.392
CTS total scores	28.27	21.86	-1.57	0.116	26.56	23.50	-0.75	0.453
Creativity	28.35	21.78	-1.62	0.106	23.85	26.10	-0.55	0.581
Algorithmic thinking	25.71	24.32	-0.34	0.732	29.83	20.36	-2.34	0.020
Critical thinking	30.92	19.32	-2.86	0.004	25.10	24.90	-0.05	0.960
Problem solving	26.23	23.82	-0.60	0.551	24.71	25.28	-0.14	0.887

The Mann-Whitney U test confirmed a significant difference in CTt scores between the groups learning as a primary tool with Scratch (mean rank 20.44) and Minecraft (mean rank 29.38): Z = –2.20, p = 0, 028. An effect size $\eta^2 = 0.1$ denotes that 10% of variance in rank was accounted by the CT learning tool used (Scratch or Minecraft).

Significantly higher scores in critical thinking were observed in Scratch group compared to Minecraft (Z = –2,86, p = 0.004, $\eta^2 = 0.17$). Interquartile range of the differences in critical thinking scores, including subgroups of male and female students are presented graphically in Fig. 2.

Significant differences for Scratch and Minecraft groups were also found in scores of CTt cognitive operation "simple If" (Z = –2.68, p = 0.007, $\eta^2 = 0.15$) and "While loop" (Z = –2.22, p = 0.027, $\eta^2 = 0.1$).

Studying the differences between groups of boys and girls, significant differences were found in algorithmic thinking scores: Z = –2.34, p = 0.020, $\eta^2 = 0.12$, with mean rank for female students 20.36 and for male students 29.83.
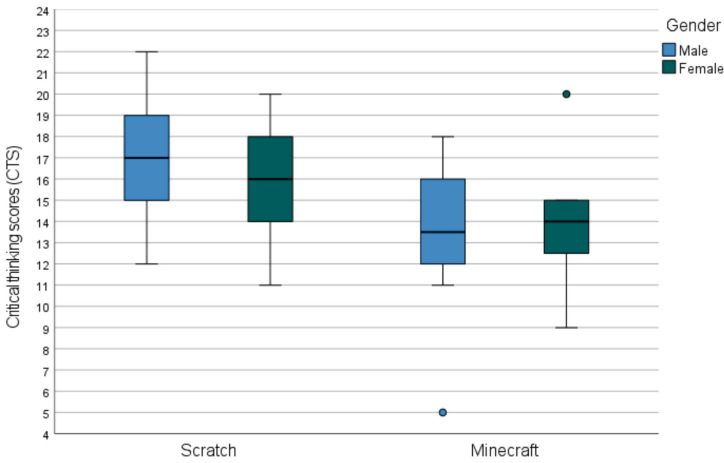
Fig. 2. Critical thinking (CTS) scores for Scratch and Minecraft groups' male and female students.

4 Discussion and Conclusion

In this study, we examined an association between the results gained by the two compu-
tational thinking assessment instruments and differences in scores based on groups of
learning tool used to develop computational thinking and gender.

4.1 Relationship Between Students' CTt and CTS Scores

Looking at the results regarding the first research question, there was no significant
relationship found between CTt and CTS general scores. This finding is in line with the
recent study by Guggemos et al. [10] and can be explained by different nature of the
instruments. However, it is interesting to note that analysis of the results of both tests in
separate groups of students by gender has shown that the tests' scores correlated with
each other in the group of male students.

Analysis, performed on the results for the individual subconstructs of the CTS test,
we also see monotonous positive relationship of the CTt scores with the algorithmic
thinking scores in CTS test, what supports the results of the study by Guggemos et al.
[10]. Thus, in response to the first research question, we can say that the tests do not
correlate from a generic point of view, but that the correlation is influenced by the gender
of the students, and to fully validate this statement, it would require research in a larger
group of students.

4.2 Differences in the CTt and CTS Tests' Results in Students' Groups
by Learning Tool and Gender

In response to the second research question, which asks how the test results were influ-
enced by the teaching tools, we can see that students who studied using "Minecraft:
Education Edition" had significantly better CTt test results than those students who

studied computational thinking in Scratch. These results could be explained by the fact that both "Minecraft: Education Edition" and CTt are based on the CSTA [3] teaching standards. The pre-designed lessons include the same elements as the CTt test (cycles, conditional sentences, etc.). This is further confirmed by the results of the separate CTt test groups, cognitive operations such as the "while loop" and the "simple if" conditions. However, students who had studied in the Scratch environment had better results on the critical thinking subconstruct of the CTt test. Critical thinking is defined as "the use of cognitive skills or strategies that increase the possibility of the desired behaviors" [11]. In the context of the definition of critical thinking, the results obtained can be explained by the fact that, unlike the pre-prepared lessons used in "Minecraft: Education Edition", in the Scratch environment, students had to create their own projects and find custom solutions to achieve the desired outcome. As we can see, different tools develop different computational thinking skills during the teaching process, and this should be considered when teaching computational thinking, so that the most versatile computational thinking skills can be developed and assessed with the widest possible range of assessment tools. There is also a need for more research in this area on which tools best develop which computational thinking skills.

In terms of gender, the significant difference was observed in the algorithmic thinking subconstruct of the CTS test. Boys showed higher scores in algorithmic thinking than girls. On a one hand side, this finding reflects the existing stereotype of computer science and engineering being more male-oriented field [15, 23]. While the findings of Ma et al. [15] study show that the CTS test algorithmic thinking scores before the intervention were slightly higher in the girls' group, after the intervention, the scores became identical, with a non-significant difference to the boys' advantage. As Groher et al. [8] mentions: "diversity among the students calls for diversity among the teaching and learning materials." This could be one of the reasons for the different tests results.

4.3 Limitations and Future Research Directions

The main limitation of this study was the relatively small sample size of students involved and obtained by the convenient sampling method. Also, the slightly different age group of the students (8th and 9th grade) might have had some influence on the results. Nevertheless, the results were in line with the findings of other related studies and provided interesting insights for further research with greater samples and other methods.

In addition to the test results, one trend was observed during the course of the study that would allow for improvements in the assessment. When taking the test, students used their hand or a computer mouse to guide the screen through the picture next to each test question in order to find the correct answer. However, this process was not logged anywhere and we only saw one of the selected answers as the test result. However, in order to assess computational thinking, we should assess the process of thinking itself. Such approach we may see in the Zoombinis game [21]. It might be possible that the student's thinking process was partly correct, e.g. right at the beginning with a slight mistake at the end, resulting in the wrong answer, but this is not what we see as a test score for diagnostic test. Such approach could also help to eliminate the cases when the student clicked the right answer by chance. In future research, we will focus on how to better assess the process thinking for the task solution, and not just the final result.

References

1. Asbell-Clarke, J., et al.: The development of students' computational thinking practices in elementary- and middle-school classes using the learning game. Zoombinis. Comput. Hum. Behav. **115**, 106587 (2021)

2. Bocconi, S., et al.: Reviewing computational thinking in compulsory education. In: Inamorato Dos Santos, A., Cachia, R., Giannoutsou, N., Punie, Y. (eds.) Publications Office of the European Union, Luxembourg (2022). https://doi.org/10.2760/126955. ISBN 978-92-76-47208-7, JRC128347

3. CSTA: K12 computer science standards (2017). https://www.csteachers.org/page/about-csta-s-k-12-nbsp-standards

4. Djambong, T., Freiman, V., Gauvin, S., Paquet, M., Chiasson, M.: Measurement of computational thinking in K-12 education: the need for innovative practices. In: Sampson, D., Ifenthaler, D., Spector, J., Isaías, P. (eds.) Digital Technologies: Sustainable Innovations for Improving Teaching and Learning, pp. 193–222. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73417-0_12

5. Dolgopolovas, V., Jevsikova, T., Dagiene, V., Savulioniene, L.: Exploration of computational thinking of software engineering novice students based on solving computer science tasks. Int. J. Eng. Educ. **32**(3), 1–10 (2016)

6. Eloy, A., Achutti, C.F., Cassia, F., Deus Lopes, R.: A data-driven approach to assess computational thinking concepts based on learners' artifacts. Inf. Educ. **21**(1), 33–54 (2022)

7. European Commission: Digital Education Action Plan 2021–2027: Resetting education and training for the digital age (2020). https://eurlex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0624

8. Groher, I, Sabitzer, B., Demarle-Meusel, H., Kuka, L., Hofer, A.: Work-in-progress: closing the gaps: diversity in programming education. In: 2021 IEEE Global Engineering Education Conference (EDUCON), pp. 1449–1453 (2021)

9. Guggemos, J.: On the predictors of computational thinking and its growth at the high-school level. Comput. Educ. **161**, 104060 (2021)

10. Guggemos, J., Seufert, S. Román-González, M.:Computational thinking assessment – towards more vivid interpretations. Tech. Know. Learn.https://doi.org/10.1007/s10758-021-09587-2

11. Halpern, D.F.: Thoughts and Knowledge: An Introduction to Critical Thinking. Lawrence Erlbaum Associates, New Jersey-London (1996)

12. ISTE: Computational thinking: leadership toolkit (2015). https://www.iste.org/computational-thinking

13. Korkmaz, Ö., Çakir, R., Özden, M.Y.: A validity and reliability study of the computational thinking scales (CTS). Comput. Hum. Behav. **72**, 558–569 (2017)

14. Labusch, A., Eickelmann, B.: Computational thinking competences in countries from three different continents in the mirror of students' characteristics and school learning. In: Kong, S.C., et al., (eds.) Proceedings of International Conference on Computational Thinking Education 2020, pp. 2–7. The Education University of Hong Kong (2020)

15. Ma, H., Zhao, M., Wang, H., Wan, X., Cavanaugh, T.W., Liu, J.: Promoting pupils' computational thinking skills and self-efficacy: a problem-solving instructional approach. Educ. Tech. Res. Dev. **69**(3), 1599–1616 (2021). https://doi.org/10.1007/s11423-021-10016-5

16. Moreno-León, J., Robles, G., Román-González, M.: Dr. Scratch: automatic analysis of scratch projects to assess and foster computational thinking. RED-Rev. Educ. Distancia **46**, 1–23 (2015)

17. Piatti, A., et al.: The CT-cube: a framework for the design and the assessment of computational thinking activities. Comput. Hum. Behav. Rep. **5**, 100166 (2022)

18. Poulakis, E., Politis, P.: Computational thinking assessment: literature review. In: Tsiatsos, T., Demetriadis, S., Mikropoulos, A., Dagdilelis, V. (eds.) Research on e-Learning and ICT in Education, pp. 111–128. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-643 63-8_7

19. 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. **72**, 678–691 (2017)

20. Román-González, M., Moreno-León, J., Robles, G.: Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In: Kong, S.-C., Abelson, H. (eds.) Computational Thinking Education, pp. 79–98. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-6528-7_6

21. Rowe, E., et al.: Assessing implicit computational thinking in Zoombinis puzzle gameplay. Comput. Hum. Behav. **120**, 106707 (2021)

22. Statter, D., Armoni, M.: Teaching abstraction in computer science to 7th grade students. ACM Trans. Comput. Educ. **20**(1), 8–837 (2020)

23. Stupurienė, G., Jevsikova, T., Juškevičienė, A.: Solving ecological problems through physical computing to ensure gender balance in STEM education. Sustainability **14**(9), 4924 (2022)

24. Sun, L., Hu, L., Zhou, D., Yang, W.: Evaluation and developmental suggestions on undergraduates' computational thinking: a theoretical framework guided by Marzano's new taxonomy. Interact. Learn. Environ. (2022). https://doi.org/10.1080/10494820.2022.2042311

25. Tang, X., Yin, Y., Lin, Q., Hadad, R., Zhai, X.: Assessing computational thinking: a systematic review of empirical studies. Comput. Educ. **148**, 103798 (2020)

26. Tikva, C., Tambouris, E.: Mapping computational thinking through programming in K-12 education: a conceptual model based on a systematic literature Review. Comput. Educ. **162**, 104083 (2021)

27. Troiano G., et al.: Is my game OK Dr. Scratch? Exploring programming and computational thinking development via metrics in student-designed serious games for STEM. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children Association for Computing Machinery, New York, NY, USA, pp. 208–219 (2019)

28. Wei, X., Lin, L., Meng, N., Tan, W., Kong, S.-C., Kinshuk.: The effectiveness of partial pair programming on elementary school students' computational thinking skills and self-efficacy. Comput. Educ. **160**, 104023 (2021)