Andragoške studije, issn 0354–5415, broj 11, decembar 2024, str. 55–83
© Institut za pedagogiju i andragogiju; Originalni naučni rad
UDK 374.72:004.9(495)
doi: 10.5937/AndStud2402055K

Kotzampasaki A. Evangelia[1], Koulaouzides A. George[2]

Computational Thinking in Adult Education: Learning Outcomes of a STEM Intervention[3]

Abstract: Computational Thinking (CT) has attracted increasing international attention, particularly in relation to younger learners. However, as a fundamental cognitive process, CT is equally relevant to adults, and its development within this demographic warrants further investigation. STEM education, with its holistic and transdisciplinary orientation, is closely associated with CT and, if appropriately designed, can serve as a powerful framework for adult education. This paper reports on an action research project that explored the dimensions of CT in the context of STEM-based adult education. For this purpose, a STEM intervention comprising eight interconnected scenarios was designed and implemented in Second Chance Schools (SCSs), whose students exemplify the essential characteristics of adult learners. A total of 48 trainees, aged 26 to 75, across four classes in two SCSs in Athens, participated in the intervention. Data was collected through pre... and post-tests and analyzed quantitatively, supplemented by qualitative observations to support interpretation. The results revealed statistically significant improvements across all levels of difficulty and dimensions examined. Specifically, gains were observed in abstraction, algorithmic thinking, decomposition, generalization, and evaluation.

Keywords: Adult Education, Computational Thinking, Second Chance Schools, STEM Education

[1] Hellenic Open University, Greece, https://orcid.org/0009-0006-9813-6273 (kotzampasaki.evangelia@ac.eap.gr)

[2] Hellenic Open University, Greece, https://orcid.org/0000-0003-4435-9800 (gkoulaouzidis@eap.gr)

[3] The research presented in this paper is based on the ongoing doctoral research dissertation of E. A. Kotzampasaki, which is supervised by G. A. Koulaouzides.

Računarsko razmišljanje u obrazovanju odraslih: ishodi učenja STEM intervencije[4]

Apstrakt: Na međunarodnom nivou se sve više pažnje posvećuje računarskom razmišljanju (u daljem tekstu: RR), posebno mlađih učenika. Međutim, RR je kao osnovni kognitivni proces podjednako važno i za odrasle, pa se njegov razvoj mora istraživati i u toj demografskoj grupi. Zahvaljujući svom holističkom i transdisciplinarnom pristupu, STEM obrazovanje je usko povezano sa RR i može biti idealno za obrazovanje odraslih, ukoliko se adekvatno osmisli. U ovom radu je predstavljeno akciono istraživanje u kojem su ispitane dimenzije RR u kontekstu STEM obrazovanja odraslih. Za potrebe istraživanja je osmišljena STEM intervencija koja se sastojala od osam međusobno povezanih scenarija. Intervencija je sprovedena u školama druge šanse (ŠDŠ) jer se njihovi polaznici odlikuju ključnim karakteristikama odraslih učenika. U intervenciji je učestvovalo ukupno 48 polaznika, uzrasta od 26 do 75 godina, koji su pohađali četiri odeljenja u dve ŠDŠ u Atini. Podaci prikupljeni pre i nakon intervencije kvantitativno su analizirani, dok su kvalitativna zapažanja iskorišćena kao pomoć u tumačenju rezultata. Statistički značajna poboljšanja su uočena nakon STEM intervencije na svim proučavanim nivoima težine i dimenzijama. Uočen je napredak u apstrakciji, algoritamskom razmišljanju, razlaganju, generalizaciji i evaluaciji.

Ključne reči: obrazovanje odraslih, računarsko razmišljanje, škole druge šanse, STEM obrazovanje

Introduction

Computational thinking (CT) is a crucial skill for adult learners, as it fosters problem-solving abilities that are applicable in various professional and everyday contexts. CT equips adults with the tools to approach complex challenges systematically, promoting skills like decomposition, abstraction, and algorithmic thinking (Yadav et al., 2017). Although adult learners need to develop or improve their computational thinking skills, challenges often arise due to the absence of suitable educational methods that consider their specific needs and motivations (El Mawas et al., 2021).To address the aforementioned challenges, we designed a small-scale STEM intervention for adult learners enrolled in Second Chance Schools (SCS) in the Prefecture of Athens, Greece. The intervention was informed by both adult education principles and the pedagogical foundations of STEM education.

[4] Istraživanje predstavljeno u ovom radu zasnovano je na istraživanju koje E. A. Kotzampasaki sprovodi u okviru rada na svojoj doktorskoj disertaciji pod mentorstvom G. A. Koulaouzides.

Computational Thinking

Computational Thinking (CT) is conceptualized as a cognitive process that facilitates the automatization of certain aspects of human thought (Selby & Woollard, 2014). Wing (2006) pointed out that CT, as a process based on the fundamental concepts of computer science, might contribute as a way of thinking and approaching several critical issues in problem solving, and might lead us in understanding both human behavior and systems design, which is why CT must be recognized as one of the fundamental life skills. Some years later, Wing (2017) explained that computational thinking was "the thought processes involved in formulating a problem and expressing its solution(s) in such a way that a computer—human or machine—can effectively carry out."

Although there is no universally accepted definition of CT, X. Tang and associates (2020) attempted to categorize some of the most common definitions. They distinguish between those emphasizing programming or computer science concepts (Brennan & Resnick, 2012; Denner et al., 2012; Weintrop et al., 2016) and those focusing on solving real-world problems through a combination of knowledge and methods from various fields (CSTA & ISTE, 2011; Selby & Woollard, 2014; Yadav et al., 2014). Moreover, in a more comprehensive approach, Denning and Tedre (2019, p. 4) explained that CT "is the mental skills and practices for (a) designing computations that get computers to do the jobs for us and (b) explaining the interpreting the world as a complex information processes […] the explanation aspect reflects the science tradition of computing in which people seek to understand how computation works and how it shows up in the world." However, in addition to efforts to define computational thinking, considerable debates about the main CT dimensions are under way in the research and educational community. Ling, et al. (2018) analyzed the relevant literature and concluded it by separating CT dimensions from CT skills. They opted for the five CT dimensions developed by Selby & Woollard (2014) that have received considerable acceptance:

- **Abstraction**: According to Wing (2011), abstraction is a crucial, high-level thinking process in computational thinking, playing a key role in identifying common properties and excluding non-essential data, thereby enhancing the ability to manage complexity. In other words, abstraction serves as a fundamental tool for handling complexity, enabling the development and understanding of simpler models to achieve a clearer comprehension of complex phenomena (Cook et al., 2012).
- **Algorithmic Thinking**: According to Hu (2011), computational thinking involves a step-by-step determination of a functional process. He

points out that individuals who study models and algorithms develop skills that can enhance critical thinking and benefit from them on multiple levels.

- **Decomposition**: Decomposition involves an initial assessment of a problem's complexity by breaking it down into smaller, more manageable components, making it easier to solve (Djambong & Freiman, 2016). It is about "…finding structure in the problem and determining how the various components will fit together in the final solution" (Csizmadia et al., 2019, p. 45).
- **Generalization**: This dimension involves recognizing the value of a solution in solving similar or other real-world problems, as well as reusing the solution either as part of another solution or in its original form (Csizmadia et al., 2019; National Research Council, 2011; Selby & Woollard, 2014).
- **Evaluation**: This dimension of computational thinking involves implementing, representing and assessing previous phases of problem-solving efforts (Fogli et al., 2017). Evaluation includes exploring alternatives, comparing options, and considering how these alternatives might function in real-life situations (Csizmadia et al., 2019).

The aforementioned dimensions of computational thinking have been applied in various research studies (Csizmadia et al., 2019; Ling et al., 2018; X. Tang et al., 2020) and form the basis of our approach in this study. At this point, the following question arises: "How can we actually foster computational thinking?". In our opinion, CT-STEM may be an answer.

Computational Thinking in STEM

Computational Thinking in STEM (CT-STEM) refers to the development of computational thinking skills through STEM education. It's important to recognize that

> STEM (Science, Technology, Engineering and Mathematics) education is not merely the accumulation of knowledge across scientific fields, but rather a holistic, transdisciplinary approach that moves beyond isolated specializations. It emphasizes the study of connections, focuses on interrelated processes, and does not concentrate on individual phenomena (Psycharis et al., 2018 as cited in Kotzampasaki & Koulaouzides, 2024, p. 183).

STEM education fosters the development of scientific, mathematical and technological practices, concepts, and insights, which enable the resolution of real-world challenges and complex questions (Man et al., 2016). Integrating computational thinking into STEM education promotes interconnectedness across scientific fields and enhances math and science lessons to reflect current professional practices (X. Tang et al., 2020). Swanson, et al. (2019) analyzed interviews with CT-STEM researchers and proposed a taxonomy with four key CT-STEM practices: data handling, simulation and modeling, systems thinking, and computational problem-solving. Research shows a positive impact of both STEM education and game-based learning on computational thinking; however, as discussed in the following paragraph, most of these studies have been conducted with younger students (Fidai et al., 2020; Lu et al., 2023; Ma et al., 2023).

Computational Thinking Evaluation: Target Populations and Pools

Interest in computational thinking (CT) has grown significantly among researchers, and recent years have seen the publication of meta-analyses on the topic. Literature reviews and meta-analyses, such as those by X. Tang and associates (2020) including 96 studies; Cutumisu and associates (2019), reviewing 39 empirical studies; and Poulakis and Politis (2021), examining 82 studies, all reach a common conclusion: most CT research focuses on primary and secondary school students, with limited work addressing adult learners. In the few studies involving adults, the participants were typically college students.

Another notable finding is that most CT evaluations focus narrowly on computing or programming skills, while CT itself encompasses a broader set of cognitive skills. X. Tang and associates (2020) propose several directions for future research: a) conducting more studies with older age groups, such as high school students, college students, vocational trainees, and learners in non-traditional educational settings; b) basing studies on clear CT definitions and exploring connections across different thematic areas; c) incorporating qualitative data like interviews or focus groups to deepen understanding of CT; d) ensuring reliability and validity in assessments; e) recognizing the unique aspects of CT that distinguish it from programming or computing; and f) designing CT assessment tools adaptable across various platforms to allow comparisons. Poulakis and Politis (2021) echo these recommendations, noting the lack of independent, validated evaluation tools for all age groups and advocating multi-method evaluations that include quantitative research as the most effective approach. A critical issue in CT research is the need for reliable assessment tools to measure whether CT skills are developed or improved. CT assessments can generally be categorized by their approach: a) assessments based on specific programming platforms, b) psychometric or scale-based tools, and c) multi-method approaches (Poulakis & Politis, 2021).

In our study, we focus on the internationally recognized Bebras CT competition (see bebras.org) to collect our main quantitative data, and we explain our choice in detail. Bebras questions are designed to assess cognitive skills useful in solving problems across various domains (X. Tang et al., 2020) and are based on the application of CT skills in daily life (Román-González et al., 2019). X. Tang and associates (2020) note that Bebras is widely used and valued by researchers due to its distinct advantages: a) no electronic devices are required; b) no prior knowledge of computers or programming is necessary; c) the questions are designed as riddles, brainteasers, or logic puzzles; d) the questions are available at varying difficulty levels; and e) each question is mapped to specific CT dimensions, aligning with Selby and Woollard's (2014) CT framework. Research suggests that Bebras questions provide a high-quality basis for CT assessment (Dagiene & Stupuriene, 2016), while psychometric analyses have been conducted to further confirm their validity (Araujo et al., 2019; Hubwieser & Mühling, 2015). Although Bebras competitions are primarily aimed at students under 18, researchers, such as López and García-Peñalvo (2016), have successfully adapted Bebras questions to assess CT in higher education students. Similarly, Lockwood and Mooney (2018) analyzed Bebras tasks and concluded that they were appropriate for college-level students as well.

CT and Adult Learners

Research on computational thinking (CT) specifically related to the general adult population remains limited (Kotzampasaki & Koulaouzides, 2024). A published content analysis of relevant research verified that there is only very limited focus on adult learners aside from educators or higher education students in the 336 studies published between 2006 and 2018 (K. Y. Tang et al., 2020). However, higher education students and undergraduate students, in particular, may not be generally classified as "adult learners" given the common understanding that they do not necessarily exhibit the key characteristics of adult learners, since adulthood is a complex concept linked to self-definition, social roles and social acceptance (Hill et al., 2023; Koulaouzides, 2019).

We should, however, take into consideration findings from research that examined computational thinking beyond schooling. For example, Lafuente-Martínez and associates (2022) attempted to create a CT test for adults, conducting reliability verification, validity, and factor analysis to develop a one-dimensional model focusing on algorithmic thinking (Algorithmic Thinking Test for Adults ... ATTA). However, we were unable to identify any published studies that utilized this tool. Moreover, Zapata-Rivera et al. (2019) studied ten educators and ten adult learners and made recommendations for programming training to support digital learning tools like alarms and notifications. This study revealed that adults primarily need support and training in CT aspects that meet lifelong learning demands, such as practical skills and time management, rather than computer use or programming (Zapata-Rivera et al., 2019). El Mawas et al. (2021) tested the adaptive educational game "AutoThinking" on 12 postgraduate students in a digital learning master's program. The "AutoThinking" labyrinth game uses flow charts and block programming; the study included questions assessing CT knowledge, such as "What is debugging?" and "What is a sequence?". The results suggest that the AutoThinking game effectively increased the students' knowledge of concepts related to computational thinking and computing (El Mawas et al., 2021). Moreover, Gao (2020) studied students in higher vocational education in China, finding that game-based learning and CT improved their learning interest and engagement; he also highlighted the need for studies involving older age groups. Similarly, Kazimoglu (2020) examined the effectiveness of a "serious game" for 151 first-year computer science students, noting its positive impact on motivation to learn programming and confidence in using CT skills. Serious games are "digital games and equipment with an agenda of educational design and beyond entertainment" (Park et al., 2012, as cited in Pliasa & Fachantidis, 2021, p. 619) and are utilized in educational research, such as that by Pliasa and Fachantidis (2021), because of their potential to develop skills, achieve cognitive objectives and modify attitudes, benefiting both typical and autistic students.

Methodology: Choices and Concerns

Research Gap and Research Questions

The above-mentioned studies revealed a significant gap in CT studies focusing on adult learners, particularly those without prior computing or programming knowledge. Therefore, we decided to focus on adult learners with low qualifications and examine what happens in formal education environments when they are given a second opportunity to participate in education (Cutumisu et al., 2019; Kotzampasaki & Koulaouzides, 2024; Ortiz et al., 2023; Poulakis & Politis, 2021; X. Tang et al., 2020). Our decision was also informed by the work of Ortiz and associates (2023), who point out the importance of research related to ways of reducing the digital divide among adults participating in basic literacy programs.

In Greece, this opportunity is offered to the general adult population in the Second Chance Schools (SCSs). The SCSs in Greece are attended by adults, who had dropped out of school before completing their nine-year compulsory education. SCSs aim to support the learners' cultural, economic and social development, and self-confidence to help them enter or advance in the workforce by teaching them essential skills (Koutouzis et al., 2023). SCS students generally exhibit the main characteristics of adult learners; they face constraints such as time limitations, work, and family responsibilities, wherefore the SCS curricula address their diverse needs by employing adaptable educational tools to make education a multidimensional endeavor. Our main research question related to the research presented in this paper is:

- What is the impact of a designed STEM didactic intervention on the dimensions of computational thinking on Second Chance School adult students?

and more specifically:

- What is the impact of the STEM intervention on each of five CT dimensions (abstraction, decomposition, algorithmic thinking, evaluation, generalization)?
- What is the impact of the STEM intervention on CT's varying levels of difficulty (easy, medium, difficult)?

Procedure

To answer the research questions, we conducted a mixed-method action research study. We designed a STEM intervention featuring a sequence of interconnected educational scenarios centered on common transdisciplinary concepts. The intervention included seminars with a variety of activities, such as educational robotics, physical computing, micro-construction, experiments and simulations, alongside diverse teaching techniques. For instance, participants worked with the Edison robot, micro:bit, and various extensions, such as motors, sensors and LEDs, while also engaging in chemistry experiments and simulations, in order to explore the concept of "color" from multiple perspectives. Detailed information on the design and content of the STEM intervention will be published in a separate article.

Our STEM intervention was implemented four times at the two SCSs in Greece. The selection of the SCSs was guided by their expressed willingness to participate and the practical feasibility of integrating the intervention in their curriculum. We secured the necessary permissions in communication with these schools and ensured that the study was implemented in a manner that respected their schedules and supported their smooth operation.

Participants first completed the Bebras CT pretest that is explained in the *Instruments* section. This was followed by the implementation of the intervention consisting of eight STEM educational scenarios. The seminars were generally three hours long and conducted once a week, unless there were scheduling conflicts with the SCSs, such as participation in other activities or an educational visit, in which case they were held bi-weekly or as two-hour sessions. Some of the didactic scenarios required less time and it was possible to implement two of them within a three-hour seminar. In each group, the educational intervention required between 16 to 19 teaching hours, with variations due to the available time in the SCSs, adjustments to the specific needs of each group, the number of learners, possible absences and the time required for brief repetitions. Implementation of the intervention lasted approximately three months in each SCS. Two weeks after the intervention was completed, the participants completed the post test, along with additional questionnaires addressing various research questions. To accommodate reading difficulties faced by some SCS adult learners, we read all the questions out loud during the process.

Our experimental research primarily combined quantitative correlation and descriptive methods through pretests and posttests. We also included qualitative data from observations during the intervention and responses to CT tests, documented in a research diary.

Population and Sample

Forty-eight adult learners aged 26...75 participated voluntarily in our interventions. At the start, we observed that most (41.7%) belonged to the 40...49 age group, followed by 33.3% in the 50...59 age group; a smaller proportion (20.8%) were in the 30...39 age group, one participant (2.1%) belonged to the 18...29 age group, while another (2.1%) belonged to the 70+ age group. The estimated average age of the trainees stood at approximately 45.85 years, which may differ from the actual average, which could only be determined if we had exact age data for each participant. Twenty-seven (56.25%) of the participants were women and 21 were men (43.75%). While the findings may not be broadly generalized to all populations, they may be applicable to those with similar sociocultural characteristics, warranting further investigation in comparable settings.

Instruments

To assess the CT dimensions, we designed a questionnaire based on the Bebras CT competition. We selected questions from the United Kingdom's 2015...2018 contests, which were developed in collaboration with Oxford University and Google, with contributions from global experts. Table 1 provides a list of questions along with references to detailed descriptions, answers and clarifications. The selection process involved three phases, and they were conducted with the assistance of experts, educators, and SCS students. The following criteria were used to select the questions: (a) questions needed to be evenly distributed across difficulty levels, (b) they should not require complex mathematical calculations, (c) they should not assume prior programming knowledge, and (d) the set of questions should sufficiently represent the five CT dimensions.

A native speaker fluent in both English and Greek reviewed the translation and the questions were piloted. The final selection included 12 questions covering all five CT dimensions across three levels of difficulty, comprising eight multiple-choice and four open-ended questions. Cronbach's alpha at pretest (.844) indicated a reliable test with high internal consistency; however, posttest Cronbach's alpha decreased (.532), potentially due to the participants' newly acquired knowledge, shifts in thinking, or changes in behavior or engagement. Further research could clarify these changes.

Ethical Issues

The research was approved by the Ethics and Research Committee of the Hellenic Open University (51/2024). Furthermore, it received approval for implementation in the Second Chance Schools from the General Secretariat of Vocational Education, Training, and Lifelong Learning, the competent authority of the Greek Ministry of Education. Participants voluntarily participated in the study and provided their consent by signing a Declaration of Consent after being fully informed.

Findings

Analysis of the 12 CT Questions

We aimed to examine the potential relationship between pretest and posttest performance on the responses to the 12 Computational Thinking (CT) test questions (categorized as correct, no answer, or incorrect) (Table 1) by conducting a chi-square (χ^2) test in SPSS. However, the χ^2 test could not be applied because its assumptions were violated. Specifically, the expected frequencies in some cells of the 3x3 contingency table were below the threshold of 5, a condition that undermines the validity of the test. This violation occurs because the χ^2 test relies on sufficient sample sizes in each cell to ensure accurate statistical inference. Application of the test under these conditions could lead to biased or unreliable results. To address this issue and maintain the robustness of the analysis, we combined the adjacent response categories "no answer" and "incorrect". This approach is commonly used to meet statistical assumptions while preserving the interpretability of the data. All variables were recoded (1 = "correct", 0 = "incorrect or no answer") and McNemar' s tests for equality of proportions were performed to assess the independence of two correlated samples. The necessary prerequisites for conducting the McNemar test were met, as it involves a pair of binary variables for ten questions (excluding Q12 and Q8, which require a different approach; further explanations are provided five paragraphs below). The null hypothesis assumes that there is no difference in the proportions of answers between the pretest and posttest. For questions 1, 2, 3, 4, 5, 6, 9, 10 and 11, McNemar' s test ($N = 48$, $p < .05$) led to rejection of the null hypothesis. This suggests a significant difference between pretest and posttest responses. Considering the data in Table 1 as well, it may be inferred that a higher percentage of learners are expected to provide correct answers to these CT questions after the STEM intervention.

Eight of the questions were multiple-choice and not analyzed further. However, questions 6, 7, 8 and 12 were open-ended, warranting additional analysis. This is crucial for examining not only participants who provided fully correct answers but also those whose responses closely approximate the correct solution.

Question 6 is considered correct only if learners determine the correct order of four colors, achieved by blending colors and combining various data. Only one out of 48 learners provided a completely correct answer in the pretest, whereas this ratio improved to one out of three in the posttest (Table 1). All incorrect answers were analyzed since even a minor error during the solution process results in an incorrect color. Notably, two learners in the pretest and four in the posttest managed to correctly identify three out of four colors in the correct order, suggesting that they possess the logical approach required to solve Question 6. Among those who answered incorrectly, 14.6% of participants in the pretest demonstrated reasoning far removed from the problem-solving logic, a percentage that decreased to 8.3% in the posttest.

The percentage of correct answers to Question 7 more than doubled (18.8% in the pretest, 41.7% in the posttest). However, this is the first question where the proportion of incorrect answers increased after the intervention (25%

in the pretest, 47.9% in the posttest), although a larger number of participants provided an answer (no answer: 27% in the pretest, 5% in the posttest). Despite the increase in correct answers, the McNemar Test (N = 48, p = .007, binomial distribution) suggests that we should adopt the null hypothesis, indicating no significant difference in the proportions of answers.

To explain the rise in incorrect answers to Question 7, and considering that the question was open-ended, we analyzed all the incorrect responses. We found that some participants approached the problem with the correct reasoning but narrowly missed the exact solution. Specifically, they arrived at the correct answer for a satisfaction level of 13, whereas the precise answer was for a satisfaction level of 14. As a result, they provided a second-best solution. When we combined those who gave either the best or second-best solution, the proportion increased from 27.1% in the pretest to 70.8% in the posttest. We recorded the data (1="best or 2^{nd} best solution"; 0 = incorrect), and the result of the McNemar test (N = 48, p <.001, binomial distribution) indicates a statistically significant difference between the pretest and posttest. Therefore, it may be concluded that learners were more likely to approach the solution more accurately after the STEM intervention, providing either the best or second-best solution.

Question 8 was also open-ended, requiring participants to identify the least expensive solution in a path graph. No participants provided the correct answer in the pretest, but 43.8% succeeded in finding the exact solution in the posttest. The most cost-effective solution is a path with a length of 41, although many participants identified the second-best path (42). Despite an increase in incorrect answers, more participants were closer to the correct solution. The combined proportion of those providing either the correct or the second-best solution was 12.5% in the pretest, which improved to 72.9% in the posttest.

Similarly, no participant provided the correct answer to the open-ended Question 12 in the pretest. The answer is considered absolutely correct if participants decode the Kix-Code and identify the four letters or numbers in the correct order. Analysis of the incorrect answers reveals that 2% of participants provided the best or second-best solution (three or four correct elements) in the pretest, while this percentage increased to 27% in the posttest.

As mentioned in the first paragraph of this section, the McNemar test should not be applied to Questions 8 and 12 using the codes "correct" and "incorrect or no answer," since no one provided the correct answer in the pretest. A new recoding scheme was applied for more statistically reliable results: "1 = best or 2^{nd} best solution" and "2 = incorrect or no answer." For both questions, the McNemar test results (N = 48, p <.001, binomial distribution) show that the null hypotheses are rejected. Data in Table 1 lead to the conclusion that a higher proportion of participants are expected to provide the best or second-best solutions to Questions 8 and 12 after the STEM intervention.

Table 1. Outcomes of CT Tests, CT dimensions per question and McNemar

		No answer		Correct		Incorrect	
		n	**%**	**n**	**%**	**n**	**%**
Q1: Balls	Pretest	10	20.80%	13	27.10%	25	52.10%
(A) AB, AL, EV	Posttest	4	8.30%	34	70.80%	10	20.80%
	McNemar: p = .004						
Q2: Blossom	Pretest	22	45.80%	10	20.80%	16	33.30%
(B) EV, GE	Posttest	2	4.20%	39	81.30%	7	14.60%
	McNemar: p <.001						
Q3: Five Sticks	Pretest	13	27.10%	25	52.10%	10	20.80%
(B) AB, AL, DE	Posttest	0	0.00%	39	81.30%	9	18.80%
	McNemar: p = .004						
Q4: Superpower Family	Pretest	20	41.70%	15	31.30%	13	27.10%
(A) AB, AL, DE, EV	Posttest	3	6.30%	33	68.80%	12	25.00%
	McNemar: p = .001						
Q5: Beaver Lunch	Pretest	21	43.80%	6	12.50%	21	43.80%
(B) AB, DE, EV, GE	Posttest	0	0.00%	35	72.90%	13	27.10%
	McNemar: p <.001						
Q6: Theater*	Pretest	36	75.00%	1	2.10%	11	22.90%
(B) AL, DE, EV	Posttest	23	47.90%	16	33.30%	9	18.80%
	McNemar: p <.001						
Q7: Soda Shop*	Pretest	27	56.30%	9	18.80%	12	25.00%
(B) AB, AL, EV	Posttest	5	10.40%	20	41.70%	23	47.90%
	McNemar: p = .007 (For the best or 2nd best solution McNemar p <.001)						
Q8: Toll roads*	Pretest	32	66.70%	0	0.00%	16	33.30%
(B) AB, AL, EV, GE	Posttest	5	10.40%	21	43.80%	22	45.80%
	(For best or 2nd best solution McNemar: p <.001)						
Q9: Gifts	Pretest	40	83.30%	1	2.10%	7	14.60%
(C) AL, EV	Posttest	10	20.80%	26	54.20%	12	25.00%
	McNemar: p <.001						
Q10: Bowl Factory	Pretest	37	77.10%	1	2.10%	10	20.80%
(C) AB, AL, DE, EV	Posttest	21	43.80%	7	14.60%	20	41.70%
	McNemar: p = .031						

		No answer		Correct		Incorrect	
		n	%	n	%	n	%
Q11: Icon Image Reduction	Pretest	43	89.60%	3	6.30%	2	4.20%
(C) DE, EV, GE	Posttest	21	43.80%	16	33.30%	11	22.90%
	McNemar: p <.001						
Q12: Kix Code*	Pretest	44	91.70%	0	0.00%	4	8.30%
(C) AL, GE	Posttest	23	47.90%	13	27.10%	12	25.00%
	(For the best or 2nd best solution McNemar: p <.001)						

Mc Nemar N=48 Binominal distribution

n = number of participants who provided each type of response.

(A)= Easy, (B)= Medium, (C)= Difficult

Abstraction (AB), Algorithmic Thinking (AL), Decomposition (DE), Generalization (GE), Evaluation (EV)

*Open-ended questions

For a more detailed description of questions with answers and clarifications see Blokhuis et al. (2015), Blokhuis et al., (2016), Blokhuis et al., (2017), Blokhuis et al., (2018).

Table 1 also provides statistical data on the performance of the 12 CT test questions before and after the intervention. Notably, there is an observed improvement in correct responses across all CT questions following the STEM learning program. For example, only 20.8% of participants answered Q2 correctly in the pretest, compared to 81.3% in the posttest. Figure 1 reveals that a greater number of participants answered correctly across all questions after the STEM intervention, without exception.

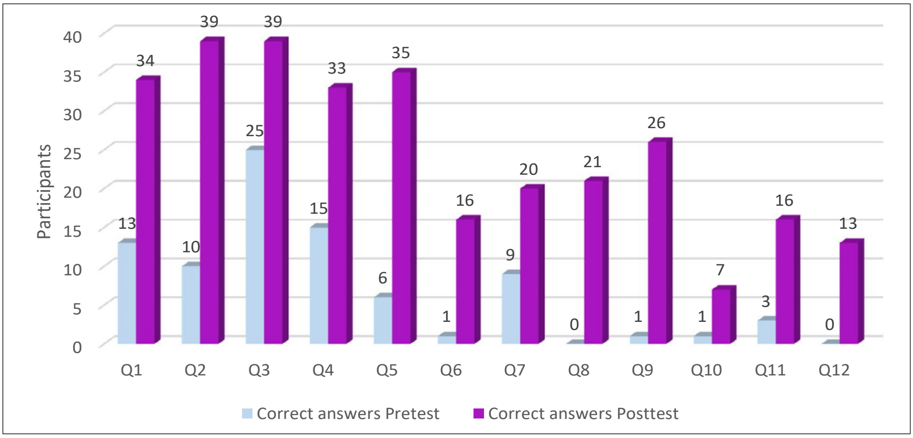

Figure 1. Number of Participants Who Correctly Answered Each Question

Further Analysis of the Number of Correct CT Answers

To conduct a more in-depth analysis, we calculated the number of correct answers provided by each participant and applied both tests. For the variables "number of correct answers in the pretest" and "number of correct answers in the posttest", a paired-sample *t*-test was considered to assess the equality of the two means. While the *t*-test assumptions appeared to be satisfied in terms of quantitative variables, sample derivation from the same population and the assumption of normality based on the Central Limit Theorem, the results of the Shapiro-Wilk test (p <.001 for the pretest, p = .036 for the posttest) indicated that the assumption of normality was violated. Consequently, a Wilcoxon Signed-Rank Test was conducted, revealing a statistically significant difference between the two measurements (z = −5.795, p <.001). Specifically, the median number of correct answers increased from *Mdn* = 1 in the pretest to *Mdn* = 6 in the posttest, with positive ranks (N = 44) demonstrating that the STEM intervention had a substantial positive effect on the participants' ability to provide correct answers.

Additionally, the mean number of correct answers increased from M = 1.75 (pretest) to M = 6.23 (posttest), with a mean improvement of 4.48 correct answers (*SD* = 2.41). This demonstrates that the STEM intervention had a substantial positive effect on the participants' ability to provide correct answers. A line graph of the number of correct answers, with the Standard Error of the Mean (*SEM* = 0.35 for pretest and *SEM* = 0.4 for posttest), is presented in Figure 2.

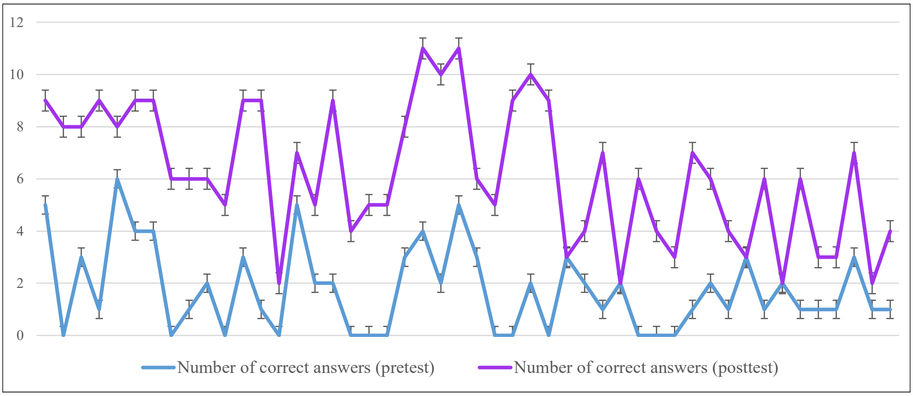

Figure 2. Number of Correct Answers with the Standard Error of the Mean (SEM)

As per the total number of correct answers of the 48 participants, they provided 84 correct answers in the pretest and 299 in the posttest. This represents a 256% improvement over the initial number of correct answers. Alternatively, the posttest performance corresponds to 356% of the pretest performance in terms of group success. It is important to note that these percentages do not account for the improvement achieved by participants who provided second-best solutions after the intervention, wherefore the actual improvement is likely even more substantial.

The Spearman Rank Correlation Coefficient was used to evaluate the relationship between the number of correct answers in CT before and after the intervention, as the variables did not follow a normal distribution. The results indicate a significant positive correlation, $\rho\ (46) = .374$, $p = .009$. This finding suggests that the participants who performed well before the intervention tended to perform well afterwards.

Abstention Rate

This section examines the abstention rate, focusing on the number of questions each participant chose not to answer. Figure 3 provides a visual representation of this data. Before the intervention, only two participants answered all questions, whereas this number increased to ten after the intervention. Moreover, 18 participants left between nine and 12 questions unanswered in the pretest, while in the posttest, no participant left more than eight questions unanswered.

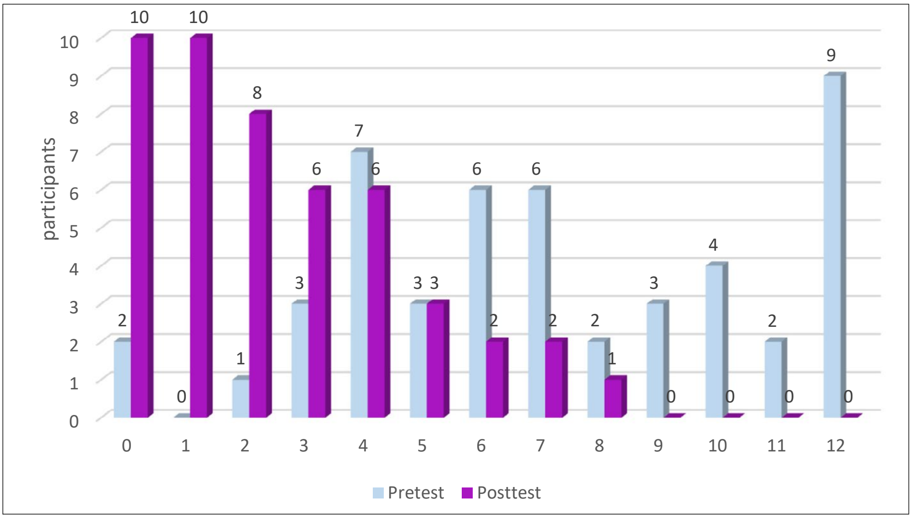

Figure 3. Number of Unanswered Questions

The requirements for a Wilcoxon Signed-Rank Test for the variables "Number of unanswered questions in the pretest" and "Number of unanswered questions in the posttest" were met, as the Shapiro-Wilk test (p <.023 for the pretest, p = .001 for the posttest) revealed non normality. The results of the Wilcoxon Signed-Rank Test (z = −5.66, p <.001) revealed a statistically significant difference between the two variables. Specifically, there was a significant decrease in the number of "unanswered questions" in the posttest compared to the pretest. The negative ranks (N = 42) indicated that most participants had fewer unanswered questions after the STEM intervention. The median number of unanswered questions decreased from Mdn = 7 in the pretest to Mdn = 2 in the posttest. These findings indicate that learners are significantly more likely to attempt answering CT questions after STEM lessons.

Additionally, a t-Test was conducted to confirm the findings, as the t-test is known to be robust to violations of normality in larger samples. The t-test results (t[47] = ...10.40, p <.001) reveal a highly significant difference between the number of unanswered questions in the pretest (M = 7.19, SD = 3.44) and the posttest (M = 2.44, SD = 2.14). These results further support the robustness of the t-test despite the violation of normality and confirm the observed improvement following the STEM intervention.

A Spearman's correlation was conducted in order to explore the correlation between pretest and posttest abstention. The results reveal a significant positive relationship between the total number of unanswered questions in the pretest and the posttest, ρ(46)= .451, p = .001. Participants who had more "no answers" in the pretest tended to have more "no answers" in the posttest as well.

Calculation of Performances According to Bebras

Evaluation of CT performance requires calculation of the score based on the Bebras competition's scoring rules for the 12 questions, given that Bebras questions are used in this study. As shown in Table 2, the three levels of difficulty yield different point values (zero, positive, or negative). Additionally, a bonus starting point must be applied to the scoring system to ensure that negative final scores are impossible.

Table 2. Information for Score Calculation According to the Bebras Competition

	"Incorrect" Points	"No Answer" Points	"Correct" Points	Questions	Start Points	Maximum Points
(A) Easy	0	0	6	1, 2, 3, 4	0	24
(B) Medium	-2	0	9	5, 6, 7, 8	8	36
(C) Difficult	-4	0	12	9, 10, 11, 12	16	48
Total score					24	132

For the variables "Bebras score pretest" and "Bebras score posttest," the Shapiro-Wilk test revealed non-normality for the pretest scores (p = .001), whereas the posttest scores were normally distributed (p = .15). A Wilcoxon Signed-Rank Test was conducted for these variables, with the results (z = ...5.97, p <.001) indicating a significant increase in the participants' scores on the CT test following the intervention (z = ...5.97, p <.001). The median score on the Bebras measuring score was *Mdn* = 30 before the intervention and increased to *Mdn* = 64 after the intervention. The positive rank (N = 44) shows that the participants scored higher in the posttest compared to the pretest and the negative rank (N = 0) shows that none of the participants scored lower after the intervention.

The mean pretest scores (M = 31.71, SD = 9.98) and posttest scores (M = 67.50, SD = 24.97) show that both the mean and median scores more than doubled after the intervention.

Spearman's correlation was conducted to explore the correlation between the pretest and posttest CT scores according to Bebras. The results revealed a significant positive relationship between the pretest and posttest scores, ρ (46) = .492, p <.000. Participants who performed better in the pretest tended to perform better in the posttest as well.

In developing a mathematical model to analyze the correlation between the variables, a simple linear regression was found to provide a clear and interpretable equation, F(1,46) = 15.35, β = 0.500, p = .000. The R^2 was 0.250, indicating that the Bebras score pretest explained approximately 25% of the variance in the Bebras score posttest. The regression equation was:

[Bebras score posttest] = [27.881] + 1.25 [Bebras score pretest]

This implies that the predicted Bebras score on the posttest increases by approximately 1.25 points for each 1-point increase in the Bebras score on the pretest. Confidence intervals indicate that we can be 95% certain that the slope to predict the learners' scores on the Bebras test after the STEM intervention, based on their scores before the intervention, lies between 0.608 and 1.891.

In addition, the possibility of more complex relationships was explored and a third-degree polynomial model explained 32.4% of the variance (R^2 = .324, p <.001). The polynomial model was found to be statistically significant, with F(3, 44) = 7.015, β = 0.211, p <.001. While this model showed an improvement in explaining the variance, it includes more terms and is less straightforward in interpreting the results. Therefore, we present the equation but will not elaborate this model further.

$[\text{Posttest}]=-7.748+1.764\times[\text{Pretest}]-0.0275\times[\text{Pretest}]^2+0.0001\times[\text{Pretest}]^3$

Difficulty Level Scores and Improvement in CT

The function for calculating CT scores at each difficulty level follows the Bebras scoring rules, as outlined in Table 2. Table 3 presents the mean scores for each difficulty level before and after the intervention, along with the corresponding percentage increases in performance. The analysis demonstrates significant improvements across all levels of difficulty: a 230% increase at the easy level, a 265% improvement at the medium level, and a 176% increase at the difficult level. Overall, the participants achieved a 213% improvement in their mean performance across all levels combined.

Table 3. Difficulty Level Scores According to the Bebras Calculation method

N=48	Shapiro-Wilks		Mean score		Percentage Mean Increase
	Pretest	Posttest	Pretest	Posttest	
Easy	$p < .001$	$p <.001$	7.88	18.13	**230.08%**
Medium	$p = .021$	$p =.021$	8.50	22.46	**265.24%**
Difficult	$p <.001$	$p = .030$	15.33	26.92	**175.60%**
Total			31.71	67.5	**212.87%**

A more detailed analysis was conducted for each difficulty level (easy, medium, difficult) to assess the statistical significance of the differences between pretest and posttest scores, as well as the correlations between them. According to the data presented in Table 3, the results of the Shapiro-Wilk test indicate that the distribution is not normal at any level of difficulty, both in the pretest and the posttest. Therefore, the most appropriate test for analysis is the Wilcoxon Signed-Rank Test and Spearman's correlation.

Easy Level: A Wilcoxon Signed-Rank Test ($z = -5.36$, $p <.001$) indicates that scores on easy CT questions were significantly higher in the posttest than in the pretest. The median score on the "easy" questions posttest was *Mdn* = 95, compared with *Mdn* = 84 for the pretest. These results suggest that the participants showed significant improvement in answering the easy CT questions after the STEM intervention. However, the Spearman's correlation rank test shows that there is no significant relationship between these variables, $r_s(46) = .149$, $p = .313$.

Medium Level: Similarly, a Wilcoxon Signed-Rank Test ($z = -5.73$, $p <.001$) reveals that scores on medium level questions were significantly higher in the posttest than in the pretest. The median score on the medium level questions after the STEM intervention was *Mdn* = 75, compared to *Mdn* = 65 before the intervention. This indicates that the intervention had a positive effect on the participants' CT performance on medium level questions. A Spearman's correlation was conducted to evaluate the relationship between the medium level question scores in the pretest and the posttest, and there was a significant positive relationship between the two variables, $\rho(46) = .413$, $p = .004$. This suggests that, similar to the easy level, we can expect better scores after the STEM lessons at the medium level, with performance being correlated to the pretest score.

Difficult Level: Scores on difficult questions were significantly higher in the posttest than in the pretest, as the results of a Wilcoxon Signed-Rank Test ($z = -5.73$, $p <.001$) reveal. The median score on the "difficult" questions posttest was *Mdn* = 60, compared with *Mdn* = 44 for the pretest. These findings suggest that participants were able to significantly improve their CT performance on the difficult questions following the STEM intervention. A Spearman's correlation revealed no significant relationship between the difficult question scores in the pretest and the posttest, $\rho(46) = .203$, $p = .167$. Thus, while improved scores are expected on the difficult CT questions after the intervention, there is no significant correlation between pretest and posttest performance, similar to the easy level.

Regarding Each Computational Thinking Dimension

One of the main reasons for choosing Bebras tasks is that they provide clarity regarding the specific dimensions of CT required to solve each question, as outlined in Table 1. This allows us to apply scoring functions for each CT dimension using functions based on the Bebras rules, as shown in Table 2. For example, the dimension of abstraction (AB) is necessary for solving questions 1, 3, 4, 5, 7, 8 and 10. These questions vary in difficulty and different points are awarded for correct or incorrect answers. Table 4 presents the key statistical data for the CT dimension scores.

Table 4. Statistical Analysis of CT Dimensions

	Shapiro – Wilks		*Wilcoxon Signed Ranks Test*	Descriptives		Spearman
	Pretest	Posttest		Pretest	Posttest	
Abstraction Start point = 10 Max = 67	$p = .021$	$p = .119$	$z = ...5.36$ $p <.001$	$M = 16.81$ $SD = 7.96$ $Mdn = 16$	$M = 35.17$ $SD = 14.34$ $Mdn = 31$	$r_s(46) = .381$ $p = .008$
Algorithm Start point = 18 Max = 99	$p = .006$	$p = .281$	$z = ...5.94$ $p <.001$	$M = 35.96$ $SD = 8.89$ $Mdn = 36$	$M = 47.52$ $SD = 21.36$ $Mdn = 45$	$r_s(46) = .461$ $p = .001.$

	Shapiro – Wilks		*Wilcoxon Signed Ranks Test*	Descriptives		Spearman
	Pretest	Posttest		Pretest	Posttest	
Decomposition Start point = 12 Max = 57	p = .001	p = .067	z = ...5.30 p <.001	M = 17.25 SD = 6.44 Mdn = 16	M = 30.19 SD = 10.71 Mdn = 27	$r_s(46)$ = ...009 p = .954.
Evaluation Start point = 20 Max = 110	p = .005	p = .071	z = ...5.94 p <.001	M = 24.92 SD = 8.88 Mdn = 21	M = 56.38 SD = 22.16 Mdn = 56	$r_s(46)$ = .485 p = .000
Generalization Start point = 12 Max = 60	p = .001	p = .690	z = ...5.66 p <.001	M = 13.08 SD = 5.58 Mdn = 12	M = 31.25 SD =13.43 Mdn = 32	$r_s(46)$ = .226 p = .122

As shown in Table 4, Wilcoxon Signed-Rank Tests conducted between pretest and posttest scores across all five dimensions reveal statistically significant differences (p <0.001). These findings indicate substantial improvements in performance following the STEM intervention across all CT dimensions. To provide further insight into the magnitude of these improvements, we calculated the percentage mean increase. The largest percentage increases, with more than double the improvement, were observed in the dimensions of generalization (239%), evaluation (226%) and abstraction (209%). Decomposition (175%) and algorithmic thinking (132%) also showed significant improvements, albeit with comparatively smaller percentage gains.

Additionally, we examined the statistical correlation between the pretest and posttest for each dimension separately. As presented in Table 4, a statistically significant medium correlation is observed for the dimensions of abstraction, algorithmic thinking and evaluation, according to Spearman's method (.38 <ρ<.49, p <.05). In contrast, for decomposition and generalization, the null hypothesis cannot be rejected using Spearman's test (p> .05), indicating no statistically significant correlation between the pretest and posttest for these dimensions.

Observation Data – Interpretation

Observation data during the STEM intervention and CT test-taking process were recorded in a research diary. In this paper, we focus on a few key issues related to CT. During the CT pretest, many participants reported that the questions were extremely complicated and difficult and that they were unsure how to approach them. They appeared discouraged, with a tendency to abandon the CT questions. However, the learners gradually became more engaged throughout the STEM lessons, approaching progressively complex activities requiring computational thinking.

For example, discussions and reflections were encouraged during STEM tasks with questions such as: "How can we break the problem and solution into smaller parts?", "What do we know about this?", "Which information is useful?", "What steps must we follow to solve this, and in what order?", "Are our solutions adequate, or can we improve them?", and "Where else can we apply our solutions?". These questions helped foster a deeper understanding of computational thinking and problem-solving strategies.

Overall, the learners seemed to enjoy their progress and appeared more confident in problem-solving. It is important to note that the participants were not taught any of the Bebras CT questions during the intervention. Nevertheless, after the intervention, when the participants took the same CT questionnaire in the posttest, it was evident that they had adopted a computational thinking approach. They viewed the questions as achievable challenges, believed that they could solve most problems and demonstrated increased persistence, ultimately solving more problems than before.

Discussion and Conclusions

Computational thinking is essential for adults but identifying an effective educational program to help them develop CT remains a challenge (Zapata-Rivera et al., 2019). Furthermore, experts emphasize the need for research of adult learners' CT, as studies focusing on older populations are limited (Cutumisu et al., 2019; Kotzampasaki & Koulaouzides, 2024; Ortiz et al., 2023; Poulakis & Politis, 2021; X. Tang et al., 2020). This study aims to address this research gap by investigating the impact of a carefully designed and implemented STEM intervention on the CT skills of adult learners.

Our action research is aligned with experts' recommendations for CT studies (Poulakis & Politis, 2021; X. Tang et al., 2020). Specifically: a) Bebras items were used as they do not require any prior programming knowledge; programming questions were intentionally excluded to consciously distinguish computer science knowledge from the unique computational way of thinking about problems and their solutions, whether specific or broad; b) Computer-based tools were not used for administering the CT test; instead, we opted for printed, colored questionnaires; c) A mixed-methods evaluation was employed, primarily quantitative, but also incorporating qualitative data for a deeper understanding; d) Our sample focused on Second Chance Schools learners, as they represent adult learners with distinct characteristics and are highly relevant to adult education. Furthermore, we were able to study an age range of 26...75 years, all without prior programming or computer science knowledge; e) Content validity and internal consistency were assessed for reliability; and f) The research focused on defining CT and its dimensions. Specifically, we successfully "measured" the five dimensions both before and after the intervention, across three levels of difficulty.

Overall, we observed an improvement in the adult learners' computational thinking because of the STEM intervention. Our results indicate that a greater number of learners answered all the questions correctly after the STEM lessons. Additionally, McNemar tests for equality of proportions, conducted on nine closed-ended questions [1, 2, 3, 4, 5, 9, 10, 11] and open-ended question 6, reveal a statistically significant difference, suggesting that a higher percentage of participants answered these questions correctly after the STEM intervention. Three open-ended questions [7, 8, 12] were recoded and McNemar tests between the pretest and posttest show a statistically significant improvement in solving these questions, either by providing the best or second-best solution. Thus, for all questions, a greater number of learners are expected to approach CT problems more accurately after the STEM intervention, offering either the correct answer or at least the second-best solution. Moreover, the significant correlation between the number of correct answers in CT before and after the intervention potentially reflects the effectiveness of the intervention in maintaining or improving the participants' performance.

The STEM lessons have also had a significant positive impact on the number of questions participants answered correctly (pretest Mdn = 1, posttest Mdn = 6). Additionally, there was a decrease in the abstention from answering questions, with learners demonstrating an increased tendency to attempt to reply to more CT questions after the STEM intervention. This improvement can be further explained by our observational data, which reflects an increase in learner confidence and engagement. In addition, despite the significant decrease in the number of unanswered questions, the Spearman correlation results show that the participants who had more "no answers" in the pretest tended to continue facing difficulties in the posttest. This may suggest that these participants require more targeted and personalized interventions to improve further their performance.

According to the Bebras calculation method, the overall CT score doubled (213%), while the paired sample t-tests reveal a statistically significant difference. Furthermore, Pearson's test shows a medium correlation between the pretest and posttest CT scores. We also identified two mathematical models that can estimate the improvement in 25% of the cases (linear regression) or 32.4% (third-degree polynomial), which is noteworthy, considering that it applies to a human science mathematical formula where thinking patterns and behaviors are inherently complex. Along with the other evidence, the highly statistically significant difference observed between pretest and posttest scores reinforces the conclusion that the STEM intervention had a positive impact on the learners' CT.

Regarding difficulty levels, paired sample t-tests indicate that we expect greater performance after the STEM intervention at all CT levels. The percentage improvements for each difficulty level are as follows: easy (230%), medium (265%), and difficult (176%). However, Pearson's method reveals no correlation between pretest and posttest scores for easy and difficult questions. To clarify, although there is a statistically significant difference in mean scores, indicating improved performance after the intervention, there is no evidence suggesting that the participants' performance on the easy pretest questions predicts their performance on the easy posttest. In fact, we observed learners who scored extremely low on the pretest achieving excellent scores on the easy questions after the STEM lessons. Similarly, improved scores were expected on the difficult CT questions after the intervention, but no significant correlation was found with the pretest scores. On the one hand, this suggests that anyone could potentially solve difficult questions, regardless of their previous performance; on the other hand, we must acknowledge that only a few participants attempted the difficult questions on the pretest. For the medium difficulty level, performances on the pretest and posttest are related, as indicated by Pearson's correlation test. To illustrate, participants with higher CT scores on the medium-level questions in the pretest were expected to achieve better outcomes on the same questions in the posttest.

In addition, a statistically significant improvement is observed across all CT dimensions following the implementation of the STEM education program. The percentages of improvement are as follows: generalization (239%), evaluation (226%), abstraction (209%), decomposition (175%) and algorithmic thinking (132.24%). Pearson's method indicates a medium correlation between the pretest and posttest for algorithmic thinking, evaluation and decomposition. This suggests that improvements in these CT dimensions were moderately linked to the participants' initial performance.

We sought to compare our qualitative findings with other studies involving adults, but we were unable to find previous research that aligns with our sample, method and analysis. Although wary of comparing different age groups, we will refer to the study by Psycharis and Kotzampasaki (2019), which involved 115 Greek 5th-6th grade students, as it follows a similar research method and analysis. In that study, the students showed a statistically significant increase in their CT performance on the Bebras Test after the implementation of a STEM inquiry-based game. Specifically, their mean percentages of improvement were as follows: generalization (624%), algorithmic thinking (622%), abstraction (578%), decomposition (265%) and evaluation (221%). We observe that, except for evaluation, where improvement is similar to our findings, the students demonstrated much greater percentages of improvement in the other dimensions. Interestingly, the students showed statistically significant improvement only at the easy and medium difficulty levels, whereas adult learners, with smaller mean improvements, achieved statistically significant improvement at all difficulty levels.

In conclusion, we can affirm that the STEM intervention designed and implemented for this action research has had a positive impact on the adult learners' computational thinking (CT), encompassing all CT dimensions and levels of difficulty. We argue that the STEM intervention helps learners gradually internalize the thought patterns of CT dimensions, enabling them to approach problems with a computational way of thinking. We believe that the positive correlation between CT, general skills and thinking styles (Durak & Saritepeci, 2018) is crucial for understanding the world, as Denning and Tedre (2019) highlight, and furthermore, for gaining insight into how people think. It is important to recognize that thinking is a multifaceted process and that the changes we have observed in attitude and behavior should not be underestimated. The interpretation of CT results is closely tied to other research questions we aim to address, which will be discussed in the following chapter on further research.

Further Research

This article presents the results of our intervention on adult learners' (CT) as part of a broader mixed-methods action research project. Interviews and additional questionnaires were used to explore further research questions related to critical reflection, the potential for transformative learning, and the correlation with CT (Kotzampasaki & Koulaouzides, 2024). By addressing these supplementary questions, we aim to offer a deeper and more comprehensive understanding of how adults apply computational thinking. We are convinced that further research on CT-STEM in adult education is essential, particularly for learners in Second Chance Schools, as the outcomes there carry significant implications.

References

Araujo, A. L. S. O., Andrade, W. L., Guerrero, D. D. S., & Melo, M. R. A. (2019). How many abilities can we measure in computational thinking? A study on Bebras challenge. *Paper presented at SIGCSE '19: The 50th ACM Technical Symposium on Computer Science Education*. https://doi.org/10.1145/3287324.3287405

Blokhuis, D., Csizmadia, A., Millican, P., Roffey, C., Schrijvers, E., & Sentence, S. (2017). *UK Bebras Computational Thinking Challenge 2017 Answers.* University of Oxford.

BLOKHUIS, D., CSIZMADIA, A., MILLICAN, P., ROFFEY, C., SCHRIJVERS, E., & SENTENCE, S. (2018). *UK Bebras Computational Thinking Challenge 2018 Answers.* University of Oxford.

BLOKHUIS, D., MILLICAN, P., ROFFEY, C., SCHRIJVERS, E., & SENTENCE, S. (2015). *UK Bebras Computational Thinking Challenge 2015.* University of Oxford.

BLOKHUIS, D., MILLICAN, P., ROFFEY, C., SCHRIJVERS, E., & SENTENCE, S. (2016). *UK Bebras Computational Thinking Challenge 2016 Answers.* University of Oxford.

BRENNAN, K., & RESNICK, M. (2012, April). New Frameworks for Studying and Assessing the Development of Computational Thinking. *Proceedings of the 2012 Annual Meeting of the American Educational Research Association, Canada*, *1*, 1–25.

COOK, C. T., DRACHOVA, S., HALLSTROM, J. O., HOLLINGSWORTH, J. E., JACOBS, D. P., KRONE, J., & SITARAMAN, M. (2012). A systematic approach to teaching abstraction and mathematical modeling. *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education – ITiCSE '12.* https://doi.org/10.1145/2325296.2325378

CSIZMADIA, A., STANDL, B., & WAITE, J. (2019). Integrating the Constructionist Learning Theory with Computational Thinking Classroom Activities. *Informatics in Education*, *18*(1), 41−67. https://doi.org/10.15388/infedu.2019.03

CSTA, & ISTE. (2011). *Operational Definition of Computational Thinking for K–12 Education*. http://csta.acm.org/Curriculum/sub/CurrFiles/ CompThinkingFlyer.pdf

CUTUMISU, M., ADAMS, C., & LU, C. (2019). A Scoping Review of Empirical Research on Recent Computational Thinking Assessments. *Journal of Science Education and Technology*, *28*(6), 651...676. https://doi.org/10.1007/s10956...019...09799...3

DAGIENĖ, V., & STUPURIENE, G. (2016). Bebras-A Sustainable Community Building Model for the Concept Based Learning of Informatics and Computational Thinking. *Informatics in education*, *15*(1), 25–44. https://doi.org/10.15388/infedu.2016.02

DENNER, J., WERNER, L., & ORTIZ, E. (2012). Computer games created by middle school girls: Can they be used to measure understanding of computer science concepts? *Computers & Education*, *58*(1), 240–249.

DENNING, P. J., & TEDRE, M. (2019). *Computational Thinking*. MIT Press.

DJAMBONG, T., & FREIMAN, V. (2016). Task-Based Assessment of Students' Computational Thinking Skills Developed through Visual Programming or Tangible Coding Environments. *International Association for Development of the Information Society*, 41−51.

DURAK, H. Y., & SARITEPECI, M. (2018). Analysis of the relation between computational thinking skills and various variables with the structural equation model. *Computers & Education*, *116*, 191–202. https://doi.org/10.1016/j.compedu.2017.09.004

EL MAWAS, N., HOOSHYAR, D., & YANG, Y. (2021). Effects of Adaptive Educational Games on Adults' Computational Thinking. In *Computer Supported Education: 12th International Conference, CSEDU 2020, Virtual Event, May 2–4, 2020, Revised Selected Papers 12* (pp. 162−176). Springer International Publishing. https:doi.org/10.1007/978...3...030...86439...2_9

Fidai, A., Capraro, M. M., & Capraro, R. M. (2020). "Scratch"-ing computational thinking with Arduino: A meta-analysis. *Thinking Skills and Creativity*, *38*, 1–14. https://doi.org/10.1016/j.tsc.2020.100726

Fogli, D., Danesi, F., Malizia, A., Turchi, T., & Bell, D. (2017). Sustaining Cultures of Participation by Fostering Computational Thinking Skills through Game-Play. *CEUR in Proceedings of GHItaly*, 1–6.

Gao, W. (2020). *Research on cultivation of computational thinking based on game-based teaching: Taking the basic course of computer application in higher vocational education as an example.* Master's Thesis, Hunan Normal University. CNKI. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202101&filename=1020320653.nh

Hill, L. H., Rogers-Shaw, C. A., & Carr-Chellman, D. J. (2023). But, *is* it Adult Education? Disciplinary Boundaries of Adult Education and Higher Education. *Adult Learning*, *34*(2), 59–67. https://doi.org/10.1177/10451595231159430

Hu, C. (2011, June). Computational thinking: what it might mean and what we might do about it. In *Proceedings of the 16th annual joint conference on Innovation and technology in computer science education* (pp. 223–227). ACM.

Hubwieser, P., & Mühling, A. (2015, April). Investigating the Psychometric Structure of Bebras Contest: Towards Measuring Computational Thinking Skills. In *2015 International Conference on Learning and Teaching in Computing and Engineering* (pp. 62–69). IEEE.

Kazimoglu, C. (2020). Enhancing Confidence in Using Computational Thinking Skills via Playing a Serious Game: A Case Study to Increase Motivation in Learning Computer Programming, IEEE Access 2020, 1831–1851.

Kotzampasaki, E., & Koulaouzides, A. (2024). Computational thinking and critical reflection through STEM education at second chance school. In G. Koulaouzides & D. Soeiro (Eds), *Reimagining Transformative and Emancipatory Adult Education for a World to Come*. Proceedings of the 1st conference of the ESREA Transformative and Emancipatory Adult Education Network (pp. 183–190). Fylatos Publishing.

Koulaouzides, G. (2019). *A Dialogue for Adult Education* [In Greek]. Metaichmio Publications.

Koutouzis, M., Papadopoulou, A., & Chalkiadaki, A. (2023). Lifelong Learning and Adult Education. *Reimagining Education for the Second Quarter of the 21st Century and Beyond: Dilemmas, Challenges, Advancements and Innovations*, *8*, 214–225.

Lafuente Martínez, M., Lévêque, O., Benítez, I., Hardebolle, C., & Zufferey, J. D. (2022). Assessing Computational Thinking: Development and Validation of the Algorithmic Thinking Test for Adults. *Journal of Educational Computing Research*, *60*(6), 1436–1463. https://doi.org/10.1177/07356331211057819

Ling, U. L., Saibin, T. C., Naharu, N., Labadin, J., & Aziz, N. A. (2018). An Evaluation Tool to Measure Computational Thinking Skills: Pilot Investigation. *National Academy of Managerial Staff of Culture and Arts Herald*, *1*, 606–614.

Lockwood, J., & Mooney, A. (2018). Developing A Computational Thinking Test Using Bebras Problems. In A. Piotrkowicz, R. Dent-Spargo, S. Dennerlein, I. Koren, P. Antoniou, P. Bailey, T. Treasure-Jones, I. Fronza, C. Pahl (Eds.), Joint Proceedings of the CC-TEL 2018 and TACKLE 2018 Workshops, co-located with *13th European Conference on Technology Enhanced Learning* (EC-TEL 2018). https://ceur-ws.org

López, A. R., & García-Peñalvo, F. J. (2016, November). Relationship of knowledge to learn in programming methodology and evaluation of computational thinking. In *TEEM '16: Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 73–77). https://doi.org/10.1145/3012430.3012499

Lu, Z., Chiu, M. M., Cui, Y., Mao, W., & Lei, H. (2023). Effects of Game-Based Learning on Students' Computational Thinking: A Meta-Analysis. *Journal of Educational Computing Research*, *61*(1), 235–256. https://doi.org/10.1177/07356331221100740

Ma, J., Zhang, Y., Zhu, Z., Zhao, S., & Wang, Q. (2023). Game-Based Learning for Students' Computational Thinking: A Meta-Analysis. *Journal of Educational Computing Research*, *61*(7), 1430–1463. https://doi.org/10.1177/07356331231178948

Man, K., Goździk, A., & Korda, M. (2016). *Edu-Arctic ... Innovative educational program attracting young people to natural sciences and polar research.* [Report on desk research #710240], European Union's Horizon 2020 research. https://edu-arctic.eu/images/project_reports/EDU-ARCTIC_D3.1_v7_19...07...2016_KM.pdf

National Research Council (2011). *Report of a Workshop of Pedagogical Aspects of Computational Thinking.* https://www.nap.edu/catalog.php?record_id=13170

Ortiz, J. S., Moreira, C., Menezes, K., Ferrari, B., Silva Junior, D., & Pereira, R. (2023). Computational Thinking and Mental Models: Promoting Digital Culture in the Youth and Adult Education. *Interacting with Computers*, *35*(2), 91–104. https://doi.org/10.1093/iwc/iwac028

Pliasa, S., & Fachantidis, N. (2021). Mobile Technologies Serious Games for the Development of Social Skills in Children with Autism Spectrum Disorders, in Enhanced with Socially Assistive Robots Interventions. In M. E. Auer, T. Tsiatsos (Eds.), *Internet of Things, Infrastructures and Mobile Applications. IMCL 2019. Advances in Intelligent Systems and Computing*, *1192*, 618–628. Springer. https://doi.org/10.1007/978...3...030...49932...7_58

Poulakis, E., & Politis, P. (2021). Computational Thinking Assessment: Literature Review. In T. Tsiatsos, S. Demetriadis, A. Mikropoulos, V. Dagdilelis (Eds.), *Research on E-Learning and ICT in Education* (pp. 111–128). Springer. https://doi.org/10.1007/978...3...030...64363...8_7

Psycharis, S., & Kotzampasaki, E. (2019). The Impact of a STEM Inquiry Game Learning Scenario on Computational Thinking and Computer Self-confidence. *Eurasia Journal of Mathematics, Science and Technology Education*, *15*(4), 1–18. https://doi.org/10.29333/ejmste/103071

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining Assessment Tools for a Comprehensive Evaluation of Computational Thinking Interventions. In S-C. Kong & H. Abelson (Eds.), *Computational Thinking Education* (pp. 79–98). Springer.

Selby, C., & Woollard, J. (2014). Refining an Understanding of Computational Thinking. *Author's Original*, 1–23.

Swanson, H., Anton, G., Bain, C., Horn, M., & Wilensky, U. (2019). Introducing and Assessing Computational Thinking in the Secondary Science Classroom. In S-C. Kong & H. Abelson (Eds.), *Computational Thinking Education* (pp. 99–117). Springer.

Tang, K. Y., Chou, T. L., & Tsai, C. C. (2020). A Content Analysis of Computational Thinking Research: An International Publication Trends and Research Typology. *The Asia-Pacific Education Researcher*, *29*(1), 9–19.

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, 1–64. https://doi.org/10.1016/j.compedu.2019.103798

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*, 33–35.

Wing, J. (2011). Research notebook: Computational Thinking – What and Why? *the link*, *6*, 20–23. https://people.cs.vt.edu/~kafura/CS6604/Papers/CT-What-And-Why.pdf

Wing, J. M. (2017). Computational thinking's influence on research and education for all. *Italian Journal of Educational Technology*, *25*(2), 7–14. https://doi.org/10.17471/2499...4324/922

Yadav, A., Good, J., Voogt, J., & Fisser, P. (2017). Computational Thinking as an Emerging Competence Domain. In *Competence-based Vocational and Professional Education Bridging the Worlds of Work and Education*, 1051–1067. http://doi.org/10.1007/978...3...319...41713...4_49

Yadav, A., Mayfield, C., Zhou, N., Hambrusch, S., & Korb, J. T. (2014). Computational Thinking in Elementary and Secondary Teacher Education. *ACM Transactions on Computing Education*, *14*(1), 1–16. https://doi.org/10.1145/2576872

Zapata-Rivera, D., Forsyth, C., & Molloy, H. (2019). Supporting Computational Thinking Skills for Adults. In *SLLL@ AIED* (pp. 28–31).