education sciences

MDPI

Article

Understanding Developmental Trajectories of Computational Thinking Concepts in Primary School: An Empirical Study of Sequences, Loops, and Conditionals

Ioannis Vourletsis

Pedagogical Department of Primary Education, University of Thessaly, 38221 Volos, Greece; vourlets@uth.gr

Abstract

Computational thinking (CT) is increasingly recognized as a foundational skill in primary education, yet its developmental progression in the early school years remains underexplored. This study examined CT as a competence comprising three core concepts—sequences, loops, and conditionals—through a cognitive developmental lens. A total of 517 students in Grades 1 to 3 in Greece were assessed using the Greek adaptation of the Beginners Computational Thinking Test (BCTt), a validated tool for young learners. To examine performance trends, conceptual interrelations, and learner profiles, we employed repeated-measures ANOVAs, correlation analysis, and cluster analysis. The results showed that students performed highest in sequences, followed by loops and conditionals, with statistically significant differences across concepts. This pattern was also reflected in the cluster analysis, which identified three distinct student profiles differing in both overall performance and conceptual emphasis. Overall, the findings underscore the progressive nature of CT development and highlight the need for instruction aligned with students' cognitive readiness and conceptual growth.

Keywords: computational thinking (CT); CT skill development; developmental differences; primary education; Beginners Computational Thinking Test (BCTt)

1. Introduction

Over the past two decades, computational thinking (CT) has emerged as a central concept in educational research and curriculum development. Far from being limited to discipline-specific technical skills, CT constitutes a cross-disciplinary cognitive ability that supports logical reasoning, problem-solving, and conceptual understanding across diverse learning domains (Grover & Pea, 2018). This perspective is reflected in Wing's (2011) influential definition of CT as "the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent," positioning CT as both a cognitive and practical capacity essential for 21st-century learning.

The increasing attention to CT has prompted growing academic discussion regarding its conceptual foundations, particularly in relation to its conceptualization as a cognitive construct (Kafai et al., 2020), emphasizing the underlying mental processes involved in algorithmic reasoning and problem-solving. This view aligns with developmental theories, particularly Piaget's (1964) model of cognitive development, which frames learning as a progression of cognitive capacities. From this perspective, CT is understood not as a single, uniform skill but as a multifaceted system composed of interrelated concepts—such as

Check for updates

Academic Editor: Diego Vergara

Received: 11 March 2026
Revised: 6 April 2026
Accepted: 8 April 2026
Published: 10 April 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

sequences, loops, and conditionals—that develop alongside broader aspects of children's cognitive growth.

Despite the growing scholarly attention to CT, there remains a limited understanding of how it develops during the early years of formal education. Previous reviews highlight the diversity of instructional approaches and emphasize the need to align content with children's cognitive capacities at different developmental stages (Hsu et al., 2018). More recent research reinforces this perspective, underscoring the importance of early CT experiences and the need for age-appropriate pedagogical strategies (Yang et al., 2024; Román-González & Pérez-González, 2024).

In line with this cognitive conceptualization, CT is increasingly framed as a cross-disciplinary competence that supports structured problem-solving across domains (Grover & Pea, 2018). In the present study, CT is operationalized through programming-related constructs, focusing specifically on sequences, loops, and conditionals. This approach is consistent with the cognitive orientation described by Kafai et al. (2020) and emphasizes students' ability to recognize and apply fundamental computational structures rather than engage with programming syntax.

In light of this conceptualization, the present study examines how key CT concepts—sequences, loops, and conditionals—emerge and evolve during the early years of formal education. Focusing on students in Grades 1 to 3, the study employs the Greek adaptation of the validated Beginners Computational Thinking Test (BCTt) to trace conceptual progression across early primary schooling. By adopting a developmental lens, the study aims to provide empirically grounded insights into age-related patterns in CT acquisition and inform the design of pedagogical strategies aligned with young learners' cognitive profiles.

2. Theoretical Framework and Related Work

2.1. Operationalizing Computational Thinking Through Sequences, Loops, and Conditionals

Although CT has gained considerable attention in recent years, several of its core ideas can be traced back to earlier educational models. Polya's (1945) problem-solving model, for example, highlighted the role of structured processes and step-by-step thinking—features that later became integral to CT. Similarly, Papert (1980), drawing on constructionist principles, proposed programming as a means for developing thinking skills, highlighting both its cognitive and social dimensions.

The renewed focus on CT following Wing's (2006) influential article spurred efforts to define and operationalize the construct. As Tang et al. (2020) note, two principal orientations have since emerged: one rooted in computer science, emphasizing algorithmic logic, data structures, and programming constructs (Weintrop et al., 2016), and another adopting a broader cognitive framing, treating CT as a general problem-solving disposition involving abstraction, decomposition, pattern recognition, and evaluation (CSTA & ISTE, 2011; Selby & Woollard, 2013). This latter perspective has been particularly influential in general education, where CT is increasingly approached as a cross-curricular, domain-general way of thinking. Expanding on this view, Annamalai et al. (2022) identified five core CT elements—abstraction, decomposition, debugging and evaluation, algorithmic reasoning, and generalization—highlighting their role in flexible and effective problem solving. Importantly, these thought processes are not confined to digital environments; as Hazzan et al. (2020) point out, CT can be fostered through unplugged, non-digital learning experiences as well.

Building on this broader conceptual landscape, one of the most influential educational frameworks is that of Brennan and Resnick (2012), which conceptualizes CT in terms of three interconnected dimensions: computational concepts, computational practices, and computational perspectives. Within this model, sequences, loops, and conditionals are

defined as core elements of the computational concepts dimension. Sequences involve organizing a series of instructions in the correct order to complete a task. Loops represent repeated execution of commands, enabling iteration. Conditionals support decision-making based on whether specific criteria are met.

The central role of sequences, loops, and conditionals in CT education is corroborated by Zhang and Nouri (2019) in their systematic review of 55 empirical studies in K-9 settings. Their analysis revealed that these three concepts are the most frequently identified in research and constitute core structural components of programming languages. Furthermore, the researchers highlight differences in conceptual complexity, noting that sequences are generally more accessible to learners, whereas loops and conditionals introduce additional cognitive demands.

Building on this foundation, Luo et al. (2022) conceptualize sequence as the "springboard" for learning additional CT concepts, emphasizing its function as a scaffold for the development of iteration and conditional logic. Their work emphasizes the foundational role of sequence as a basis for understanding more complex computational structures. As the authors note, "instruction provides an onramp for the students to learn other CT concepts by using the sequence concept as a 'springboard'" (Luo et al., 2022, p. 21).

Empirical research highlights the centrality of these three concepts in early CT education. A systematic review of 42 studies in early childhood CT education further confirms this pattern (Zeng et al., 2023), as sequences were addressed in 73.8% of the studies, loops in 42.9%, and conditionals in 23.8% of them. These concepts also form the exclusive focus of the BCTt—a validated, age-appropriate, multiple-choice assessment designed specifically for young learners. The BCTt evaluates understanding of sequences, loops (simple and nested), and conditionals (if-then, if-then-else, while), without requiring prior coding experience, thereby allowing for the measurement of CT independently of programming skills (Zapata-Cáceres et al., 2020, 2021).

While sequences, loops, and conditionals originate from computer science and procedural programming, they have been widely adopted in educational contexts as foundational structures supporting structured thinking and problem-solving. The decision to focus on these three concepts, rather than the broader set proposed by Brennan and Resnick (2012), was based on their developmental appropriateness for early primary education and their prominence in empirical research involving young learners. In addition, sequences, loops, and conditionals constitute core structural elements of algorithmic thinking and can be assessed in a relatively discrete and observable manner, making them particularly suitable for empirical investigation in early educational contexts. These concepts represent developmentally accessible entry points into CT and provide a suitable basis for examining conceptual progression during the early school years.

Consistent with this conceptual and methodological framework, the present study employed the Greek adaptation of the BCTt (Vourletsis & Politis, 2025), which has demonstrated strong psychometric properties, including high content validity, internal consistency, and test–retest reliability. Drawing on this tool, the study examined how young learners in early primary school conceptualize and differentiate between sequences, loops, and conditionals, aiming to shed light on the developmental trajectory of these foundational CT concepts.

2.2. Cognitive and Developmental Foundations of Computational Thinking

The framing of CT as a cognitively grounded skill is deeply informed by foundational developmental theories, particularly Piaget's (1964) theory of cognitive development. According to this framework, children progress from sensorimotor and preoperational reasoning to concrete operational thought (ages 7–11), and eventually to formal operational

reasoning, characterized by abstract and hypothetical thinking. Given this developmental progression, children's ability to engage with CT concepts is likely to vary according to their level of cognitive maturity. Sequences, being more concrete, may be more accessible to younger learners, while loops and conditionals could place greater demands on capacities that typically mature later, including logical reasoning and cognitive flexibility. More specifically, sequence understanding primarily involves organizing actions in a correct temporal order, whereas loops require maintaining and mentally repeating structured instructions, placing additional demands on working memory. Conditionals, in turn, involve evaluating alternative outcomes based on logical relations (e.g., if–then structures), requiring more advanced reasoning and control processes. While Piaget's framework provides a useful foundation for understanding cognitive development, it is used in the present study as a general interpretive framework rather than a strict developmental model.

This developmental perspective is supported by an expanding body of empirical evidence linking CT performance to core cognitive abilities. For example, Román-González et al. (2017) found strong correlations between CT performance and core cognitive domains such as fluid reasoning, spatial ability, and general problem-solving, suggesting that CT skills are not isolated but are closely related to broader cognitive abilities. Similarly, Gerosa et al. (2021), in a study involving kindergarten children, demonstrated that both sequencing ability and symbolic number comparison significantly predicted CT scores, even after accounting for general intelligence. These findings imply that early numeracy and symbolic reasoning may serve as foundational scaffolds for the emergence of computational thinking during early childhood.

More specifically, at the level of individual CT concepts, Jiang and Wong (2021) emphasized that understanding conditionals and logical operators relies heavily on logical reasoning—a cognitive capacity that develops progressively throughout primary education. Complementing this view, Zeng et al. (2023) noted that loops and conditionals tend to impose higher cognitive loads, as they require children to mentally construct and manipulate abstract rule-based structures, often involving numerical or logical parameters. These elevated demands on working memory and abstraction may partly account for the observed variability in how young learners approach different CT constructs during early schooling.

Beyond concept-specific reasoning demands, additional evidence highlights broader cognitive mechanisms that influence young learners' engagement with CT tasks. For instance, Angeli and Valanides (2020) found that preschool children experienced difficulties with spatial orientation when using robotics tools like the Bee-Bot, particularly when the robot's direction differed from their own. Such challenges in spatial perspective-taking point to concrete cognitive constraints in early CT learning and emphasize the importance of visuospatial transformation for understanding algorithmic processes. Similarly, executive functions, including planning, inhibition, and working memory, have also been shown to play a pivotal role in the acquisition of CT skills. Robledo-Castro et al. (2023), studying students aged 10–11, found that CT interventions significantly improved visuospatial working memory, sequential planning, and cognitive inhibition—core components of goal-directed, structured thinking. Supporting this, Tsarava et al. (2022) observed significant correlations between CT performance and higher-order cognitive abilities, including verbal reasoning, visuospatial processing, and complex numeracy, while basic arithmetic showed weaker associations. These findings suggest that CT proficiency depends more on integrative and flexible cognitive skills than on procedural knowledge alone.

In addition to these cognitive mechanisms, recent longitudinal research advocates for a person-centered approach to capture the inherent heterogeneity in how students acquire CT skills. According to Cheng et al. (2025), children's CT development follows distinct developmental trajectories, which can be categorized into three profiles: "Steady

Climbers" (showing an upward trend), "Consistent Performers" (maintaining stable levels), and "Gradual Decliners" (exhibiting a downward trend). This trajectory-based perspective suggests that CT development is not a uniform maturation process but rather a context-dependent process, significantly influenced by the interaction of cognitive, technological, and teacher-related factors.

Beyond individual cognition, a range of social and contextual factors also influence how CT skills are acquired and expressed. Drawing on social cognitive theory, Lai et al. (2023) argued that cognitive ability, prior experience, self-efficacy, and collaborative learning environments interact dynamically in shaping CT outcomes. Peer interaction, verbalization, and negotiated meaning-making can amplify or compensate for cognitive differences among learners. This perspective is supported by Kjällander et al. (2021), whose case study with first graders revealed that skills in sequencing, repetition, and debugging emerged through multimodal and embodied interactions using visual block-based languages such as ScratchJr. These early learning experiences suggest that CT acquisition in young children is not strictly a function of internal maturation, but also of instructional design, representation formats, and collaborative activity.

2.3. Empirical Patterns in the Development of Computational Thinking

An expanding body of empirical research indicates that CT skills show consistent age-related differences across primary education. Age-related gains have been consistently documented across CT domains, particularly in areas such as sequencing, abstraction, loops, and conditional logic. These findings suggest that differences in CT performance reflect both cognitive maturation and increased exposure to formal instruction.

Large-scale assessments have provided robust evidence for this trend. For instance, Román-González et al. (2017) found statistically significant differences in CT performance across grade levels, with older students exhibiting stronger abilities across multiple CT dimensions. Similarly, H. S. Kim et al. (2021), in a national study of elementary school students, reported a steady increase in performance from Grades 1 to 6, particularly in abstraction and automation. However, their findings also noted a plateau effect in middle school, suggesting that further gains may depend on sustained instructional scaffolding beyond early exposure.

Other studies, such as those by Rijke et al. (2018), confirm that students' performance in abstraction and decomposition tasks improves significantly with age. Although younger students were capable of engaging with CT tasks, older students demonstrated more sophisticated problem representations and solution strategies. Interestingly, younger learners did not report greater difficulty or cognitive load, indicating that early exposure to CT—especially through unplugged or tangible tasks—is both feasible and meaningful.

Curriculum-based intervention studies further support a staged developmental trajectory. Rodríguez-Martínez et al. (2020) and Tengler et al. (2022) showed that even short-term, structured engagements using block-based environments like Scratch significantly enhanced students' performance in sequences, loops, and conditionals. These improvements were more pronounced when instruction was aligned with learners' mathematical development and cognitive readiness.

More specifically, empirical studies consistently report differences in performance across CT concepts, with sequencing typically being more accessible than loops and conditionals. An (2022), in a large-scale study of over 1200 children aged 5 to 7, documented a sharp increase in sequencing ability between ages 5 and 6, followed by more moderate gains between 6 and 7. Similarly, Elkin et al. (2016) found that preschoolers were generally successful in constructing simple, syntactically correct sequences using robotics kits, but showed decreased accuracy on tasks involving repeat loops.

Additional evidence supporting these performance differences is provided by Luo et al. (2022), who compared the CT performance of third- and fourth-grade students. Third-grade learners, who were introduced to sequencing, demonstrated the ability to generate complete, ordered instructions both in coding and problem-solving contexts. In contrast, fourth-grade students, despite exposure to conditionals, showed no consistent understanding of evaluating true/false statements or articulating outcomes. Students also experienced difficulties constructing loop instructions in word problems, indicating that increasing conceptual abstraction and cognitive demands may constrain the acquisition of more complex CT constructs.

Additional evidence from validated assessment instruments reinforces this developmental pattern. Zapata-Cáceres et al. (2020), in their validation of the BCTt, reported significant performance differences across age groups, with younger learners demonstrating stronger mastery of foundational constructs such as sequences, whereas older students showed gradual improvement in more cognitively demanding concepts such as loops and conditionals. These findings provide further empirical support for a hierarchical progression in the acquisition of CT concepts during primary education. de Ruiter and Bers (2022), through the development and validation of the Coding Stages Assessment (CSA), confirmed that learning to code follows a structured developmental trajectory in which mastery of simpler procedural structures, such as sequences, both cognitively and chronologically precedes the acquisition of more complex constructs, including repeat loops and conditional logic.

Despite this general developmental progression, the complexity of loops and conditionals is a persistent challenge. No measurable improvement in children's understanding of "when" conditionals has been observed even after short-term interventions (Pila et al., 2019). Similarly, significantly lower performance has been reported in sensor-based repeat loop tasks compared to simpler sequencing tasks (Sullivan & Bers, 2016). These findings also reveal developmental and gender-related differences, with second graders outperforming first graders and boys demonstrating higher performance in certain tasks. In addition, difficulties in applying conditional logic—particularly when evaluating true/false statements in numerical or input-based contexts—have been documented, reflecting the high cognitive load associated with these constructs (Luo et al., 2022). Instructional design and exposure also play a critical role. Kong and Wang (2023), in a longitudinal study with over 13,000 students, found that sustained engagement with cognitively demanding CT tasks significantly enhanced abstraction and algorithmic thinking. In contrast, Cui and Ng (2021) reported that even upper-primary students had difficulty translating intuitive mathematical strategies into executable code when support was limited. Their findings illustrate that age alone does not ensure the development of CT competence. Rather, structured instructional support and the use of clear, accessible representations play a critical role in helping students transition from intuitive forms of reasoning to more formal computational thinking.

Finally, while a general developmental trend is evident, the field still lacks a universally accepted learning progression model for CT. Fagerlund et al. (2021) emphasized the absence of a consistent taxonomy, particularly in environments such as Scratch, where CT concepts may emerge in nonlinear or overlapping ways. This lack of consensus underscores the need for continued empirical investigation into the developmental trajectories of specific CT concepts, particularly during the early years of formal education, when foundational cognitive and computational competencies are actively emerging.

3. Methodology

3.1. Research Questions

This study forms part of a broader postdoctoral research project conducted at the Pedagogical Department of Primary Education, University of Thessaly, Greece, focusing on the adaptation and implementation of the BCTt in the Greek educational context. While previous work has addressed the test's psychometric validation and grade- and gender-based performance differences, the present study examines the developmental interrelations among core CT constructs and the identification of learner profiles. Specifically, it investigates how primary school students' performance varies across grade levels and foundational CT concepts. Beyond identifying performance differences, the study also examines the extent to which performance across individual CT concepts is interrelated and whether distinct student profiles can be identified. Accordingly, the study addressed the following research questions, focusing on students' performance in three foundational CT concepts (sequences, loops, and conditionals):

1. Are there statistically significant differences in students' performance across the three CT concepts?
2. Are there statistically significant within-grade differences in students' performance across the three CT concepts?
3. Are there statistically significant between-grade differences in students' performance across the individual CT concepts?
4. Are students' performances across the three CT concepts statistically significantly correlated at both the overall sample and grade-specific levels?
5. Are there distinct student profiles based on their performance across the three CT concepts?

By addressing these research questions, the study aimed to provide a comprehensive understanding of how CT skills develop, differentiate, and interrelate during early primary education. In doing so, it also sought to inform the design of targeted instructional practices that align with students' developmental trajectories and cognitive diversity. To support these aims, the study adopted a descriptive-comparative research design (Cantrell, 2011), emphasizing the comparison of student performance across CT concepts without manipulating variables or randomly assigning participants. Rather than pursuing causal or predictive inference, the focus was placed on describing and comparing group characteristics within a naturalistic educational context.

3.2. Participants

Participants were recruited within the broader postdoctoral project involving the Greek adaptation of the BCTt. Participants were drawn from primary schools within the Attica Regional Directorate of Education, the largest and most demographically diverse educational region in Greece. Although not nationally representative, this region includes urban, suburban, and mixed-income areas, providing a diverse cross-section of early primary education contexts. More specifically, a two-stage probability sampling procedure was employed to ensure systematic and unbiased participant selection. In the first stage, we applied a probability proportional to size (PPS) method, which is mostly used when sampling units vary in size and need to be represented proportionally (Cheung, 2014). We used the 13 Regional Directorates of Primary and Secondary Education in Greece as our primary sampling units, sorted them by the number of schools they oversee, and then randomly selected the Attica Regional Directorate. Moving to the second stage, we used a simple random sampling (SRS) approach, where schools were chosen randomly with equal probability (Singh, 2003). Using this method, we compiled a comprehensive list of schools

in the Attica region and, with the help of a random number generator, selected five schools from this list.

To ensure the sample size would be robust enough for meaningful analysis, we aimed to follow the established guidelines in existing research. Sample sizes of 200–300 participants, or around 10 participants per scale item, are often seen as minimums for reliable factor analysis (Boateng et al., 2018), or recommendations often suggest a respondent-to-item ratio of 5:1, up to 30:1, as a safe range (Tsang et al., 2017). Based on these guidelines, we aimed for a solid sample size to support psychometric analysis of the BCTt scale and also used the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy to confirm sample suitability (Arafat et al., 2016).

Initially, data were collected from 673 students (228 males and 445 females) across Grades 1 to 4. For the purposes of the present study, we focused on Grades 1 to 3, resulting in a final sample of 517 students (228 males and 289 females). This decision was informed by preliminary findings from a separate phase of the broader project, in which Grade 4 students showed a ceiling effect, with more than 15% reaching the maximum score, exceeding the commonly accepted threshold (Terwee et al., 2007). This pattern limited the instrument's ability to discriminate between different levels of performance at this grade level and suggests that many students had already consolidated their understanding of the assessed CT concepts. As such, the observed ceiling effect may also reflect a developmental shift, whereby these foundational constructs become less sensitive indicators of variation at higher grade levels.

3.3. Data Collection

3.3.1. Measures

Our measurement of students' CT skills was based on the Greek adaptation of the BCTt (Vourletsis & Politis, 2025), a validated instrument designed for younger learners. The BCTt (Zapata-Cáceres et al., 2020, 2021) builds upon key ideas from earlier CT assessment tools, such as the CTt (Román-González, 2015; Román-González et al., 2017), and is designed to evaluate core CT constructs in early education.

The test consists of 25 multiple-choice items, organized into six item sets targeting three fundamental CT concepts. Specifically, the test includes 6 items assessing sequences, 12 items assessing loops (both simple and nested), and 7 items assessing conditionals, including if-then, if-then-else, and while structures. Importantly, the BCTt does not require prior experience with any programming environment, enabling a focused evaluation of students' CT abilities that is independent of coding background or exposure. The instrument was designed with minimal textual demands in order to be appropriate for young learners. Tasks rely primarily on visual representations and simple instructions, thereby reducing the potential influence of reading ability or language proficiency on students' performance. Representative sample items for each CT concept (sequences, loops, and conditionals) can be found in the original validation study of the BCTt. Higher scores indicate greater mastery of CT concepts.

We employed the Greek-translated and culturally adapted version of the BCTt, which has demonstrated strong psychometric properties in previous studies. To evaluate the reliability of the instrument in the present analytic sample, internal consistency indices (KR-20 coefficients) were calculated for the total scale and each subscale across grade levels. The total scale demonstrated high internal consistency (KR-20 = 0.89 for Grade 1, 0.88 for Grade 2, and 0.80 for Grade 3). Subscale reliabilities were also acceptable: Sequence (KR-20 range = 0.62–0.69), Simple Loop (0.66–0.72), and Nested Loop (0.63–0.74). Conditional subscales, despite containing fewer items, demonstrated adequate internal coherence based

on inter-item consistency. These results confirm the reliability of the instrument for the present study sample.

The Greek adaptation of BCTt has also demonstrated strong psychometric properties in prior validation analyses. Content validity was established through expert review, confirming alignment between test items and intended CT constructs. Empirical analyses showed appropriate model fit indices and discrimination parameters, supporting the instrument's ability to differentiate across levels of student ability. Furthermore, test–retest reliability over a 2–3 week interval confirmed the temporal stability of student performance. Taken together, these findings support the use of the Greek BCTt as a reliable and developmentally appropriate measure of early CT competencies.

3.3.2. Procedure

We conducted the data collection during May and June 2022, following the necessary approvals and in accordance with ethical research standards. The study received official approval from the Directorate of Primary Education of Eastern Attica, which granted permission to carry out the research in the participating public schools. To ensure consistency across classrooms and minimize procedural variation, all sessions were conducted by the same administrator. Before each administration, we provided students with clear and structured instructions, including one example for each CT concept assessed, as specified in the official BCTt protocol. These examples were presented using digital slides to maintain uniform delivery across the sample. During the assessment, students were encouraged to ask clarification questions. We responded only to procedural queries, avoiding any feedback or cues that might influence their answers. The full 25-item version of the BCTt was administered in a single session per class. On average, students in Grades 1 and 2 required about 55 min to complete the test, while those in Grade 3 completed it in approximately 35 min.

Immediately after each session, we entered the students' responses into a spreadsheet and applied a binary scoring system: a value of 1 was assigned to each correct response, and 0 to each incorrect or missing response. This dichotomous coding allowed total scores to range from 0 to 25, and enabled the normalization of subscale scores (sequences, loops, and conditionals) to a 0–1 scale, supporting comparability in subsequent analyses. All sessions were conducted under the researcher's supervision, and students were reminded to review any unintentionally skipped items without receiving any feedback or assistance regarding the correct answer. If a child appeared to skip a question unintentionally, the facilitator gently encouraged them to return to it. This process aimed to ensure that omissions were due to lack of certainty rather than fatigue or disengagement. Since missing responses were not explicitly recorded as a separate category, omissions were coded as incorrect. While this may limit post hoc analysis of missingness, the observed rates of skipped items were minimal.

It is important to note that the study did not involve any instructional intervention. The BCTt was administered as a stand-alone diagnostic assessment to capture students' pre-existing understanding of CT concepts, without prior teaching or training provided by the researchers.

3.4. Data Analysis

Before addressing the research questions, we conducted preliminary analyses to explore the distribution of students' total CT scores and their performance across the three CT concepts—sequences, loops, and conditionals. Given the unequal number of items per concept (6, 12, and 7 respectively), concept-level scores were normalized to a 0–1 scale to allow comparability across constructs. For overall CT performance, we used the total

score (out of 25). All statistical analyses were conducted using IBM SPSS Statistics (version 29.0.1.0; IBM Corp., 2022).

Prior to inferential testing, we assessed key statistical assumptions. Independence was ensured by assigning each student to only one grade group and using a single score per analysis. Outliers were examined using boxplots, with certain extreme values observed in Grade 3 sequence scores. These data points were retained, as they were deemed plausible based on established criteria (Hoaglin et al., 1986; Tukey, 1977). Normality was evaluated using standardized skewness and kurtosis z-scores (H. Y. Kim, 2013), all of which fell within acceptable limits (West et al., 1995). Linearity and bivariate normality were visually confirmed via scatterplots. Overall, the assumptions for parametric analyses were satisfactorily met.

We began our inferential analyses by comparing student performance across the three CT concepts using one-way repeated measures ANOVAs with normalized scores (0–1). These analyses were conducted both at the full-sample level and within each grade level to examine general and grade-specific patterns. When the assumption of sphericity was violated, we applied Greenhouse–Geisser corrections (Maxwell & Delaney, 2004), and Bonferroni-adjusted pairwise comparisons were used for follow-up testing.

To determine whether students' performance in each individual CT concept differed significantly across grade levels, we carried out three separate one-way ANOVAs. In cases where Levene's test indicated heterogeneity of variances (for sequences and loops), Welch's ANOVA and Games–Howell post hoc tests were applied. For conditionals, where the homogeneity assumption was met, we used standard ANOVA with Tukey's HSD. For between-subject analyses, omega squared (ω^2) was reported as the primary effect size index, whereas partial eta squared (ηp^2) was reported for repeated-measures analyses. Effect sizes were interpreted following conventional benchmarks (Cohen, 1988; Lakens, 2013).

Next, we explored the interrelationships among the three CT concepts by conducting Pearson correlation analyses using normalized scores. These analyses were run both for the overall sample and within each grade level. Prior to computation, we verified linearity and the absence of significant outliers through scatterplots. Correlation coefficients were interpreted using Cohen's (1988) benchmarks.

Finally, we sought to identify distinct student profiles based on performance patterns. We used hierarchical clustering (Ward's method, squared Euclidean distance) to explore the cluster structure and inform the selection of the number of clusters; the final solution was then estimated using K-means on standardized (z-scored) concept scores. Standardization ensured that each CT concept contributed equally to the clustering outcome (Kaufman & Rousseeuw, 1990). Cluster solutions with two, three, and four clusters were examined and compared in terms of interpretability, conceptual differentiation, and group size. Based on these criteria, the three-cluster solution was selected as the most appropriate representation of the data. To evaluate the robustness of the clustering solution, the analysis was repeated using alternative initial seeds, yielding highly comparable cluster structures, indicating stability of the identified profiles. Although the sampling included multiple schools, class- and school-level identifiers were not retained during data preparation, as multilevel modeling was not part of the original analytic plan. Consequently, clustering effects could not be modeled, and the analyses assume independence of observations.

4. Results

4.1. Computational Thinking Performance Overview

4.1.1. Total Computational Thinking Score and Grade-Level Differences

We began our analysis by calculating descriptive statistics to provide a general overview of students' overall performance in CT across grade levels (see Table 1). Our

results indicated a steady and gradual increase in total CT scores as students progressed through the primary grades. Specifically, students in Grade 1 achieved an average score of 12.44 (SD = 6.22), those in Grade 2 scored slightly higher with a mean of 13.49 (SD = 5.96), while students in Grade 3 reached a considerably higher mean of 16.35 (SD = 5.03).

Table 1. Descriptive statistics for total CT scores across grade levels.

Grade Level	N	Mean	Median	Std. Deviation	Minimum	Maximum
1	160	12.44	13.00	6.22	0.00	24.00
2	172	13.49	14.00	5.96	0.00	24.00
3	185	16.35	16.00	5.03	7.00	25.00
Total	517	14.19	15.00	5.96	0.00	25.00

In addition to examining differences in average scores, we also explored the variability of total CT scores within each grade level to examine score variability across grade levels. Our findings revealed a gradual reduction in standard deviation from Grade 1 to Grade 3, suggesting that older students not only achieved higher scores but also exhibited lower variability. This pattern suggests a gradual convergence in performance across grade levels.

To examine grade-level differences in total CT scores, we conducted Welch's ANOVA due to violations of the homogeneity of variances. The analysis revealed a statistically significant effect of grade level, Welch's $F(2, 331.74) = 23.69$, $p < 0.001$, with a moderate effect size ($\omega^2 = 0.08$). Post hoc comparisons using the Games–Howell procedure indicated that Grade 3 students scored significantly higher than those in Grade 1 (mean difference = 3.91, 95% CI [2.46, 5.36], $p < 0.001$) and Grade 2 (mean difference = 2.86, 95% CI [1.48, 4.24], $p < 0.001$).

4.1.2. Computational Thinking Concepts Scores

We next examined students' performance across the three core CT concepts assessed in the study: sequences, loops, and conditionals. As shown in Table 2, students achieved the highest average scores in sequences (M = 0.78, SD = 0.25), followed by loops (M = 0.57, SD = 0.27), while conditionals received the lowest scores overall (M = 0.39, SD = 0.31). These descriptive results suggest differences in average performance across CT concepts.

Table 2. Descriptive statistics for student performance across the three CT concepts (N = 517).

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
Sequences	517	0.00	1.00	0.78	0.25	−1.13	0.40
Loops	517	0.00	1.00	0.57	0.27	−0.30	−0.65
Conditionals	517	0.00	1.00	0.39	0.31	0.20	−1.18
Valid N (listwise)	517						

To better understand these differences, we also examined the distribution of scores for each concept using skewness and kurtosis values. Sequence scores exhibited a pronounced negative skew (skewness = −1.13), indicating that scores were concentrated toward the higher end of the scale. In terms of kurtosis, sequences showed a slightly peaked distribution (kurtosis = 0.40), indicating a modest concentration of scores around the mean. Loop and conditional scores were more symmetrically distributed, with skewness values of −0.30 and 0.20, respectively. These two concepts also displayed flatter distributions

(kurtosis = −0.65 for loops and −1.18 for conditionals), indicating greater variability and wider score dispersion.

4.2. Inferential Analyses

4.2.1. Concept-Level Differences

As presented in Section 4.1.2, we observed substantial variation in students' performance across the three CT concepts. The highest mean score was found in sequences (M = 0.78, SD = 0.25), followed by loops (M = 0.57, SD = 0.27), while the lowest performance was recorded in conditionals (M = 0.39, SD = 0.31). To evaluate the significance of these differences, we conducted a one-way repeated measures ANOVA.

Before interpreting the results, we tested the assumption of sphericity using Mauchly's test, which indicated a significant violation, $\chi^2(2) = 40.04$, $p < 0.001$. We therefore applied the Greenhouse–Geisser correction ($\varepsilon = 0.93$) to adjust the degrees of freedom. The ANOVA revealed a significant main effect of CT concept, $F(1.86, 960.18) = 624.28$, $p < 0.001$, with a large effect size ($\eta p^2 = 0.55$). These results indicate that a substantial proportion of the variance in students' performance was explained by differences among the CT concepts.

Pairwise comparisons showed that performance in sequences was significantly higher than in loops (mean difference = 0.21, 95% CI [0.19, 0.24], $p < 0.001$) and conditionals (mean difference = 0.39, 95% CI [0.36, 0.42], $p < 0.001$). Additionally, performance in loops was significantly higher than in conditionals (mean difference = 0.18, 95% CI [0.15, 0.21], $p < 0.001$).

Overall, these findings reveal a clear performance gradient among the three CT concepts. As illustrated in Figure 1, students demonstrated the strongest performance in sequences and the weakest in conditionals, with loops occupying an intermediate position.

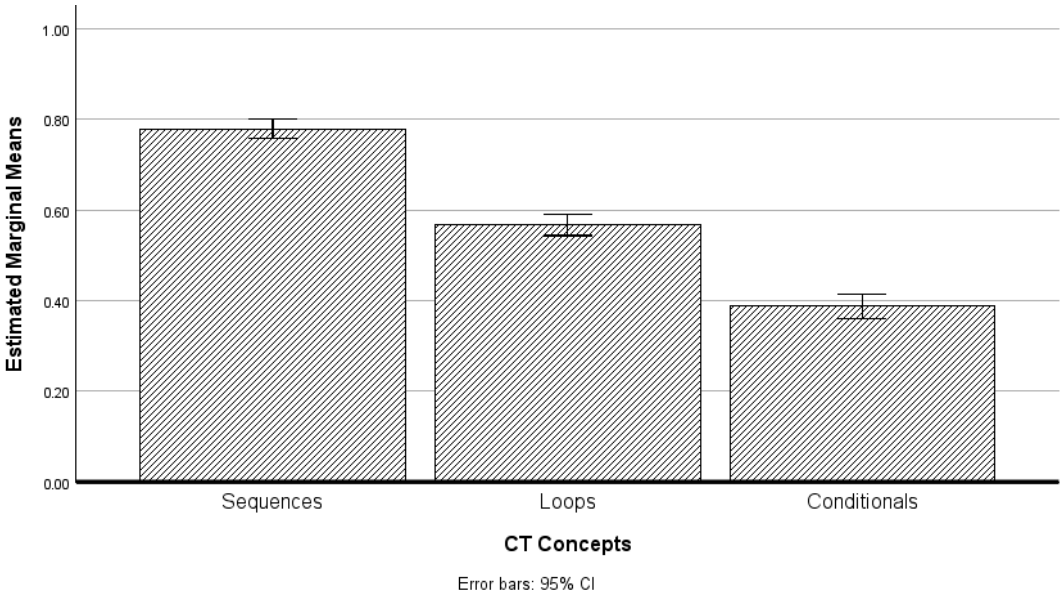


Figure 1. Estimated marginal means for each CT concept with 95% confidence intervals (N = 517).

4.2.2. Within-Grade Comparisons

To investigate whether students' performance differed significantly across the three CT concepts within each grade level, we conducted separate one-way repeated measures ANOVAs for Grades 1, 2, and 3. In each analysis, we treated students' scores across sequences, loops, and conditionals as related measures, and we tested all relevant assumptions, applying corrections where necessary. As shown in Figure 2, students across all grade levels demonstrated the highest mean scores in sequences, followed by loops,

and then conditionals. This consistent performance pattern was statistically confirmed in all three grades through repeated measures ANOVAs and Bonferroni-adjusted pairwise comparisons ($p < 0.001$ in all cases).

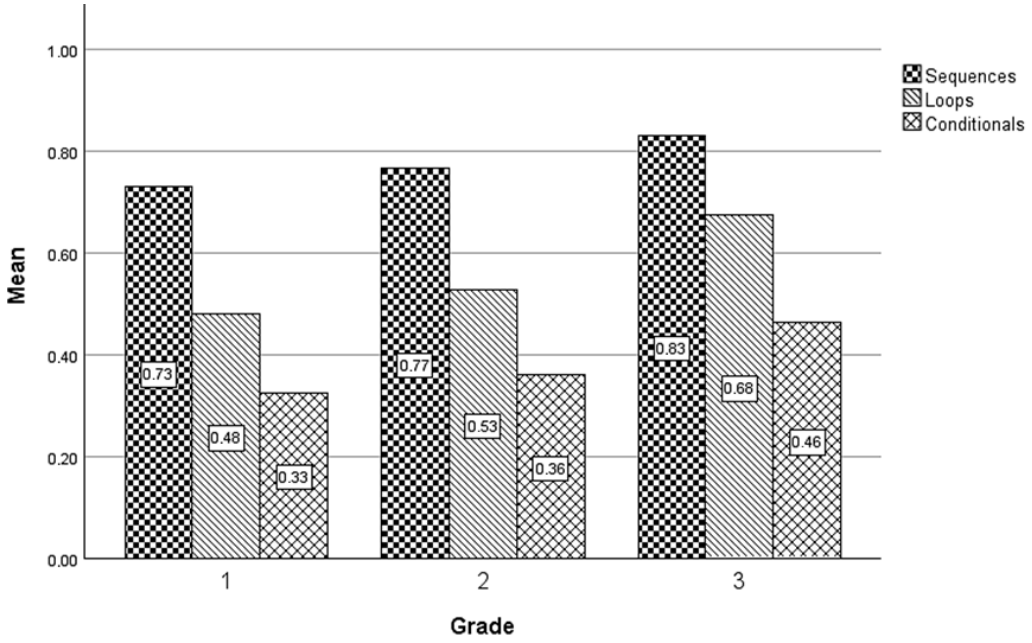


Figure 2. Mean CT concept scores by grade level with 95% confidence intervals (N = 517).

In Grade 1, we found that students performed highest in sequences (M = 0.73, SD = 0.27), followed by loops (M = 0.48, SD = 0.27), and lowest in conditionals (M = 0.33, SD = 0.30). The repeated measures ANOVA with Greenhouse–Geisser correction indicated a statistically significant difference among the three concepts, F(1.81, 288.33) = 234.01, $p < 0.001$, with a large effect size ($\eta p^2 = 0.60$). Post hoc comparisons confirmed significantly better performance in sequences over loops (mean difference = 0.25, 95% CI [0.21, 0.29], $p < 0.001$) and conditionals (mean difference = 0.41, 95% CI [0.35, 0.46], $p < 0.001$), and also in loops over conditionals (mean difference = 0.16, 95% CI [0.11, 0.20], $p < 0.001$).

In Grade 2, the same trend persisted, with mean scores of 0.77 for sequences (SD = 0.26), 0.53 for loops (SD = 0.27), and 0.36 for conditionals (SD = 0.30). The repeated measures ANOVA revealed a significant effect of concept, F(1.87, 320.10) = 247.16, $p < 0.001$, with a large effect size ($\eta p^2 = 0.59$). Bonferroni-adjusted comparisons again showed higher performance in sequences compared to loops (mean difference = 0.24, 95% CI [0.20, 0.28], $p < 0.001$) and conditionals (mean difference = 0.41, 95% CI [0.36, 0.46], $p < 0.001$), as well as in loops over conditionals (mean difference = 0.17, 95% CI [0.13, 0.21], $p < 0.001$).

In Grade 3, students achieved the highest scores in sequences (M = 0.83, SD = 0.22), followed by loops (M = 0.68, SD = 0.23), and then conditionals (M = 0.46, SD = 0.31). The repeated measures ANOVA again yielded a significant effect of concept, F(1.81, 332.81) = 172.37, $p < 0.001$, accompanied by a large effect size ($\eta p^2 = 0.48$). Post hoc analyses revealed that sequences significantly outperformed loops (mean difference = 0.16, 95% CI [0.12, 0.20], $p < 0.001$) and conditionals (mean difference = 0.37, 95% CI [0.32, 0.42], $p < 0.001$), while loops also exceeded conditionals (mean difference = 0.21, 95% CI [0.16, 0.26], $p < 0.001$).

4.2.3. Between-Grades Differences

To assess whether the grade-related patterns observed in the total CT scores (Section 4.1.1) were also evident within individual CT concepts, we conducted separate one-way ANOVAs

for sequences, loops, and conditionals. As descriptive statistics by grade level were presented in Section 4.2.2, we focus here exclusively on the inferential findings.

For the sequences concept, Levene's test indicated a violation of the homogeneity of variances assumption. We therefore applied Welch's ANOVA, which revealed a statistically significant effect of grade level, Welch's $F(2, 332.04) = 7.84$, $p < 0.001$. The effect size was small ($\omega^2 = 0.02$), indicating modest but meaningful differences across grades. Post hoc comparisons using the Games–Howell procedure showed that Grade 3 students scored significantly higher than those in Grade 1 (mean difference = 0.10, 95% CI [0.04, 0.16], $p < 0.001$) and Grade 2 (mean difference = 0.06, 95% CI [0.01, 0.13], $p = 0.028$). No significant difference was found between Grades 1 and 2 ($p = 0.424$).

For loops, Welch's ANOVA again revealed a significant grade-level effect, Welch's $F(2, 332.71) = 30.09$, $p < 0.001$. The effect size was moderate ($\omega^2 = 0.09$), suggesting substantial differences among grades. Post hoc Games–Howell comparisons indicated that Grade 3 students outperformed those in Grade 1 (mean difference = 0.19, 95% CI [0.13, 0.26], $p < 0.001$) and Grade 2 (mean difference = 0.15, 95% CI [0.09, 0.21], $p < 0.001$), with no significant difference observed between Grades 1 and 2 ($p = 0.241$).

For conditionals, Levene's test confirmed homogeneity of variances, allowing us to proceed with a standard one-way ANOVA. The analysis indicated a significant grade-level effect, $F(2, 514) = 9.82$, $p < 0.001$, with small-to-moderate effect size ($\omega^2 = 0.03$). Tukey HSD post hoc tests showed that Grade 3 students scored significantly higher than those in Grade 1 (mean difference = 0.14, 95% CI [0.06, 0.22], $p < 0.001$) and Grade 2 (mean difference = 0.10, 95% CI [0.03, 0.18], $p = 0.005$), while the difference between Grades 1 and 2 was not significant ($p = 0.522$).

4.3. Correlational Analyses

We conducted a series of Pearson correlation analyses to investigate the interrelationships among the three CT concepts assessed in the present study: sequences, loops, and conditionals. Our objective was to determine the extent to which students' performance in one concept was associated with their performance in the others, thereby shedding light on the structural cohesion of CT skills at the primary education level.

As a first step, we examined correlations within the full sample (N = 517), as presented in Table 3. All pairwise relationships were found to be statistically significant and positive ($p < 0.001$), indicating that students who performed well in one CT domain tended to perform well in the others. Based on conventional benchmarks for interpreting correlation strength (Cohen, 1988), the correlation between loops and conditionals (r = 0.64) and between sequences and loops (r = 0.63) can be considered strong, while the correlation between sequences and conditionals (r = 0.50) falls within the moderate-to-strong range. These findings suggest a considerable degree of interconnectedness among the three CT constructs, reinforcing the notion that CT domains may develop in parallel during early schooling.

Table 3. Pearson Correlation Coefficients Among CT Concept Scores (N = 517).

	Sequences	Loops	Conditionals
Sequences	—	0.628 **	0.503 **
Loops		—	0.641 **
Conditionals			—

**. Correlation is significant at the 0.01 level (2-tailed).

To further explore the relationships among CT concepts, we conducted separate Pearson correlation analyses for each grade level, as presented in Table 4. In Grade 1 (n = 160),

all correlations among scores for sequences, loops, and conditionals were statistically significant and positive ($p < 0.001$). The strongest association emerged between loops and conditionals (r = 0.70), followed by sequences and loops (r = 0.69), while the correlation between sequences and conditionals, although comparatively lower, remained substantial (r = 0.55). This pattern indicates moderate to strong conceptual relationships, with loops continuing to play a central role.

Table 4. Pearson Correlation Coefficients Among CT Concept Scores (N = 517).

CT Concept Pair	Grade 1 (n = 160)	Grade 2 (n = 172)	Grade 3 (n = 185)
Sequences—Loops	0.69 **	0.64 **	0.49 **
Sequences—Conditionals	0.55 **	0.53 **	0.39 **
Loops—Conditionals	0.70 **	0.68 **	0.49 **

**. Correlation is significant at the 0.01 level (2-tailed).

In Grade 2 (n = 172), we observed statistically significant and positive correlations among all three CT concept scores ($p < 0.001$). The strongest relationship was found between loops and conditionals (r = 0.68), followed by sequences and loops (r = 0.64). The correlation between sequences and conditionals was also moderate to strong (r = 0.53). This pattern closely mirrors that observed in Grade 1, with loops and conditionals showing the strongest association, followed by sequences and loops, and sequences and conditionals.

In Grade 3 (n = 185), we found statistically significant and positive correlations among all three CT concept scores ($p < 0.001$). The strongest association was observed between loops and conditionals (r = 0.49), followed closely by sequences and loops (r = 0.49), while the correlation between sequences and conditionals was slightly lower but still meaningful (r = 0.39). Although the correlations in Grade 3 were slightly lower than in earlier grades, they remained statistically significant and fell within the moderate-to-strong range, suggesting continued interconnection among CT domains, although the relative strength of associations was lower than in earlier grades.

4.4. Cluster Analysis

To investigate whether distinct profiles of CT performance could be identified among students, we conducted K-means cluster analyses using standardized (z-scored) scores for sequences, loops, and conditionals. We explored K-means solutions with two, three, and four clusters to identify patterns of conceptual performance.

The two-cluster solution yielded a clear dichotomy between high- and low-performing students across all three CT domains. Although statistically distinct and balanced in size, this solution offered limited insight into the diversity of performance patterns present in the sample. The four-cluster solution provided more nuanced profiles—differentiating, for example, between students showing general underperformance and those with specific conceptual difficulties—but introduced smaller subgroups that were less interpretable and unevenly distributed. As a result, we selected the three-cluster solution for further analysis, as it provided a meaningful balance between interpretability, conceptual differentiation, and group size, while also ensuring adequate separation among clusters and sufficient representation of students within each profile. The resulting clusters reflected three distinct student profiles:

- Cluster 1 (n = 99) included students with consistently low scores across all three concepts, indicating overall lower performance across CT domains.
- Cluster 2 (n = 195) consisted of high-performing students who demonstrated strong scores in sequences, loops, and conditionals.

- Cluster 3 (n = 223) comprised students with moderately high performance in sequences but below-average scores in loops and conditionals, suggesting a foundational understanding of CT with difficulties in more complex constructs.

To further characterize the identified profiles, cluster centroids based on standardized concept scores are presented in Table 5. Cluster 1 demonstrated consistently below-average performance across all CT concepts (z = −1.61 to −1.00), Cluster 2 demonstrated consistently high performance (z = 0.68 to 1.02), and Cluster 3 showed relatively stronger performance in sequences (z = 0.13) compared to loops and conditionals (z = −0.17 and −0.45). Distances between cluster centers ranged from 1.86 to 3.70, indicating clear differentiation among clusters in the multidimensional performance space.

Table 5. Cluster Centroids (Standardized Scores).

Cluster	Sequences (z)	Loops (z)	Conditionals (z)
1	−1.61	−1.25	−1.00
2	0.68	0.83	1.02
3	0.13	−0.17	−0.45

Cluster differentiation was further supported by significant between-cluster differences across all CT concepts, as indicated by descriptive one-way ANOVA results: sequences, $F(2, 514) = 546.30$, $p < 0.001$; loops, $F(2, 514) = 346.73$, $p < 0.001$; and conditionals, $F(2, 514) = 534.79$, $p < 0.001$. To examine whether cluster membership merely reflected grade-level differences, we analyzed the distribution of grade levels within each cluster (Table 6). Although cluster membership was significantly associated with grade level, $\chi^2(4) = 20.22$, $p < 0.001$, all three clusters included students from Grades 1, 2, and 3. This indicates that the identified profiles capture within-grade variability and are not reducible solely to age-related progression.

Table 6. Grade-Level Distribution Across Clusters.

Cluster	Grade 1	Grade 2	Grade 3	Total
1	43	36	20	99
2	47	60	88	195
3	70	76	77	223

The clustering solution was re-estimated using alternative initial seeds, yielding highly comparable cluster assignments across alternative initializations, supporting the stability of the three-cluster solution.

5. Discussion

This study examined how early primary school students performed across three foundational CT concepts—sequences, loops, and conditionals—and explored their interrelations and performance profiles. Overall, the findings revealed a consistent pattern in which sequences were the strongest-performing concept, followed by loops and then conditionals. This gradient suggests that these constructs differ in their relative accessibility, with sequencing likely benefiting from its alignment with familiar step-by-step reasoning, while loops and conditionals place greater demands on abstraction and rule-based thinking. This is further reflected in the distribution of sequence scores, which exhibited negative skewness, indicating that a substantial proportion of students performed at the higher end of the scale. As a result, reduced score variability may have limited the sensitivity

of the ANOVA to detect differences related to this construct. These findings align with prior research identifying sequencing as an early-accessible construct and highlighting iteration and conditional reasoning as more cognitively demanding for young learners (Elkin et al., 2016; Sullivan & Bers, 2016; Zeng et al., 2023). Similarly, sequencing has been conceptualized as a foundational "springboard" for the development of more complex computational structures (Luo et al., 2022), while loops and conditionals have been characterized as progressively more challenging constructs in early CT development (Zhang & Nouri, 2019). From a practical perspective, the observed effect sizes indicate that the differences among CT concepts are not only statistically significant but also educationally meaningful. The consistently higher performance in sequences compared to loops and conditionals suggests that students may require more structured instructional support when transitioning to more complex constructs. For teachers, this implies the need to introduce loops and conditionals gradually, using scaffolding strategies that build on students' existing understanding of sequencing. For curriculum designers, the magnitude of these differences supports the sequencing of CT content in a way that aligns with learners' cognitive readiness, ensuring that more abstract concepts are introduced only after foundational procedural understanding has been established.

This conceptual ordering remained stable within each grade, indicating that the relative structure of CT performance was consistent across early primary education. Although older students generally demonstrated stronger overall performance, this consistency suggests that the relationships among the examined constructs are preserved across grade levels. This pattern supports the interpretation that sequences, loops, and conditionals represent distinguishable yet developmentally related components of CT competence. Previous studies have similarly reported that foundational procedural structures tend to be acquired earlier, while more abstract forms of reasoning continue to develop over time (de Ruiter & Bers, 2022; Román-González et al., 2017). From a cognitive developmental perspective, this pattern is consistent with theoretical accounts emphasizing the gradual emergence of abstract reasoning capacities during middle childhood (Piaget, 1964), particularly for constructs such as iteration and conditional logic (Jiang & Wong, 2021). However, the present findings suggest that such improvements may not be uniform across all learners, further supporting the need to consider individual variability in developmental trajectories.

Differences between grades were more pronounced for loops and conditionals than for sequences, suggesting that more complex CT constructs continue to develop throughout early primary education. This pattern is consistent with empirical evidence showing that performance in iteration and conditional reasoning improves with age and cognitive maturity, while sequencing tends to stabilize earlier (An, 2022; Zapata-Cáceres et al., 2020). Similarly, Román-González et al. (2017) and H. S. Kim et al. (2021) documented progressive improvements in CT performance across grade levels, particularly in constructs involving abstraction and algorithmic reasoning. At the same time, the relatively modest differences observed between adjacent grades highlight that development may occur gradually rather than in discrete stages, and that individual variability remains substantial during this period (Fagerlund et al., 2021). The absence of statistically significant differences between Grades 1 and 2 further supports this interpretation, suggesting that the development of foundational CT concepts may follow a gradual progression during the early stages of primary education, with more pronounced differentiation emerging in later grades.

The correlation analyses further indicated that performance across CT concepts was positively associated, suggesting that these constructs share common cognitive foundations. This finding is consistent with research linking CT performance to broader cognitive abilities, including fluid reasoning, working memory, and general problem-solving (Román-González et al., 2017; Tsarava et al., 2022). The observed associations support the

interpretation that CT competence involves overlapping cognitive processes rather than fully independent skills. At the same time, the somewhat weaker associations observed in the higher grade may reflect emerging differentiation as students develop more specific conceptual strengths and weaknesses. This interpretation aligns with developmental perspectives suggesting that cognitive abilities may initially appear more unified and become more differentiated with increasing experience and cognitive maturation (Gerosa et al., 2021; Piaget, 1964), although further longitudinal research is needed to clarify this process in the context of CT.

A particularly noteworthy finding concerns the observed decline in the correlation between sequences and loops across grade levels. This pattern may indicate that, as students progress, these concepts become increasingly differentiated rather than being processed as part of a unified skill set. In the earlier grades, sequences and loops may rely on similar underlying cognitive processes, such as procedural ordering and pattern recognition. However, with increasing experience and cognitive maturation, loops may require more specialized reasoning related to iteration and control structures, leading to a gradual decoupling from basic sequencing skills. This trend provides further support for the view that CT development involves a shift from more integrated to more differentiated cognitive representations over time.

The cluster analysis provided additional insight into the heterogeneity of student performance, revealing distinct profiles characterized by differences in overall performance and conceptual emphasis. In addition to broadly low- and high-performing groups, a distinct profile emerged characterized by relatively strong sequencing alongside weaker loops and conditionals. This pattern suggests that some students may develop competence in concrete procedural structures while still encountering difficulty with more abstract constructs. Similar variability in developmental trajectories has been documented in longitudinal and person-centered research, which highlights that CT development does not follow a uniform pathway but instead reflects diverse patterns shaped by cognitive and contextual factors (Cheng et al., 2025). The presence of these profiles across all grades further indicates that differences in conceptual understanding exist not only between age groups but also within them, reinforcing the importance of considering individual variability in early CT development. Notably, the distribution of students in Cluster 3 across all grade levels suggests that difficulties with more complex constructs, such as loops and conditionals, are not confined to younger learners but may persist across early primary education, indicating variability in individual developmental trajectories.

From an instructional perspective, these profiles suggest the need for differentiated teaching approaches. In particular, students in Cluster 3, who demonstrate relatively strong sequencing skills but weaker performance in loops and conditionals, may benefit from targeted scaffolding that explicitly supports the transition from linear to more complex control structures. This could involve the use of visual representations, step-by-step decomposition of iterative processes, and guided practice with conditional reasoning tasks. More broadly, the identification of distinct learner profiles highlights the importance of adapting instruction not only to students' overall performance levels but also to their specific conceptual strengths and difficulties.

Despite its contributions, this study has several limitations. The sample was drawn from a single geographic region, which may restrict the generalizability of the findings to other contexts with different curricular or technological conditions. Although the Attica region includes a diverse mix of urban, suburban, and socioeconomically varied school settings, it may not fully reflect the educational characteristics of other regions in Greece, particularly rural or less resourced areas. As such, caution is warranted when generalizing the findings beyond similar educational contexts. The cross-sectional design does not

allow causal or developmental inferences at the individual level. Although the BCTt demonstrated satisfactory reliability, formal measurement invariance across grades was not tested, and therefore, potential differences in how items function across grade levels cannot be fully ruled out, so some observed grade differences may partly reflect variations in item functioning. In addition, the multiple-choice format limits insight into students' reasoning processes, and school- and class-level identifiers were not retained, preventing multilevel analyses. Given the nested structure of the data, this may have affected the estimation of standard errors and associated *p*-values and should therefore be considered when interpreting the results. Finally, missing responses were not systematically logged, restricting more detailed analysis of test-taking behavior. Future research should address these issues through longitudinal, multilevel, and process-oriented designs.

Future research should also employ longitudinal and multilevel designs to trace individual developmental trajectories and better account for contextual influences. The use of mixed-methods approaches, such as think-aloud protocols and performance-based assessments, would also help illuminate the reasoning processes underlying students' engagement with more complex constructs like loops and conditionals. Expanding sampling across regions and systematically incorporating classroom- and school-level factors would further strengthen the generalizability and explanatory depth of future findings.

6. Conclusions

This study examined how foundational computational thinking concepts—sequences, loops, and conditionals—develop and interrelate during the early years of primary education. The findings indicate that students' understanding of these concepts follows a structured pattern, with simpler procedural structures appearing more accessible and more complex constructs presenting greater challenges. At the same time, the observed relationships among concepts and the identification of distinct performance profiles suggest that computational thinking is not a uniform ability but a differentiated competence that develops progressively. These results support the view of computational thinking as a cognitively grounded and developmentally contingent construct, reinforcing the importance of examining its emergence through a developmental lens rather than treating it as a single, static skill.

Within the broader educational context, these findings underscore the importance of developmentally aligned instructional and assessment practices that recognize both shared progression patterns and individual variability among learners. While the study provides meaningful insights, its scope is constrained by its cross-sectional design, regional sampling, and reliance on structured assessment formats, highlighting the need for further research using longitudinal, multi-regional, and process-oriented approaches. Future investigations should examine how computational thinking evolves over time and how instructional, cognitive, and contextual factors interact to support or constrain its development. By clarifying how foundational computational concepts emerge in early schooling, this study contributes to ongoing efforts to design educational practices that support coherent and equitable computational thinking development during the formative stages of education.

Funding: This research received no external funding.

Institutional Review Board Statement: Approval to conduct the study in public primary schools was granted by the Directorate of Primary Education of Eastern Attica (May 2022). The study involved anonymous educational assessment, and no personally identifiable data were collected.

Informed Consent Statement: Informed consent was obtained from the parents or legal guardians of all students involved in the study. Participation was voluntary, and students were informed that they could withdraw at any time without any consequences.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of Variance
BCTt	Beginners Computational Thinking Test
CSA	Coding Stages Assessment
KMO	Kaiser-Meyer-Olkin
KR-20	Kuder–Richardson Formula 20
PPS	Probability Proportional to Size
SRS	Simple Random Sampling

References

An, M. (2022). CTST: Development and validation of an sequence ability in computational thinking in early childhood education. In *Proceedings of the 5th international conference on big data and education* (pp. 241–247). Association for Computing Machinery. [CrossRef]

Angeli, C., & Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. *Computers in Human Behavior*, *105*(2), 105954. [CrossRef]

Annamalai, S., Che Omar, A., & Abdul Salam, S. N. (2022). Review of computational thinking models in various learning fields. *International Journal of Education, Psychology and Counseling*, *7*(48), 562–574. Available online: https://gaexcellence.com/ijepc/article/view/3599 (accessed on 7 April 2026).

Arafat, S., Chowdhury, H., Qusar, M., & Hafez, M. (2016). Cross cultural adaptation and psychometric validation of research instruments: A methodological review. *Journal of Behavioral Health*, *5*(3), 129. [CrossRef]

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. [CrossRef] [PubMed]

Brennan, K., & Resnick, M. (2012, April 13–17). *New frameworks for studying and assessing the development of computational thinking*. Annual American Educational Research Association Meeting (pp. 1–25), Vancouver, BC, Canada. Available online: https://scratched.gse.harvard.edu/ct/files/AERA2012.pdf (accessed on 7 April 2026).

Cantrell, M. A. (2011). Demystifying the research process: Understanding a descriptive comparative research design. *Pediatric Nursing*, *37*(4), 188–189. [PubMed]

Cheng, M., Lai, X., Chiu, T. K., Feng, Y., & Sun, D. (2025). Unveiling children's computational thinking developmental trajectories: Profiles and contextual influences in programming education. *Education and Information Technologies*, *30*(18), 26631–26660. [CrossRef]

Cheung, A. K. L. (2014). Probability proportional sampling. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 5069–5071). Springer. [CrossRef]

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Inc.

CSTA & ISTE. (2011). *Operational definition of computational thinking for K–12 education*. Available online: https://cdn.iste.org/www-root/Computational_Thinking_Operational_Definition_ISTE.pdf (accessed on 7 April 2026).

Cui, Z., & Ng, O.-L. (2021). The interplay between mathematical and computational thinking in primary school students' mathematical problem-solving within a programming environment. *Journal of Educational Computing Research*, *59*(5), 988–1012. [CrossRef]

de Ruiter, L. E., & Bers, M. U. (2022). The Coding Stages Assessment: Development and validation of an instrument for assessing young children's proficiency in the ScratchJr programming language. *Computer Science Education*, *32*(4), 388–417. [CrossRef]

Elkin, M., Sullivan, A., & Bers, M. U. (2016). Programming with the KIBO robotics kit in preschool classrooms. *Computers in the Schools*, *33*(3), 169–186. [CrossRef]

Fagerlund, J., Häkkinen, P., Vesisenaho, M., & Viiri, J. (2021). Computational thinking in programming with Scratch in primary schools: A systematic review. *Computer Applications in Engineering Education*, *29*(1), 12–28. [CrossRef]

Gerosa, A., Koleszar, V., Tejera, G., Gómez-Sena, L., & Carboni, A. (2021). Cognitive abilities and computational thinking at age 5: Evidence for associations to sequencing and symbolic number comparison. *Computers and Education Open*, *2*, 100043. [CrossRef]

Grover, S., & Pea, R. (2018). Computational Thinking: A competency whose time has come. In S. Sentence, E. Barendsen, & C. Schulte (Eds.), *Computer science education: Perspectives on teaching and learning in school* (pp. 19–38). Bloomsbury Academic. [CrossRef]

Hazzan, O., Ragonis, N., & Lapidot, T. (2020). Computational thinking. In O. Hazzan, N. Ragonis, & T. Lapidot (Eds.), *Guide to teaching computer science* (pp. 57–74). Springer. [CrossRef]

Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *81*(396), 991–999. [CrossRef]

Hsu, T.-C., Chang, S.-C., & Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education*, *126*, 296–310. [CrossRef]

IBM Corp. (2022). *IBM SPSS statistics for windows* (Version 29.0.0.0) [Computer software]. IBM Corp.

Jiang, S., & Wong, G. K. W. (2021). Exploring age and gender differences of computational thinkers in primary school: A developmental perspective. *Journal of Computer Assisted Learning*, *38*(1), 60–75. [CrossRef]

Kafai, Y., Proctor, C., & Lui, D. (2020). From theory bias to theory dialogue: Embracing cognitive, situated, and critical framings of computational thinking in K-12 CS education. *ACM Inroads*, *11*(1), 44–53. [CrossRef]

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.

Kim, H. S., Kim, S., Na, W., & Lee, W. J. (2021). Extending computational thinking into information and communication technology literacy measurement. *ACM Transactions on Computing Education*, *21*(1), 1–25. [CrossRef]

Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52–54. [CrossRef]

Kjällander, S., Mannila, L., Åkerfeldt, A., & Heintz, F. (2021). Elementary students' first approach to computational thinking and programming. *Education Sciences*, *11*(2), 80. [CrossRef]

Kong, S.-C., & Wang, Y.-Q. (2023). Monitoring cognitive development through the assessment of computational thinking practices: A longitudinal intervention on primary school students. *Computers in Human Behavior*, *145*, 107749. [CrossRef]

Lai, X., Ye, J., & Wong, G. K. W. (2023). Effectiveness of collaboration in developing computational thinking skills: A systematic review of social cognitive factors. *Journal of Computer Assisted Learning*, *39*(5), 1418–1435. [CrossRef]

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 62627. [CrossRef]

Luo, F., Israel, M., & Gane, B. (2022). Elementary computational thinking instruction and assessment: A learning trajectory perspective. *ACM Transactions on Computing Education*, *22*(2), 1–26. [CrossRef]

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data*. Routledge. [CrossRef]

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.

Piaget, J. (1964). Cognitive development in children: Development and learning. *Journal of Research in Science Teaching*, *2*, 176–186. [CrossRef]

Pila, S., Aladé, F., Sheehan, K. J., Lauricella, A. R., & Wartella, E. A. (2019). Learning to code via tablet applications: An evaluation of Daisy the Dinosaur and Kodable as learning tools for young children. *Computers & Education*, *128*, 52–62. [CrossRef]

Polya, G. (1945). *How to solve it*. Princeton University Press. [CrossRef]

Rijke, W. J., Bollen, L., Eysink, T. H. S., & Tolboom, J. L. J. (2018). Computational thinking in primary school: An examination of abstraction and decomposition in different age groups. *Informatics in Education*, *17*(1), 77–92. [CrossRef]

Robledo-Castro, C., Hederich-Martínez, C., & Castillo-Ossa, L. F. (2023). Cognitive stimulation of executive functions through computational thinking. *Journal of Experimental Child Psychology*, *235*, 105738. [CrossRef]

Rodríguez-Martínez, J. A., González-Calero, J. A., & Sáez-López, J. M. (2020). Computational thinking and mathematics using Scratch: An experiment with sixth-grade students. *Interactive Learning Environments*, *28*(3), 316–327. [CrossRef]

Román-González, M. (2015). Computational thinking test: Design guidelines and content validation. In *EDULEARN15 proceedings* (pp. 2436–2444). IATED. Available online: https://library.iated.org/view/ROMANGONZALEZ2015COM (accessed on 7 April 2026).

Román-González, M., & Pérez-González, J.-C. (2024). Computational thinking assessment: A developmental approach. In H. Abelson, & S.-C. Kong (Eds.), *Computational thinking curricula in K–12: International implementations* (pp. 121–142). The MIT Press. [CrossRef]

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior*, *72*, 678–691. [CrossRef]

Selby, C., & Woollard, J. (2013). *Computational thinking: The developing definition* [Project report]. University of Southampton Institutional Repository. Available online: http://eprints.soton.ac.uk/id/eprint/356481 (accessed on 7 April 2026).

Singh, S. (2003). Simple random sampling. In *Advanced sampling theory with applications* (pp. 71–136). Springer. [CrossRef]

Sullivan, A., & Bers, M. U. (2016). Girls, boys, and bots: Gender differences in young children's performance on robotics and programming tasks. *Journal of Information Technology Education: Innovations in Practice*, *15*, 145–165. [CrossRef]

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, 103798. [CrossRef]

Tengler, K., Kastner-Hauler, O., Sabitzer, B., & Lavicza, Z. (2022). The effect of robotics-based storytelling activities on primary school students' computational thinking. *Education Sciences*, *12*(1), 10. [CrossRef]

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. [CrossRef] [PubMed]

Tsang, S., Royse, C., & Terkawi, A. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia*, *11*(5), 80. [CrossRef]

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., & Ninaus, M. (2022). A cognitive definition of computational thinking in primary education. *Computers & Education*, *179*, 104425. [CrossRef]

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Vourletsis, I., & Politis, P. (2025). Greek translation, cultural adaptation, and psychometric validation of beginners computational thinking test (BCTt). *Education and Information Technologies*, *30*(2), 2211–2235. [CrossRef]

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. [CrossRef]

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Sage Publications, Inc.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35. [CrossRef]

Wing, J. M. (2011). Research notebook: Computational thinking—What and why? *The Link Magazine*. Available online: https://openlab.bmcc.cuny.edu/edu-210211-summer-2023-j-longley/wp-content/uploads/sites/3085/2023/06/CT-What-And-Why-copy.pdf (accessed on 7 April 2026).

Yang, W., Su, J., & Li, H. (2024). Empowering young minds: The future of computational thinking and AI education in early childhood. *Future in Educational Research*, *2*(4), 312–317. [CrossRef]

Zapata-Cáceres, M., Martín-Barroso, E., & Román-González, M. (2020, April 27–30). *Computational thinking test for beginners: Design and content validation*. 2020 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal. [CrossRef]

Zapata-Cáceres, M., Martín-Barroso, E., & Román-González, M. (2021). BCTt: Beginners computational thinking test. In *Understanding computing education (Vol. 1). Proceedings of the raspberry Pi foundation research seminar series*. Raspberry Pi Foundation. Available online: www.rpf.io/seminar-proceedings-2020 (accessed on 7 April 2026).

Zeng, Y., Yang, W., & Bautista, A. (2023). Computational thinking in early childhood education: Reviewing the literature and redeveloping the three-dimensional framework. *Educational Research Review*, *39*, 100520. [CrossRef]

Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*, *141*, 103607. [CrossRef]

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