TISER
TECHNOLOGY, INNOVATION AND SPECIAL EDUCATION RESEARCH

www.tiserjournal.com

Computational Thinking Assessment in Education: A Systematic Review

Faysal ÇEKER*, Mardin Artuklu University, Faculty of Economic and Administrative Sciences, Türkiye, ORCID ID: 0009-0005-0297-4890, faysalceker@outlook.com
Bahadır YILDIZ, Hacettepe University, Faculty of Education, Türkiye, ORCID ID: 0000-0003-4816-3071, bahadir@bahadiryildiz.net

Abstract: In this study, empirical studies on the assessment of computational thinking in educational contexts were systematically reviewed. Within this scope, assessment tools, assessed components, indicators, and the cognitive processes associated with these indicators were examined through an integrated analytical framework. The review was conducted in two stages. First, 20 systematic reviews were examined to identify the general tendencies and gaps in the field. Second, 47 empirical studies focusing on CT assessment in educational settings were analyzed through structured data extraction and thematic content analysis in accordance with the PRISMA and Kitchenham guidelines. The findings showed that tests and programming-based tools have a dominant place in the literature. Algorithmic thinking, decomposition, abstraction, and pattern recognition were found to be the most frequently assessed components. However, it was observed that the cognitive processes associated with CT assessment indicators were rarely made explicit. By examining indicators together with related cognitive processes, the study made the relationship between observable assessment measures and underlying cognitive operations more traceable. The results also indicated that process-oriented mathematical tasks involving multi-step problem solving, multiple representations, structural reasoning, and conditional decision-making may provide suitable contexts for making CT-related indicators and cognitive processes more visible. In this respect, the review offers a systematic synthesis of CT assessment research and may provide a basis for more interdisciplinary and context-sensitive assessment approaches.

Keywords: *Computational Thinking, Assessment, Mathematics Education, Systematic Review.*

1. INTRODUCTION

Computational thinking (CT) has become an important construct in contemporary education, and its assessment across different disciplines has become a significant issue. Although CT has been discussed in relation to programming, problem solving, and digital learning environments, it is understood that the tools, indicators, and cognitive processes used in CT assessment need to be examined more systematically. This need is particularly evident when CT is considered beyond programming-based contexts and in relation to fields such as mathematics education. Therefore, this study presents a systematic review of empirical research on CT assessment and examines how existing assessment approaches can inform more process-oriented and context-sensitive perspectives.

2. LITERATURE REVIEW

In this section, the theoretical framework of the study, the rationale for the systematic literature review, and the purpose of the research are presented together. Thus, before the methodology is introduced, the conceptual basis of CT assessment and its relationship with mathematics education are clarified.

***Corresponding Author**
Mardin Artuklu University, Faculty of Economic and Administrative Sciences, Türkiye.

2.1. Theoretical Framework

Computational thinking, which can be simply described as "thinking like a computer scientist," has taken its place among the most important skills of this time. Grover (2018) expressed the 5Cs of the 21st century as critical thinking, communication, collaboration, creativity, and computational thinking. Although Papert (1980) was the first author to use the term CT explicitly in his work, the concept gained its current popularity in educational and scientific communities largely through the work of Wing (2006). Wing expressed the importance of CT as follows: "Ubiquitous computing was yesterday's dream that has become today's reality; computational thinking is tomorrow's reality."

Following Wing's (2006) call, studies on CT increased rapidly across different disciplines; however, "the need to establish a consensus definition of computational thinking remains" (Kalelioğlu et al., 2016; Lockwood & Mooney, 2018). Wing (2006, p. 33) initially defined the concept as follows: "Computational thinking involves using the fundamental concepts of computer science to solve problems, design systems, and understand human behavior." She later revised this definition with contributions from Aho, Cuny, and Snyder (Wing, 2011) and ultimately updated it as follows: "Computational thinking is 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" (Wing, 2014).

Just as there is a need for consensus on the definition of CT, there is also a need to identify its components. While many intertwined components have been defined in the literature, recent studies have highlighted four core components: decomposition, pattern recognition, abstraction, and algorithm design (Dong et al., 2019; Fofang et al., 2020; Google, 2020; Semiawan, 2019; Valenzuela, 2020). These four components, described by Google (2020) using the analogy of the four legs of a table, are widely regarded as essential CT skills in programming contexts and form the core of how CT is commonly assessed in education.

Various approaches and models have been proposed in the literature to assess CT skills. Studies on this topic have generally focused on analyzing the final product (e.g. a game or project). While design-based approaches are prominent, test-based instruments and rating scales have also been suggested (Yeni, 2020). Yıldız (2021) classified studies on CT assessment into three categories: test-based assessment, scale-based assessment, and product-based assessment.

The US National Research Council (NRC), in its Workshop on the Pedagogical Aspects of Computational Thinking, presented three reasons for assessing CT: (1) to review curricula, related materials, and pedagogy, (2) to assess individual progress, and (3) to support teacher training and development (NRC, 2011). However, although a significant number of studies have addressed CT skills, the debate surrounding their assessment is still in its infancy (de Araujo et al., 2016). Challenges in assessing CT make it difficult to make objective decisions about how these skills should be developed (Tosik-Gün & Güyer, 2019).

The fact that most methods used to assess CT are computer and programming-based has led researchers to seek alternative approaches. Kalelioğlu et al. (2016) emphasized that CT can be applied in much broader learning contexts than simply developing computer-based solutions, and they recommended: designing tasks for all fields and age groups; teaching CT skills pedagogically; identifying ways to assess the extent to which students are equipped with these skills.

Mathematical thinking has been conceptualized in the literature from different perspectives; however, it is frequently associated with the application of mathematical concepts, techniques, and processes in the course of solving problems (Gunhan et al., 2009). In mathematics education, problem solving is not regarded merely as a topic of instruction, but also as a method of teaching and a way of thinking (Posamentier & Krulik, 2016). In this respect, mathematical thinking may be understood as a form of cognitively organized activity in which a problem situation is interpreted, relevant relationships are identified, and a solution is constructed and evaluated. Accordingly, mathematical thinking may be considered not only in relation to mathematical content, but also in relation to the cognitive processes through which problem situations are analyzed and resolved.

Mathematical thinking and computational thinking may therefore be regarded as related forms of problem-solving activity. Although these constructs emerge from different disciplinary traditions, both involve interpreting a problem situation, organizing relevant information, identifying structures and relationships, and constructing a solution pathway. This point is particularly important because CT has often been assessed through computer-based, programming-based, or unplugged tasks in the literature, whereas its problem-solving nature suggests that it should not be restricted to digital environments alone (Tosik-Gün & Güyer, 2019). In this way, the relationship between mathematical thinking and CT provides a theoretical basis for considering broader assessment contexts in which underlying cognitive processes can become visible.

Recent studies have addressed the relationship between CT and mathematics education more explicitly. Wing (2006) argued that CT complements and integrates mathematical and engineering thinking, emphasizing that computer science is grounded in formal mathematical foundations. This view was later reinforced by the broader framing of CT as the thought processes involved in formulating problems and expressing solutions in a form that can be carried out effectively by a human or a machine (Wing, 2014). In mathematics education, this relationship has been further elaborated through studies showing that problem solving and the associated thinking processes constitute common ground between computational thinking and mathematical thinking (Kallia et al., 2021). Consequently, three important aspects of CT that need to be addressed in mathematics education have been identified: problem solving, cognitive processes, and transfer (Kallia et al., 2021). From this perspective, mathematics education offers a relevant disciplinary context for examining CT not only through digital production, but also through cognitively demanding forms of mathematical activity.

Within this context, mathematical tasks provide a meaningful basis for the assessment of CT. The notion of academic task, introduced by Doyle (1983, 1988), draws attention to cognitively demanding situations in which learners are expected to organize available conditions and resources in order to achieve a meaningful instructional goal. Adapted to mathematics classrooms, mathematical tasks have been defined as classroom activities designed to focus students' attention on particular mathematical ideas, and their cognitive demand has been treated as a central feature of task design (Stein et al., 1996; Stein & Lane, 1996; Stein & Smith, 1998). Since cognitive demand refers to the kinds of thinking required to carry out a task (Stein & Lane, 1996), mathematical tasks appear particularly important for examining CT in educational contexts. In this way, CT may be assessed not only through digital products, but also through mathematical tasks that make problem solving processes and underlying cognitive operations more visible.

These conceptual differences have direct implications for assessment practices. When CT is defined primarily through programming and computer science, assessment tends to emphasize observable constructs such as algorithms, loops, conditionals, debugging, and code-related performance. By contrast, definitions that frame CT more broadly as a form of problem-solving place greater emphasis on decomposition, abstraction, representation, pattern recognition, modeling, and logical reasoning. Therefore, differences in the conceptualization of CT influence not only which components are assessed, but also which indicators, task types, and cognitive processes are considered relevant in assessment tools. Accordingly, for CT to be effectively integrated into mathematics education, transfer between CT skills and mathematical tasks at different cognitive levels needs to be established within the shared context of problem solving.

2.2. Rationale for the Systematic Literature Review

After the concept of computational thinking (CT) was popularized by Wing (2006), it became the subject of many interdisciplinary studies; however, the measurement and assessment of CT skills have still not been addressed in a sufficiently systematic way. This situation has led to the body of knowledge in this field remaining scattered and limited in terms of transferability to practice. In order to identify these gaps and structure future research directions in the field, systematic literature reviews (SLRs) have come to the fore as an important research area.

In the first stage of this study, 20 SLRs published between 2006 and 2024 were examined and analyzed under 20 themes derived from their content. This scoping review process contributed both to the formulation of the core research questions and to the development of the hybrid methodological structure adopted in the study PRISMA, Kitchenham, and Mayring. The analyses showed that tools used to assess computational thinking were predominantly programming-based, that indicators were only rarely associated with cognitive processes, and that possible transfer to domains such as mathematics education remained limited.

Among the 20 reviewed studies, four focused directly and three indirectly on the assessment of computational thinking. The analysis showed that these reviews remained limited in several respects: assessment tools were not examined at the indicator level, measured components were not associated with mathematical tasks, and the relationships among tools, components, indicators, and cognitive processes were not addressed in an integrated way.

From this perspective, the present study contributes to the literature in several ways: by examining the assessment tools and assessment types used for CT, identifying which computational thinking components are assessed through which indicators, relating these indicators to cognitive processes, and addressing the adaptability of these indicators and processes to mathematical tasks. Thus, CT assessment tools, assessed components, indicators, cognitive processes, and mathematics education are considered together within the framework of the present systematic review.

2.3. Purpose of the Study

The purpose of this study was to examine empirical studies on the assessment of computational thinking in educational contexts in a systematic manner, with attention to the assessment tools used, the skills addressed through these tools, the assessment methods adopted, and the indicators reported in them. Although the study was not limited to mathematics education, assessment approaches developed in different educational fields were also included, provided that they met the specified criteria, since computational thinking has an interdisciplinary structure.

In the review process, the assessed components of computational thinking, the cognitive processes associated with these components, the indicators through which they were represented, and the methods by which they were measured were examined together. The assessment tools were also evaluated in terms of their application contexts, item content, and analytical approaches. At the final stage, the components, cognitive processes, and indicators obtained from the reviewed studies were considered in terms of their adaptability to mathematics education and their possible contribution to future mathematical task design. Mathematics education was therefore considered at a subsequent analytical stage, after the assessment indicators and associated cognitive processes had been identified across the broader educational literature. In this way, an attempt was made to make the gaps in the literature more visible and to provide a basis for the development of new tasks.

2.4. Research Questions

The research questions of this study were structured in line with the purpose and objectives of the research as follows:

1. What assessment tools are used to assess CT in educational contexts?
2. What types of assessment approaches are employed within these tools?
3. Which components of CT are assessed by these tools?
4. What indicators are used in these assessment tools, and which cognitive processes are associated with these indicators?
5. Which indicators are adaptable to mathematics education, and what types of mathematical tasks can be designed based on these indicators?

These research questions were prepared to guide the systematic examination of CT assessment practices. Particular attention was given to indicator-, process-, and component-oriented assessment practices in educational contexts, including authentic assessment practices based on student products.

3. METHODOLOGY

The methodology of the study was structured in two connected stages: the scoping review phase and the main systematic literature review phase. In this section, these stages are presented together with the research questions, search strategy, database selection, inclusion and exclusion criteria, study selection process, and data analysis procedures. Thus, an attempt was made to present the scope of the review, the selection process, and the analytical framework used in the study in a clear and traceable way.

3.1. Scoping Review Phase

In the first stage, 20 systematic literature reviews published or made available online between 2006 and the end of 2024 in indexed and peer-reviewed outlets were examined. These reviews addressed the teaching, integration, or assessment of CT. The corpus was composed of 18 journal articles, one conference paper included because of its influence in the field, and one study selected to represent the Turkish literature.

The reviews were examined using a structured analytical framework developed to ensure that each study was analyzed according to the same set of questions. The framework consisted of 20 categories, which also formed the headings of the data extraction table. These categories were author and publication year, study title, publication outlet, index, country, discipline, study topic, study purpose, research questions, review protocol, target group, publication-year range, databases searched, number of studies reviewed, search terms, inclusion criteria, exclusion criteria, type of analysis, main findings, and highlighted research gaps. Each review was read in full, and the relevant information was recorded under these headings. The findings obtained from the scoping review phase were used to determine the scope, purpose, research questions, and methodological structure of the main systematic literature review.

3.2. Main Systematic Literature Review Phase

In the second stage, studies focusing on the assessment of CT in educational contexts were systematically searched by following the PRISMA (Page et al., 2021) and (Kitchenham 2004; Kitchenham & Charters, 2007) protocols. PRISMA was used to ensure transparency in study selection and reporting, while the Kitchenham protocol was used to guide the processes of planning, data extraction, quality assessment, and classification.

The 20-category analytical framework constructed during the scoping review phase was taken as the starting point for this stage. However, the framework was revised in accordance with the research questions and the characteristics of the empirical studies included in the main review. Categories specific to the analysis of systematic reviews were removed or reformulated, while headings concerning assessment tools, tool types, assessed CT components, reported indicators, associated cognitive processes, validity and reliability information, item or task characteristics, and potential adaptability to mathematics education were added. The revised framework again consisted of 20 headings and was used to extract and organize data from each included study. The data recorded under these headings were then examined through thematic content analysis in accordance with the research questions. The search strategy, databases, search strings, inclusion and exclusion criteria, study selection process, and procedures for merging duplicate records were reported in accordance with the PRISMA protocol.

3.3. Search Strategy

The search strategy was constructed to identify studies addressing how CT is assessed in educational contexts. For this purpose, the terms "Computational Thinking," "Assessment," and "Education"/"Educational" were combined using Boolean logic. The following search string was used: "Computational Thinking" AND "Assessment" AND ("Education" OR "Educational").

Through this search structure, studies focusing on the relationship between CT and assessment were reached, and studies addressing this relationship within educational contexts were included. In this way, non-educational content in the broader literature pool was excluded, and the search was directed to assessment tools and indicators used in teaching, learning, and pedagogical assessment contexts.

3.4. Selection of the Database

In this study, Scopus and Web of Science (WoS) databases were preferred for the SLR. Scopus was considered suitable because it includes a comprehensive archive across a wide range of disciplines, particularly engineering, computer science, natural sciences, and health sciences, and provides detailed citation information on authors and publications. This feature made it possible to conduct a more in-depth and traceable analysis. Web of Science, with its multidisciplinary structure, was also considered important because it provides access to qualified and influential publications in social sciences, humanities, and education. In addition, its citation analysis features were useful for identifying foundational and guiding studies in specific fields. For these reasons, the combined use of Scopus and Web of Science databases was considered compatible with the scope, purpose, and interdisciplinary structure of the study.

3.5. Inclusion and Exclusion Criteria

In line with the purpose and research questions of this study, the studies to be included in the systematic literature review were selected according to predefined inclusion and exclusion criteria. The inclusion criteria consisted of empirical studies published or available online from 2006 until the search date (24 May 2025) that focused on the assessment of CT in educational contexts; studies including assessment tools, methods, or indicators related to CT; articles published in peer-reviewed journals, written in English, and accessible in full text; studies assessing at least one CT component such as abstraction, decomposition, pattern recognition, or algorithmic thinking; studies using structured assessment tools such as tests, rubrics, performance tasks, or interviews; and studies conducted in K-12 or higher education contexts.

The exclusion criteria covered studies outside educational contexts; theoretical or conceptual publications that did not include empirical data or assessment; studies without full-text access; publications written in languages other than English; conference abstracts, books or book chapters, posters, blogs, technical reports, and unpublished theses; studies involving programming or robotics activities that did not include an explicit CT assessment tool, indicator set, scoring criterion, or analyzable assessment evidence; publications that did not provide tools, indicators, or empirical data related to CT measurement; secondary studies such as systematic literature reviews or meta-analyses; and studies that reused an assessment instrument already examined within the review without providing a new adaptation, implementation context, validity or reliability evidence, indicator set, or other analytically distinct contribution.

3.6. Study Selection Process

Following the search protocol established for the study, searches were conducted in the two selected databases on 24 May 2025, yielding 1384 records (WoS: 682; Scopus: 702). The publication period used in the main systematic review extended from 2006 to the search date (24 May 2025), whereas the preceding scoping review covered studies published or made available online between 2006 and the end of 2024. Based on the predefined eligibility criteria and the filters available within the databases, 1140 records were excluded at the database level before manual screening. No external automation or machine-learning tool was used in this process. Since 59 of the remaining records were common to both databases, they were removed as duplicate records. As a result of the abstract screening of the remaining 185 studies, 128 studies were excluded in accordance with the inclusion and exclusion criteria. Full-text access was obtained for the remaining 57 studies.

During the Web of Science search process, full-text access was kept open. For this reason, some studies that could be accessed through publishers or university databases could not be reached through the search. Therefore, backward reference searching was carried out, and 36 studies were included later. After the abstract screening of these studies, six were excluded. Thus, 87 studies were selected before full-text review.

As a result of the detailed analysis carried out on the full texts of these studies, 40 studies were excluded by associating them with the exclusion reasons. Consequently, 47 studies were selected and reported for detailed analysis. The flow diagram prepared in line with the PRISMA (2020) search protocol is given in Figure 1.

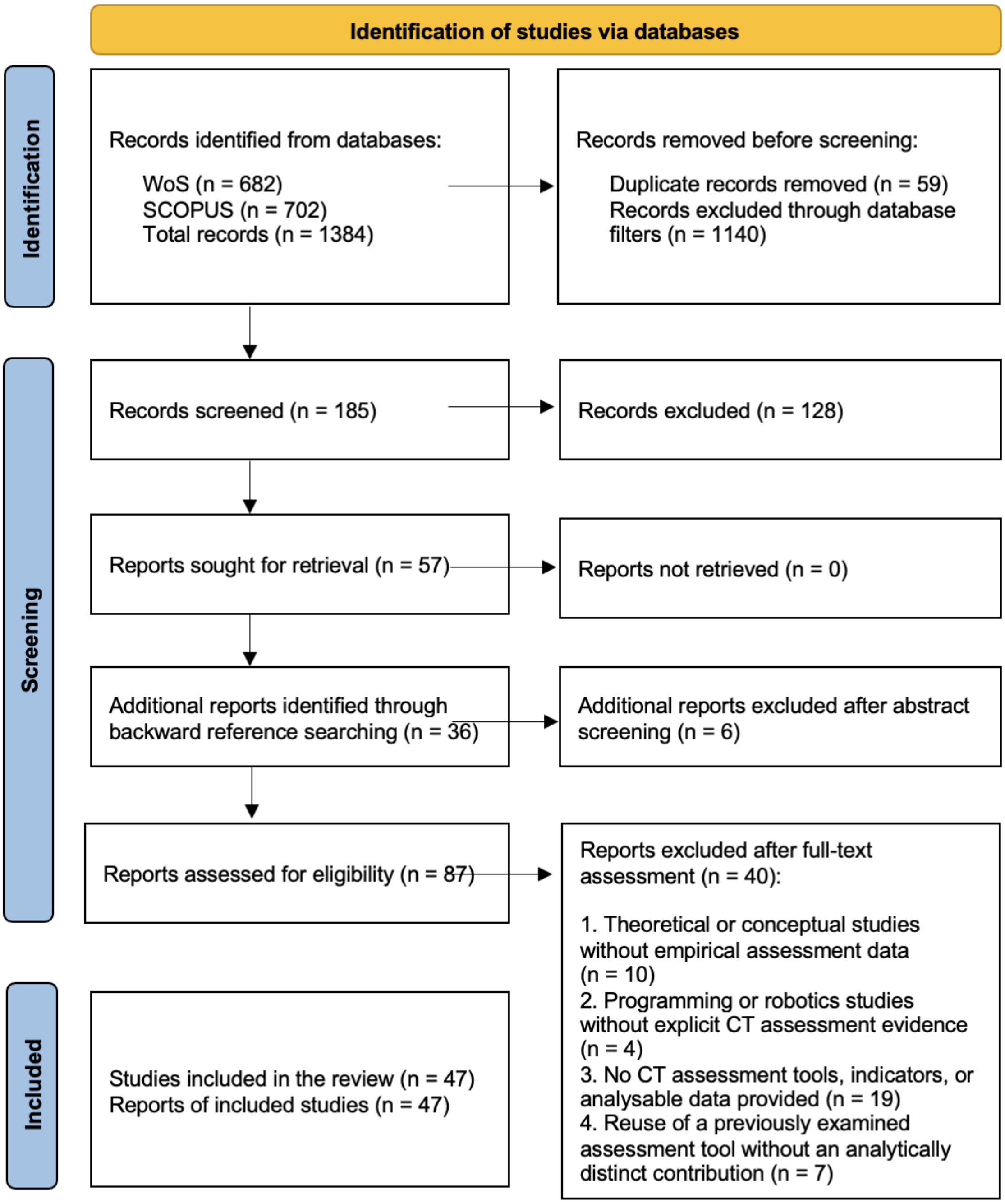

Figure 1. *PRISMA Flow Diagram of Study Selection Process*

3.7. Data Analysis, Validity and Reliability

The data obtained in both phases of the study were analyzed through structured data extraction and thematic content analysis. In the scoping review phase, each systematic review was examined according to the same 20 analytical headings, which also formed the structure of the data extraction table. In the main systematic literature review phase, this initial framework was revised in accordance with the research questions and the characteristics of the empirical studies. The updated headings were used to organize the extracted data, while recurring patterns and related categories were examined through thematic content analysis.

The analysis process was not conducted as a strictly linear procedure. Repeated reviews and revisions were carried out at different stages in order to strengthen the internal coherence of the findings. Each study was examined according to the updated 20 analytical headings, and the extracted data were recorded in structured tables. Recurring patterns were identified, and related categories were brought together in accordance with the research questions. The cognitive processes associated with the assessment indicators were not always stated explicitly in the reviewed studies. When a cognitive process was reported directly, the terminology used in the source study was recorded. When it was not stated explicitly, the process was identified through an interpretive examination of the assessment indicator, task or item structure, rubric criteria, performance description, component definition, and the observable behavior expected from the student. Accordingly, the cognitive process statements represent a structured interpretive synthesis rather than a list reproduced directly from the source studies. Descriptive statistics were used where necessary to present the findings in a clearer and more comparable form.

To strengthen the validity, consistency, and traceability of the analysis, the analytical headings, category definitions, source studies, and representative data excerpts were documented systematically. The extracted data, classifications, and interpretations were reviewed repeatedly at different points in time and compared with the original source texts. Inconsistencies identified during these checks were reconsidered and revised.

Through this integrated analytical approach, the assessment tools reported in the studies, the CT components measured, the indicators identified, and the cognitive processes associated with these indicators were examined comparatively. Particular attention was given to the relationship between the indicators embedded in assessment tools and the cognitive processes underlying them. In addition, the indicators and processes identified in the reviewed studies were considered in terms of their potential adaptability to mathematics education.

4. FINDINGS

In this section, the findings obtained from the scoping review phase and the main systematic literature review phase are presented. Detailed publication information for the studies examined in both phases and the tables containing the indicators and related cognitive processes are provided in the Supplementary Material.

4.1. Findings of the Scoping Review

Detailed publication information for the 20 systematic reviews examined in this phase is available in Supplementary Table S1. The distribution of the reviews by publication year, country, academic field, database, index, and target group is summarized in figures, and the main findings and highlighted gaps across the included reviews are briefly presented.

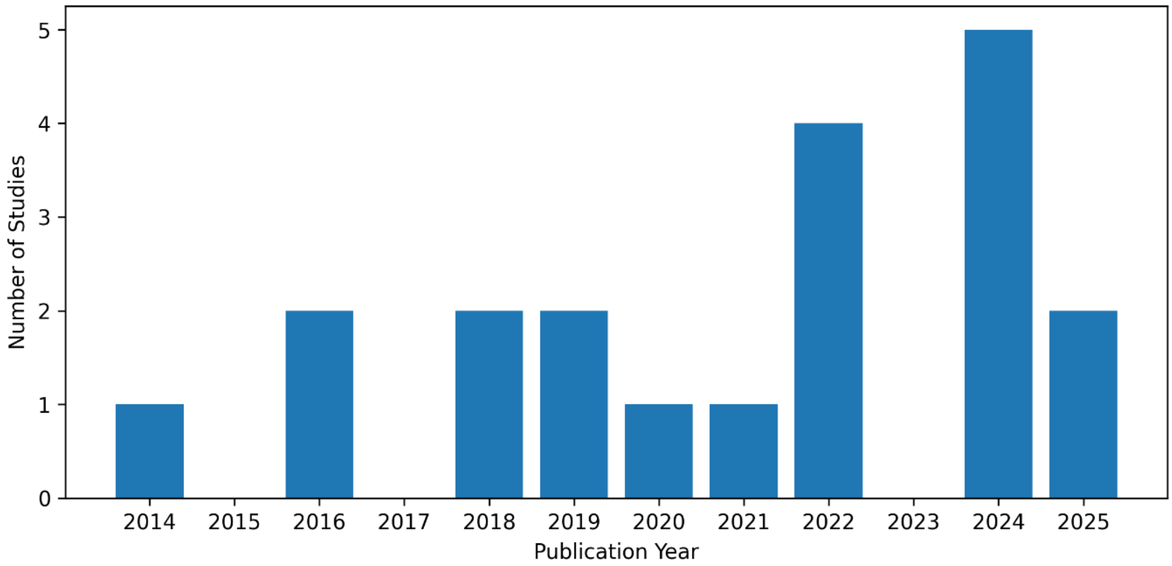

Figure 2. *Distribution of SLRs by Publication Year*

Figure 2 presents the publication-year distribution of the systematic literature reviews examined. The reviews were published between 2014 and 2025, with the two studies shown under 2025 having been accepted in 2024 and published online in 2025.

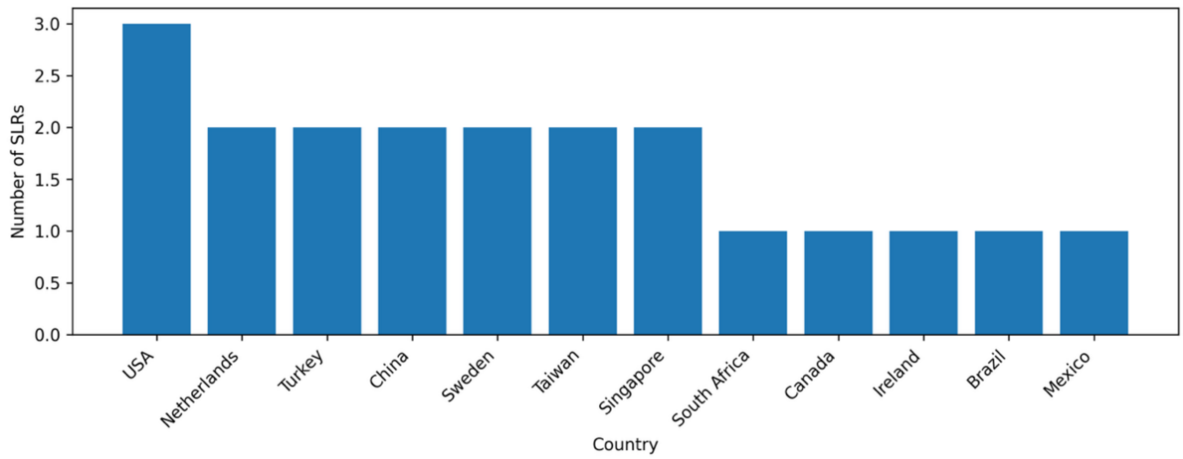

Figure 3. *Distribution of SLRs by Country*

Figure 3 presents the distribution of the systematic literature reviews examined by country. The country variable was determined primarily based on the first author's institutional affiliation at the time of publication. The findings indicate that SLRs are predominantly concentrated in certain countries. While the United States accounts for the highest number of publications, it is also observed that several countries from Europe and Asia have contributed to the literature. Africa, on the other hand, remains underrepresented.

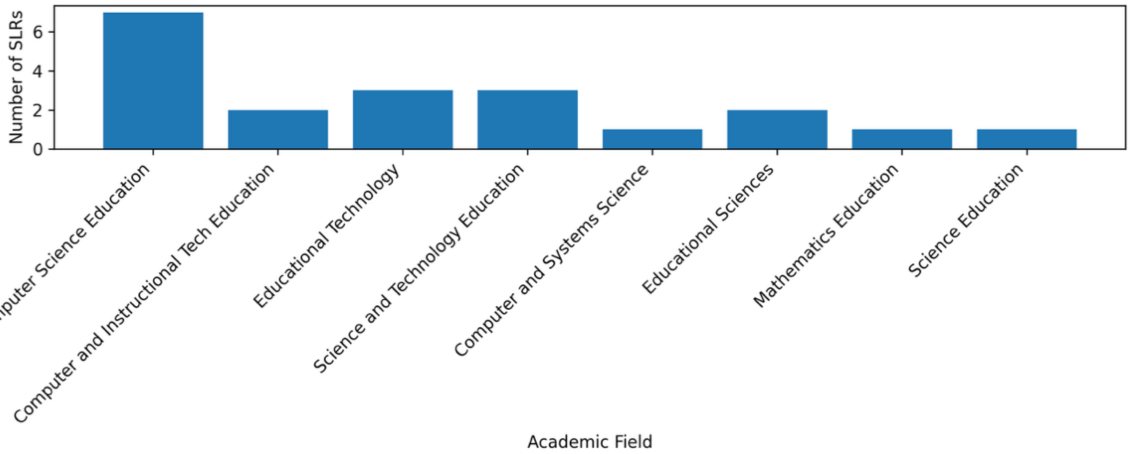

Figure 4. *Distribution of SLRs by Academic Field*

Figure 4 presents the distribution of the reviewed systematic literature reviews by academic field. The findings showed that reviews on computational thinking were mostly concentrated in computer science and related disciplines. Education sciences and science education were also represented to a certain extent, whereas mathematics education remained limited in terms of the number of studies.

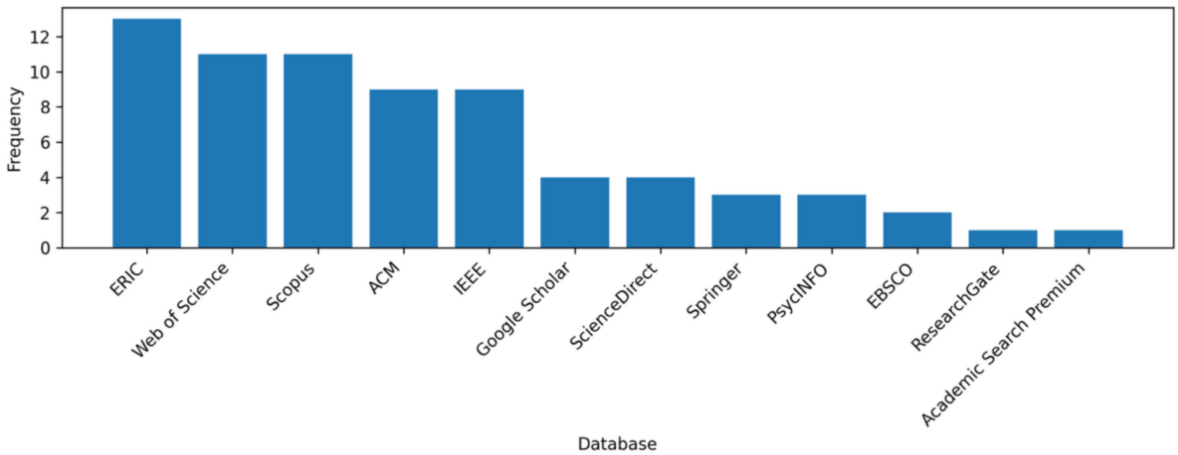

Figure 5. *Distribution of SLRs by Databases*

Figure 5 presents the distribution of databases used in the systematic literature reviews examined. The findings revealed that ERIC was the most frequently used database. WoS and Scopus followed ERIC, while ACM Digital Library and IEEE Xplore formed the next group. The remaining databases were used less frequently.

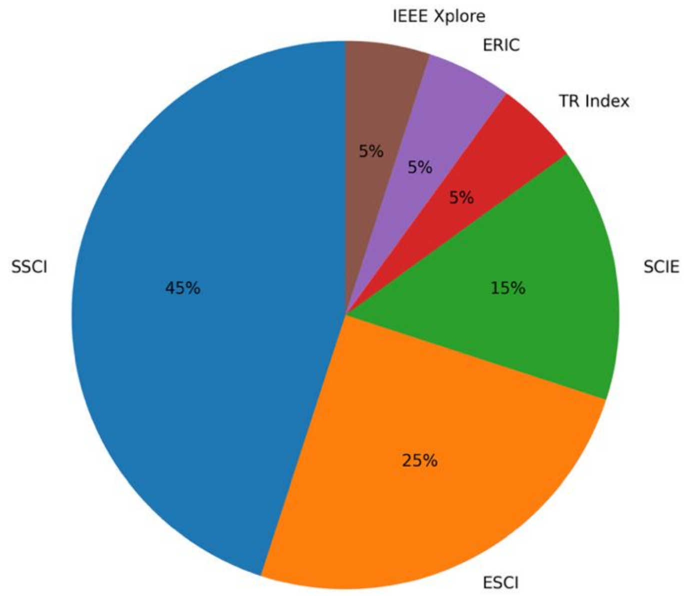

Figure 6. Distribution of SLRs by Indices

Figure 6 presents the distribution of the reviewed systematic literature reviews by indexing categories. The findings showed that approximately half of these studies were indexed in SSCI, while most of the remaining studies were indexed in ESCI and SCIE. Other indices were represented to a more limited extent.

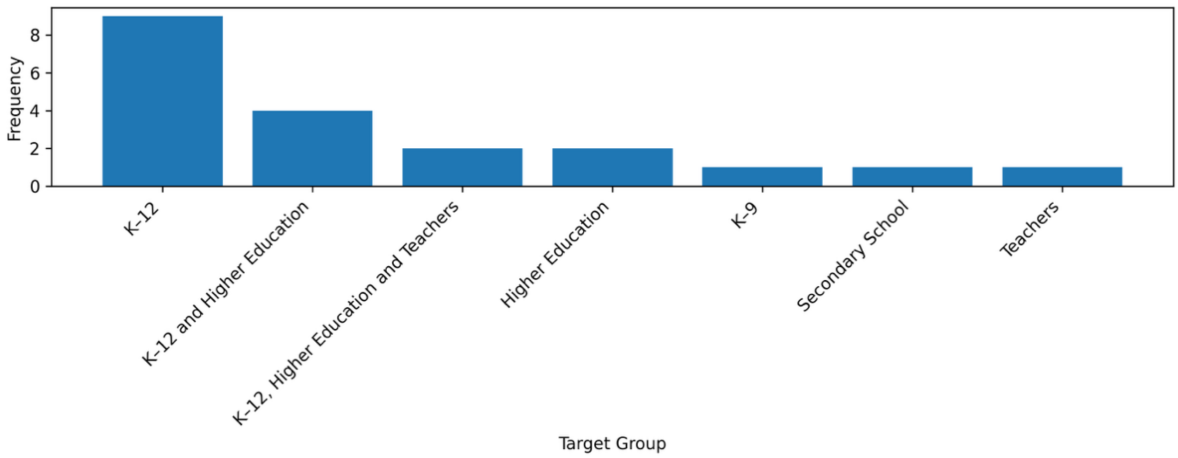

Figure 7. *Distribution of SLRs by Target Group*

Figure 7 presents the distribution of the systematic literature reviews examined by target group. The findings indicate that the reviews span different educational levels; however, the majority of SLRs targeted the K-12 level. Studies focusing on higher education and teachers were relatively limited.

The analysis of the 20 reviewed SLRs showed that both theoretical and practice-oriented developments have continued in the field of CT. The main findings of these studies and the gaps they highlighted are summarized below.

Yeni et al. (2024), who examined the integration of CT into other disciplines in order to address field-specific challenges, stated that CT integration has been carried out mainly in science and mathematics through constructivist strategies, block-based programming, and physical computing devices. Tosik-Gün and Güyer (2019) reported that the most frequently assessed components were abstraction, algorithmic thinking, and decomposition; that tasks, tests, and interviews were prominent in data collection; and that the tools used were mostly programming-based (e.g. rubrics, Dr. Scratch). When these two studies were considered together, it was understood that visibility remained limited in the social sciences, that validity and reliability were reported only to a limited extent in most studies, that process-oriented and non-programming assessments were rare, and that there was a strong concentration at the K-12 level.

Kalelioğlu et al. (2016), who stated that the CT literature is still in a stage of maturation, reported that abstraction, algorithmic thinking, and problem solving are the concepts most strongly associated with CT. They also stated that CT has been framed as a problem-solving process and argued that the current CT concept and its definition lack sufficient scientific justification. Tang et al. (2020) revealed that assessment practices have been concentrated at the elementary and middle school levels; that tests and portfolios have been the most common tools; and that validity and reliability reporting has appeared in only a limited number of studies. The common findings showed that high school, higher education, and non-formal education contexts, as well as interdisciplinary assessments and standardized and qualitative approaches, have remained insufficiently represented.

Ezeamuzie and Leung (2022) determined that there is still no consensus regarding the definition of CT. In their review, it was observed that most definitions were programming-based, while some definitions adopted a more cognitive orientation, and that this situation created a lack of conceptual clarity. The researchers also stated that the absence of a clear distinction between CT and programming points to an important gap in the literature. Åkerfeldt et al. (2024) reported that research on programming and CT has increased and that seven research strands have emerged, including assessment, instructional strategies, teacher competencies, and student attitudes. Based on these studies, it can be stated that the theoretical framework of CT needs to be strengthened and that further research is needed on data literacy, artificial intelligence integration, interdisciplinarity, and assessment methods.

Ogegbo and Ramnarain (2022) showed that abstraction, algorithmic thinking, and debugging have become prominent in science classes; that tools such as Scratch and model-, project-, and inquiry-based approaches have become widespread; and that portfolios have frequently been used. Zhang and Nouri (2019) reported that all CT skills within the framework proposed by Brennan and Resnick (2012) can be taught through Scratch, that product types have diversified, and that age-related cognitive development affects the understanding of CT. When considered together, these studies showed that there is a need to strengthen materials and assessment structures in science education, to provide clearer reporting for teachers, and to make the ethical dimension more visible.

Lu et al. (2022) reported that most studies were conducted in the United States, in STEM and undergraduate contexts; that the tools used were cognitive-focused and mostly computer-based or manual; and that algorithms, problem solving, data, logic, and abstraction were the most frequently assessed dimensions. Wang et al. (2022), on the other hand, stated that domain-specific definitions have developed, that problem-based learning has become widespread, and that CT-STEM integrated assessment and equity-focused studies have remained limited. These findings brought to the fore the need for valid tools adaptable to different contexts, the inclusion of affective dimensions, and more inclusive and experimental studies, together with the need for a common database.

Hsu et al. (2018), who stated that researchers have not yet fully determined how CT should be taught, reported that CT studies have increased rapidly; that applications have concentrated in programming, computer science, biology, and robotics; that project-, problem-, collaborative-, and game-based learning have been widely used; and that strategies such as scaffolding, storytelling, and aesthetic experience have been recommended. Lockwood and Mooney (2018), who stated that CT is a competence beyond literacy and much more than programming, reported that CT is still in an early stage in educational studies. These researchers indicated that CT can be taught in different subjects and can contribute to daily and academic life; however, teacher education and administrative barriers were reported as important limitations. Both studies showed that age-appropriate assessment tools need to be developed, educator training needs to be strengthened, and practice-oriented models need to be improved. de Araujo et al. (2016) reported that programming courses have remained the dominant approach; that Scratch has frequently been used; that abstraction, algorithm development, and problem solving have been the most frequently assessed skills; and that code analysis and multiple-choice tests have been the most common tools. Kallia et al. (2021), on the other hand, highlighted skills related to CT in mathematics education, such as abstraction, decomposition, pattern recognition, algorithmic thinking, modeling, logical reasoning, and automation, as a result of their Delphi study; they also drew attention to analytical thinking, generalization, and strategy evaluation. These findings indicated that non-programming CT approaches, methodological transparency, and authentic tasks in mathematical contexts need to be strengthened.

Lye and Koh (2014) reported that visual programming and digital story and game production are widespread in K-12; that constructivist designs are dominant; and that programming-based CT teaching has generally produced positive outcomes. Tariq et al. (2025), on the other hand, reported that project-based learning is the most effective approach; that collaborative, inquiry-based, and interdisciplinary designs improve analytical and critical thinking; and that real-world problems increase motivation. These two studies showed that there is a need for more robust effectiveness studies, including the integration of innovative approaches such as data enrichment methods (screen recording, think-aloud) and artificial intelligence.

Yun and Crippen (2025) revealed that the integration of CT into science education is aligned with contemporary scientific practices and that equipping preservice teachers with these skills is of critical importance for sustainability at the K-12 level. Yin et al. (2024), on the other hand, stated that guidance has been provided to teachers on material development and the design of rich CT experiences, and that the classroom implementation of collaborative learning has been analyzed comprehensively. When these studies were considered together, the need for experimental and large-sample studies, unplugged activities, design-based approaches, and the development of scalable models in different contexts was emphasized.

Liu et al. (2024) stated that CT occupies a central place in teacher education; that classroom integration has been supported through professional development (PD) models; and that assessment approaches have been classified according to the Kirkpatrick and Desimone frameworks. Weng et al. (2024), on the other hand, revealed that activities presenting AI and CT together support students' product and conceptual development and that CT conceptually deepens AI learning. These two studies showed that it is necessary to establish structures that will support the sustainability of PD programs and teacher communities and to develop conceptual and pedagogical frameworks for AI-CT integration.

4.2. Findings of the Main Systematic Literature Review

Detailed publication information for the 47 studies included in the main systematic literature review is provided in Supplementary Table S2. In the main text, figures summarizing their distribution by publication year, country, and discipline are presented.

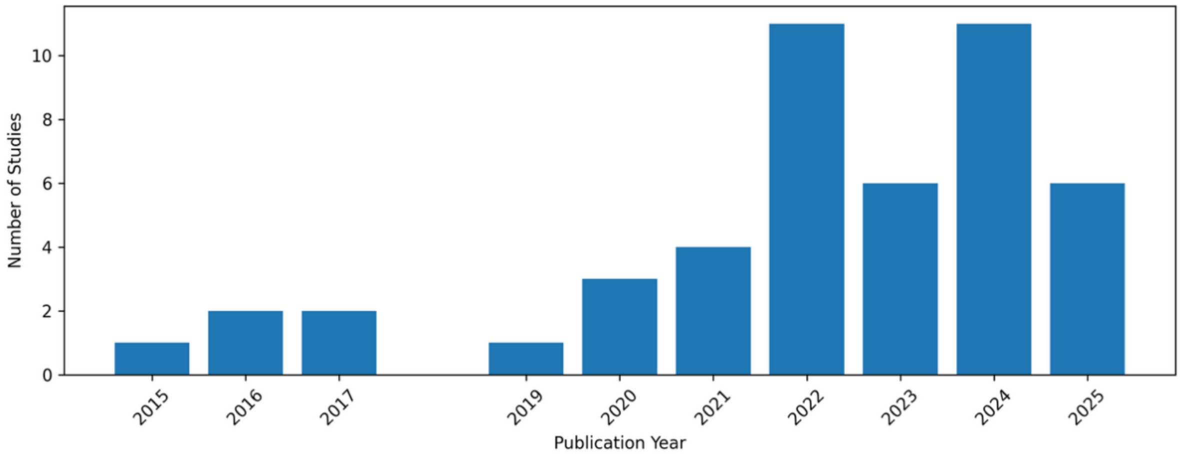

Figure 8. *Distribution of Studies by Publication Year*

Figure 8 shows the distribution of the 47 studies by publication year. The figure shows that 41 of these studies were published within the last five years. The six studies listed under 2025 consist of publications from the first five months of the year.

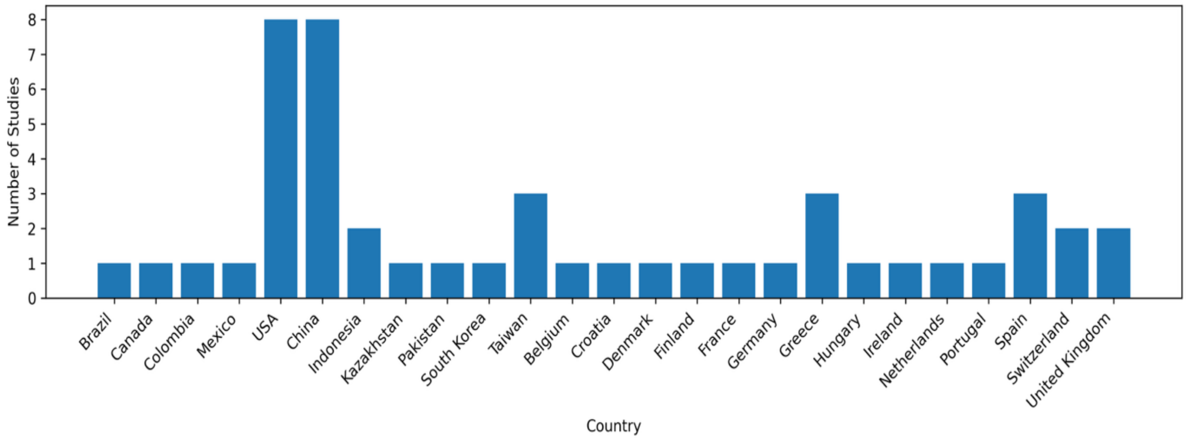

Figure 9. Distribution of Studies by Country

Figure 9 shows the distribution of the 47 studies by country. Approximately one-third of the reviewed studies were concentrated in the United States and China. In determining the country variable, the institutional affiliation of the first author at the time of publication was taken as the main criterion; where necessary, the majority principle was applied.

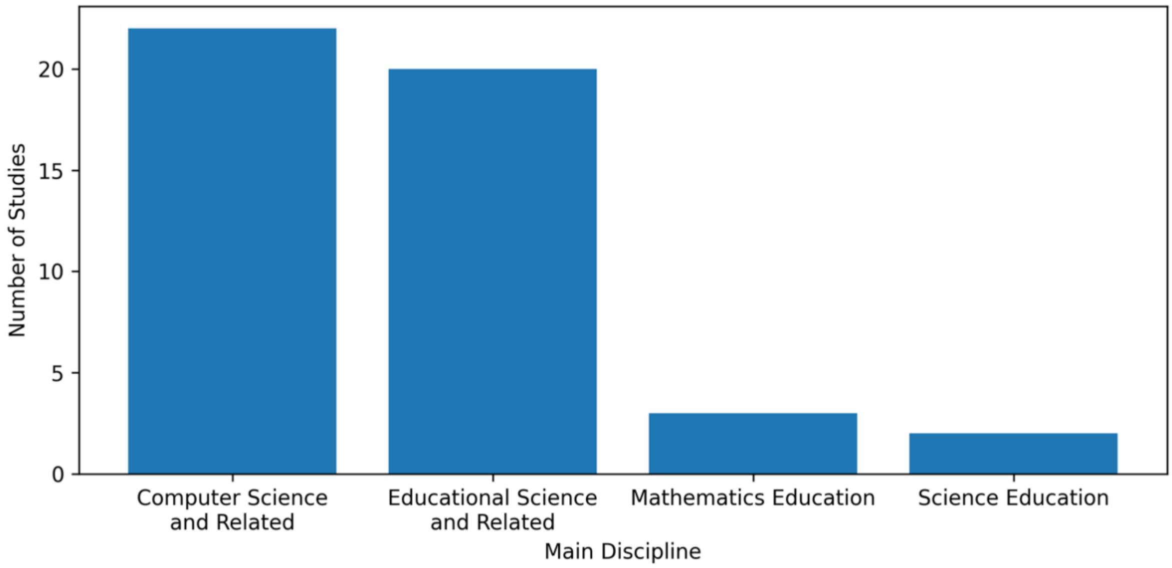

Figure 10. *Distribution of Studies by Main Discipline*

Figure 10 presents the distribution of the studies examined by main discipline. Although the fields of study represent different variations of the same discipline, the reviewed studies were grouped under four main disciplines. The findings show that the vast majority of studies are related to computer science and associated fields.

This study included only publications indexed in the WoS and Scopus databases. Since both databases apply established selection criteria for journal coverage, no additional index-based classification was used in the main systematic review.

4.2.1. Assessment Tools and Types in Computational Thinking

The 47 studies reviewed indicate that a wide range of assessment tools have been employed to assess computational thinking in educational contexts. Based on their content and mode of implementation, these tools were classified into seven categories (see Supplementary Table S3): (1) multiple-choice and diagnostic tests, (2) rubric-based assessment tools, (3) performance tasks and product analysis, (4) questionnaires and self-report scales, (5) automated analysis systems, (6) observation- and interview-based tools, and (7) open-ended questions and discourse analysis. In some studies, multiple tool types were used together in order to develop more comprehensive and multidimensional assessment approaches.

Tests, particularly multiple-choice formats, appeared mostly in studies conducted with children and adolescents (e.g. Li et al., 2021; Román-González et al., 2017; Shen et al., 2025). Diagnostic tests were used to identify students' strengths and weaknesses in CT (e.g. Zhang & Wong, 2023). Rubric-based approaches were generally preferred when student products and learning processes were evaluated (e.g. Gane et al., 2021; Zhong et al., 2016).

Performance tasks were more visible in programming-, problem-solving-, and modeling-oriented activities (e.g. Aristizábal-Zapata et al., 2024; Kaleem et al., 2024). In addition to these tools, questionnaires and self-report scales were used to examine students' perceptions, attitudes, and self-efficacy regarding CT. Within this group, the Computational Thinking Scale (CTS) developed by Korkmaz et al. (2017) was among the most frequently adapted tools (e.g. Huda & Rohaeti, 2024; Pirzado et al., 2025).

Automated analysis systems were mainly encountered in the assessment of programming-based student products. In this group, tools such as Dr. Scratch, Scrape, and Quality Hound were used especially in Scratch-based projects (e.g. Altanis & Retalis, 2019; Romero et al., 2017). Observation- and interview-based tools, however, were directed more toward the contextual and process-oriented dimensions of CT (e.g. Manfra et al., 2022; Xiang et al., 2025).

Open-ended questions and discourse analysis were used to examine students' reasoning processes and conceptual depth. In some studies, mixed assessment designs were also formed by bringing together different tool types (e.g. Altanis & Retalis, 2019; Arastoopour-Irgens et al., 2020). This methodological diversity reveals the need for holistic assessment approaches that can capture both observable performance indicators and the cognitive processes underlying CT.

4.2.2. Assessed Components in Computational Thinking

The summary frequency distribution of the CT components identified across the 47 studies is given in Table 1. The distribution of these components by study is presented in Supplementary Table S4.

Table 1. *Distribution of Assessed Components*

Component	Frequency (f)
Algorithmic Thinking	45
Decomposition	27
Abstraction	25
Pattern Recognition	19
Data Representation	15
Generalization	13
Evaluation	12
Debugging	12
Modeling	7
Logical Reasoning	5
Hardware / Software Awareness	4
Operators	2
User Interactivity	2

According to the reviewed studies, CT components were not distributed homogeneously in the literature. Some components were encountered more frequently than others. Algorithmic thinking (45), decomposition (27), abstraction (25), and pattern recognition (19) were identified as the four components most frequently assessed through measurement tools. This distribution was considered to be close to the four-component model of computational thinking proposed by Google (2020).

The frequency of algorithmic thinking was related to the fact that this component was treated in many studies as a broad mode of thinking including sequencing, loops, conditional structures, and control flow (Aristizábal-Zapata et al., 2024; El-Hamamsy et al., 2022; Eloy et al., 2022; Fagerlund et al., 2020; Gane et al., 2021). Decomposition was generally addressed as breaking complex problems into smaller and more manageable parts. Abstraction was mostly explained as focusing on the essence of a problem by leaving aside unnecessary details. Pattern recognition was also identified as an important component, especially in problem-solving contexts involving data-driven and structural analysis.

The component framework was not determined solely on the basis of frequency. Decisions were also informed by the breadth and consistency of use in the CT literature and by the way each construct was conceptually defined across studies. Constructs such as problem solving were not coded as separate components because they were treated as overarching processes that extend across the entire CT framework rather than as distinct assessment components. Similarly, broader higher-order skills such as creativity and critical thinking were not included as independent CT components, since they were regarded as related but conceptually broader capacities. By contrast, constructs were retained as separate components when they had a sufficiently distinct conceptual definition, a recognizable functional role in assessment, and consistent use across the literature.

During coding, processes expressed with different terms but having conceptual overlap were recoded under broader component categories. For example, the notion of model-based explanation used by Arastoopour-Irgens et al. (2020) was recoded under modeling. Expressions such as representation reading, representation translation, and data use were considered within data representation. In the same study, causality building was evaluated under logical reasoning. Sequencing, loops, and conditional structures were classified under algorithmic thinking, as commonly accepted in the broader CT literature.

Although problem solving is often considered a higher-order thinking skill closely related to computational thinking, it was measured directly as an independent component in some studies (Cutumisu et al., 2022; Hijon-Neira et al., 2024; Huda & Rohaeti, 2024). Since CT was mainly conceptualized in this review as a problem-solving process, problem solving was not coded as a separate component. Differences were also found across studies regarding testing, evaluation, and debugging (Atmatzidou & Demetriadis, 2016; Kong & Wang, 2023). Testing and evaluation were brought together under the broader component of evaluation, since both involve monitoring, verification, and judgment. Debugging, however, was kept as an independent component.

Hardware and software awareness, operators, and user interaction—components that have been addressed in only a few studies—were interpreted as context-sensitive assessment elements rather than as part of the core CT structure, since they were mostly specific to robotics, physical computing, or game-based learning environments.

4.2.3. Indicators Used and Associated Cognitive Processes

In the 47 studies included in the review, the indicators used in computational thinking assessment tools were examined together with the cognitive processes associated with these indicators. As a result of the analyses, 355 indicators were obtained from the assessment tools used in the examined studies. These indicators were initially recorded in their raw form. Then, considering content similarity, functional proximity, and conceptual consistency, they were brought together under thematic groups, and finally 14 indicator clusters were formed.

The 355 indicators were retained in their original form and assigned to 14 thematic clusters according to content similarity, functional proximity, and conceptual consistency. For example, indicators such as sequential task execution, describing stepwise algorithms, sequencing items, tracing instructions, sequence interpretation, step ordering, ordered steps, correct sequencing, and matching sequences were assigned to the "Sequencing and Control Flow" cluster because they referred to the organization, execution, or interpretation of ordered procedural steps. Thus, the clustering procedure did not reduce or standardize the original indicator set; it was used to organize the full set of indicators under conceptually related thematic headings. The distribution of the indicators among these clusters is presented in Table 2.

Table 2. *Clusters of Indicators Used*

Cluster	Frequency (f)
Algorithmic Structures and Steps	34
Sequencing and Control Flow	28
Loops and Repetition Patterns	26
Conditionals and Logical Decisions	20
Debugging and Error Detection	30
Decomposition and Task Structuring	29
Pattern Recognition and Structure Identification	41
Abstraction and Representation	23
Data Handling, Interpretation and Transformation	28
Spatial and Visual Reasoning	23
Problem Solving and Strategy Generation	15
Evaluation, Testing and Improvement	23
Creativity, Design and Construction	19
Automation, System Behavior and Interaction	16

The cognitive orientation of these indicators was also examined. For this purpose, cognitive process statements and process-related explanations in the methods sections, task descriptions, and assessment criteria of the studies were reviewed. In this process, 324 raw cognitive process statements related to the indicators were collected. These statements were compared in terms of similarity, synonymy, and conceptual proximity and were standardized into 73 representative cognitive processes, which were then grouped under 10 thematic cognitive process clusters through semantic clustering. Some cognitive process expressions were reported explicitly in the source studies, whereas others were identified through the interpretive analysis described in Section 3.8. The cluster names and distributions are presented in Table 3.

Table 3. *Clusters of Associated Cognitive Processes*

Cluster	Frequency (f)
Abstraction and Generalization Processes	7
Analytical and Inferential Reasoning	9
Logical and Deductive Reasoning	8
Algorithmic and Procedural Reasoning	9
Control-Flow and Structural Logic	7
Pattern Recognition and Structure Detection	8
Decomposition and Modular Structuring	7
Debugging and Error Processes	7
Modeling and System Reasoning	6
Creative, Collaborative and Reflective Thinking	5

To illustrate the relationship between indicators and cognitive processes, indicators concerning the decomposition of complex tasks and the identification of sub-units were associated with decomposition-related cognitive processes. Similarly, indicators involving the evaluation of logical expressions, the prediction of algorithm behavior, and the interpretation of conditional structures were associated with logical, algorithmic, and control-flow reasoning processes.

These examples illustrate how observable assessment indicators were interpreted in relation to the cognitive operations represented in the original study contexts.

When the indicator clusters and cognitive process themes were considered together, it was seen that the assessment tools mainly focused on core aspects of computational thinking, such as algorithmic thinking, decomposition, logical reasoning, pattern recognition, abstraction, data representation and interpretation, error analysis, and modeling. The distribution of the cognitive process clusters also showed parallelism with the thematic structure of the indicators. This alignment showed conceptual consistency between the observable behavioral indicators measured through assessment tools and the cognitive processes associated with them. In this context, it can be said that the assessment tools reported in the literature address the core components of computational thinking at both behavioral and cognitive levels.

4.2.4. Adaptable Indicators to Mathematics Education and Corresponding Task Designs

Math-adaptable indicators were developed from those extracted across the 47 studies included in the systematic literature review, to support their transferability to mathematics education and to identify potential application contexts. This analysis was conducted after the indicators and associated cognitive processes had been identified from the full set of educational studies. It therefore constituted a subsequent interpretive stage of the review rather than a restriction applied during the database search.

The adaptability decision was based not only on the wording of each indicator, but also on the assessment logic of the source study. The assessment tool, item or task structure, rubric criteria, performance descriptions, component definitions, and associated cognitive processes were examined together. First, the observable behavior expected from the student was identified. It was then considered whether this behavior and the underlying thinking process could be elicited and observed through a mathematical task. Indicators were regarded as adaptable when the same functional behavior could be represented through mathematical problem solving, sequencing, decomposition, pattern recognition, representation, data interpretation, spatial reasoning, logical decision-making, or evaluation of solution strategies. Indicators whose meaning depended exclusively on a specific programming syntax, software feature, hardware component, or system interaction were not considered directly adaptable unless the underlying behavior could be meaningfully expressed in a mathematical context.

A total of 270 indicators were adapted from the assessment tools used across the reviewed studies. All indicators were initially listed in their raw form; after removing redundant expressions, orthographic variations, and semantically identical items, 249 non-redundant adaptable indicators were obtained. These indicators were then distributed across the same 14 thematic clusters used in the indicator analysis, based on criteria of content similarity, functional proximity, and conceptual coherence. The distribution of adaptable indicators across these clusters is presented in Table 4.

Table 4. *Clusters of Adaptable Indicators*

Cluster	Frequency (f)
Algorithmic Structures and Steps	31
Sequencing and Control Flow	12
Loops and Repetition Patterns	19
Conditionals and Logical Decisions	17
Debugging and Error Detection	9
Decomposition and Task Structuring	31
Pattern Recognition and Structure Identification	24
Abstraction and Representation	16
Data Handling, Interpretation and Transformation	22
Spatial and Visual Reasoning	10
Problem Solving and Strategy Generation	39
Evaluation, Testing and Improvement	15
Creativity, Design and Construction	2
Automation, System Behavior and Interaction	2

An examination of cluster frequencies shows that indicators adaptable to mathematics education are particularly concentrated in the clusters of problem solving and strategy development (39), algorithmic structures and steps (31), and decomposition and task structuring (31). This suggests that computational thinking within mathematical tasks is most observable through multi-step problem solving, planning the solution process, and breaking down problem situations into manageable subcomponents.

Similarly, the relatively high frequencies of the pattern recognition and structure identification (24) and data processing, interpretation, and representation (22) clusters indicate that recognizing mathematical relationships, identifying patterns, and interpreting quantitative information through multiple representations play an important role in adaptable indicators. In contrast, the limited representation of indicators associated with creativity, design, and production (2), as well as automation, system behavior, and interaction (2), supports the idea that such indicators tend to emerge only in more context-specific mathematical tasks. Overall, these distributions suggest that computational thinking in mathematics education is predominantly observable through problem-solving, structural reasoning, and representation-based task features.

When the mathematical contexts associated with these adaptable indicators are examined, it is observed that they are mostly clustered around task types requiring students to structure solution processes step by step, decompose problem situations into subcomponents, and justify solution paths. In particular, functional tasks based on input-output relationships, pattern and sequence problems, tasks requiring the interpretation of representations such as tables and graphs, and problem contexts involving conditional decision-making were identified as suitable contexts for the emergence of computational thinking indicators. In addition, tasks involving geometric transformations, spatial orientation, and positioning were found to support visual-spatial reasoning and structural awareness.

These contexts make computational thinking indicators more visible in mathematics education when students are expected not only to reach a correct result, but also to structure their solution process, justify intermediate steps, and consider alternative solution paths. Overall, the findings suggest that mathematics-based CT assessment tasks may involve multi-step problem solving, representational transitions, structural reasoning, and the evaluation of solution strategies.

5. RESULTS AND DISCUSSION

In this section, the findings obtained from the study are discussed in relation to assessment tools, assessed components, the alignment between indicators and cognitive processes, and the assessment of CT in the context of mathematics education. The scope and limitations of the results are also evaluated, and a concluding synthesis is presented in connection with the existing literature.

5.1. Trends and Approaches in Assessment Tools

The results of this study show that tests and programming-based assessment tools play a dominant role in the assessment of CT. This trend supports previous systematic reviews indicating that CT integration has mostly been conducted in science and mathematics through constructivist approaches and block-based programming environments (Tosik-Gün & Güyer, 2019; Yeni et al., 2024). Similarly, it has frequently been emphasized in the literature that assessment studies have primarily focused on the K-12 level, while process-based, non-programming, and interdisciplinary assessment approaches have remained limited (Kalelioğlu et al., 2016; Tang et al., 2020). While supporting these general trends, this study examines assessment tools at the indicator and cognitive process levels and reveals structural gaps in current assessment approaches.

5.2. Distribution and Theoretical Coherence in Assessed Components

The results of this study reveal that certain components are addressed more prominently than others in the assessment of CT. In particular, the clear prominence of algorithmic thinking, decomposition, abstraction, and pattern recognition confirms that the accepted core structure of CT in the literature has largely been preserved. This distribution coincides with previous studies emphasizing that CT is generally approached as a problem-solving-based form of thinking (Kalelioğlu et al., 2016; Tang et al., 2020; Tosik-Gün & Güyer, 2019).

While earlier research has shown that these components are mostly assessed through programming and structured problem-solving activities, the current study confirms this trend by addressing the conceptual diversity observed in the literature. Recoding overlapping processes expressed under different names within a common component framework has revealed a more comprehensive and comparable structure for CT assessment. In this respect, the study has shed light not only on the frequency of CT components but also on their theoretical organization.

5.3. Alignment Between Indicators and Cognitive Processes

The results of this study show that the indicators used in CT assessment tools are conceptually aligned with the cognitive processes associated with these indicators. The concentration of indicators around core CT components, such as algorithmic thinking, decomposition, logical reasoning, pattern recognition, abstraction, data representation and interpretation, debugging, and modeling, suggests that the behaviors measured by these tools are connected to theoretically meaningful cognitive processes. In this way, the implicit relationship between CT components and assessment practices noted in previous studies has been made more visible (Lu et al., 2022; Tosik-Gün & Güyer, 2019).

CT assessment tools in the literature have mostly been addressed through product-based or performance-based indicators, while limited attention has been given to the cognitive processes represented by these indicators. In order to overcome this limitation, the indicators were first classified under thematic clusters, and the cognitive processes associated with these indicators were then examined. In this way, it was possible to clarify not only what CT assessment tools measure, but also how the indicators used in these tools are related to underlying cognitive operations.

The alignment identified between the indicator clusters and the cognitive process clusters made the implicit theoretical structure in the CT assessment literature more explicit and traceable. It can be stated that assessment tools developed in different contexts can be interpreted through shared cognitive foundations. This relationship can also be used in the analysis of existing assessment practices and in the development of future tasks and tools with stronger theoretical consistency.

From a theoretical perspective, these findings suggest that CT assessment should not be understood solely as the measurement of performance within programming environments. The dominance of programming-based tools may narrow the construct by making some components and cognitive processes more visible while leaving others underrepresented. By linking observable indicators with the cognitive processes they represent, the present study offers a more explicit account of the internal structure of CT assessment. This indicator–cognitive process relationship may provide a theoretical basis for evaluating whether existing instruments adequately represent the multidimensional nature of CT and for designing assessment approaches with stronger conceptual coherence.

5.4. Assessing Computational Thinking in the Context of Mathematics Education

According to the results of this study, CT becomes visible in mathematics education especially through problem solving, structural reasoning, and representation-based tasks. Adaptable indicators were concentrated in task characteristics such as multi-step problem solving, decomposing problem situations into subcomponents, and making transitions between representations. This finding shows that mathematical activities can provide a suitable context for CT assessment. It is also consistent with studies emphasizing that CT can be addressed through cognitive processes shared with mathematical thinking and mathematical problem solving (Kalelioğlu et al., 2016; Kallia et al., 2021).

Pattern and sequence problems, functional tasks based on input-output relationships, situations requiring the interpretation of representations such as tables and graphs, and problem contexts involving conditional decision-making were identified as contexts in which CT-related indicators may emerge more clearly. On the other hand, more technical or context-specific indicators, such as those related to creativity, design, automation, system behavior, and interaction, were represented to a limited extent. It can therefore be stated that these indicators may become meaningful only within specific mathematical task designs and learning environments.

In mathematics education, these contexts can be considered together with the notion of mathematical tasks and their cognitive demand. Tasks involving pattern generalization, input-output reasoning, transitions between representations, and conditional decisions require students not only to produce an answer, but also to structure, justify, and monitor their thinking processes. In this respect, mathematical tasks provide a meaningful basis for assessing CT. They make it possible to observe not only whether a solution is reached, but also how the underlying cognitive processes are organized throughout the solution path (Doyle, 1983, 1988; Stein et al., 1996; Stein & Lane, 1996; Stein & Smith, 1998). This issue is also important for CT assessment in non-digital contexts (Kalelioğlu et al., 2016; Tosik-Gün & Güyer, 2019). The present review does not involve the development or empirical testing of mathematics tasks; rather, it identifies task characteristics and mathematical contexts in which CT-related indicators may become observable.

Within this framework, the findings suggest that CT assessment in mathematics education should not remain limited to approaches focusing solely on correct answers. Process-oriented tasks that allow for different solution paths may make both students' mathematical thinking processes and CT-related indicators more visible. Therefore, approaches prioritizing the traceability of solution processes in mathematical task design may offer more meaningful insights for CT assessment.

5.5. Limitations of the Study

Some limitations of this study should be acknowledged. First, the review was restricted to studies indexed in WoS and Scopus and published in English, which may have led to the exclusion of relevant research published in other languages or indexed elsewhere. The use of the terms "Education" and "Educational" in the search string may also have limited the retrieval of some studies indexed under alternative educational descriptors. In addition, because the analysis was based on the assessment tools, indicators, and cognitive-process explanations reported in the reviewed studies, it should be noted that the findings remain limited to the information available in published reports. Finally, it should also be noted that the thematic clustering of indicators and cognitive processes involved an interpretive analytical step. Despite these limitations, the consistency and traceability of the analysis were strengthened through systematic coding procedures and repeated checks.

5.6. Conclusion

In this study, empirical studies on the assessment of computational thinking in educational contexts were reviewed systematically. Assessment tools, assessed components, indicators, and the cognitive processes related to these indicators were considered together. The review was carried out in two stages, and the general structure of the CT assessment literature was described on this basis.

It was seen that test-based and programming-based tools still had an important place in CT assessment. Algorithmic thinking, decomposition, abstraction, and pattern recognition were the components addressed most frequently. However, in most of the reviewed studies, the relationship between assessment indicators and cognitive processes was not clearly explained. Although there was an implicit consistency among many studies, this relationship generally remained in the background. For this reason, indicators were considered together with related cognitive processes, and the link between observable assessment indicators and cognitive operations was made more visible.

The findings also indicated that process-oriented mathematical tasks involving multi-step problem solving, multiple representations, structural reasoning, and conditional decision-making may provide suitable contexts for making CT-related indicators and cognitive processes more visible. In this sense, CT assessment does not have to remain limited to programming-based activities. Mathematical tasks may also serve as non-digital assessment contexts when they make students' solution processes and underlying cognitive operations observable.

In general, assessment tools, indicators, and cognitive processes related to CT were brought together within a systematic synthesis. In this way, an attempt was made to make the theoretical basis of current assessment practices more visible. The findings may also contribute to the reanalysis of existing instruments and to the development of assessment approaches that are more theoretically grounded and sensitive to context. In particular, the indicator-cognitive process relationship identified in the review may offer an analytical basis for examining the conceptual coverage and theoretical coherence of CT assessment instruments.

6. RECOMMENDATIONS

Based on the findings of this review, some recommendations can be made for future studies on CT assessment. First, assessment approaches should not be limited only to the products produced by students. The cognitive processes underlying these products should also be taken into account, and the relationship between indicators and cognitive operations should be explained more clearly. Second, since programming-based tools are still dominant in the literature, CT assessment needs to be considered in different disciplinary contexts. Mathematics education may be one of these contexts, especially through tasks involving multi-step problem solving, transitions between representations, pattern recognition, structural reasoning, and the traceability of solution processes. Future studies may extend these findings by developing concrete mathematics tasks based on the identified indicators and by examining their validity, reliability, and practical use through empirical applications in different educational contexts.

It was also seen that validity and reliability evidence was reported in a limited way in many studies. For this reason, clearer reporting of validity and reliability procedures is needed in future assessment studies. Another point is that the current literature is concentrated mainly at the K-12 level. More studies may therefore be conducted in high school, higher education, and teacher education contexts. Finally, the indicator-cognitive process matching used in this study may be useful for re-examining existing assessment instruments and for developing new assessment tools with a clearer theoretical basis.

7. ABOUT THE AUTHORS

Faysal ÇEKER: He is a PhD candidate in mathematics education at Hacettepe University. His research focuses on computational thinking in mathematics education, the integration of computational thinking into mathematics teaching, and teacher education. He works as a lecturer in Mardin Artuklu University.

Bahadır YILDIZ: He is a faculty member of Hacettepe University Faculty of Education, Department of Mathematics Teaching. He graduated from Hacettepe University, Elementary Education Undergraduate Program and completed her master's and doctoral education in the Department of Computer Education and Instructional Technology at the same university. He is working on computational thinking, design thinking, interdisciplinary approach and STEAM, machine learning and artificial intelligence in mathematics education, with the main focus on the integration of information and communication technologies into mathematics education. He produces national and international publications and projects on related topics.

8. References

Åkerfeldt, A., Kjällander, S., & Petersen, P. (2024). A research review of computational thinking and programming in education. *Technology, Pedagogy and Education*, *33*(3), 375-390. https://doi.org/10.1080/1475939X.2024.2316087

Altanis, I., & Retalis, S. (2019). A multifaceted students' performance assessment framework for motion-based game-making projects with Scratch. *Educational Media International*, *56*(3), 201-217. https://doi.org/10.1080/09523987.2019.1669876

Arastoopour Irgens, G., Dabholkar, S., Bain, C., Woods, P., Hall, K., Swanson, H., ... & Wilensky, U. (2020). Modeling and measuring high school students' computational thinking practices in science. *Journal of Science Education and Technology*, *29*(1), 137-161. https://doi.org/10.1007/s10956-020-09811-1

Aristizábal Zapata, J. H., Gutiérrez Posada, J. E., & Diago, P. D. (2024). Design and validation of a computational thinking test for children in the first grades of elementary education. *Multimodal Technologies and Interaction*, *8*(5), 39. https://doi.org/10.3390/mti8050039

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems*, *75*, 661-670. https://doi.org/10.1016/j.robot.2015.10.008

Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking*. Paper presented at the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Cutumisu, M., Adams, C., Glanfield, F., Yuen, C., & Lu, C. (2022). Using structural equation modeling to examine the relationship between preservice teachers' computational thinking attitudes and skills. *IEEE Transactions on Education*, *65*(2), 177-183. https://doi.org/10.1109/TE.2021.3105938

de Araujo, A. L. S. O., Andrade, W. L., & Guerrero, D. D. S. (2016). *A systematic mapping study on assessing computational thinking abilities.* In *2016* IEEE frontiers in education conference (FIE) (pp. 1-9). https://doi.org/10.1109/FIE.2016.7757678

Dong, Y., Catete, V., Jocius, R., Lytle, N., Barnes, T., Albert, J., … & Andrews, A. (2019). *PRADA: A practical model for integrating computational thinking in K–12 education*. In Proceedings of the 50th ACM Technical Symposium on Computer Science Education (pp. 906–912). https://doi.org/10.1145/3287324.3287431

Doyle, W. (1983). Academic work. *Review of Educational Research, 53*, 159-199. https://doi.org/10.3102/00346543053002159

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist, 23(2),* 167–180. https://doi.org/10.1207/s15326985ep2302_6

El-Hamamsy, L., Zapata-Cáceres, M., Marcelino, P., Bruno, B., Dehler Zufferey, J., Martín-Barroso, E., & Román-González, M. (2022). Comparing the psychometric properties of two primary school computational thinking (CT) assessments for grades 3 and 4: The beginners' CT test (BCTt) and the competent CT test (cCTt). *Frontiers in Psychology*, *13*, 1082659. https://doi.org/10.3389/fpsyg.2022.1082659

Eloy, A., Achutti, C. F., Fernandez, C., & de Deus Lopes, R. (2022). A data-driven approach to assess computational thinking concepts based on learners' artifacts. *Informatics in Education*, *21*(1), 33-54. https://doi.org/10.15388/infedu.2022.02

Ezeamuzie, N. O., & Leung, J. S. (2022). Computational thinking through an empirical lens: A systematic review of literature. *Journal of Educational Computing Research*, *60*(2), 481-511. https://doi.org/10.1177/07356331211033158

Fagerlund, J., Häkkinen, P., Vesisenaho, M., & Viiri, J. (2020). Assessing 4th grade students' computational thinking through scratch programming projects. *Informatics in Education*, *19*(4), 611-640. https://doi.org/10.15388/infedu.2020.27

Fofang, J. S., Weintrop, D., Walton, M., Elby, A., & Walkoe, J. (2020). Mutually supportive mathematics and computational thinking in a fourth-grade classroom. In M. Gresalfi & I. S. Horn (Eds.), *The interdisciplinarity of the learning sciences:* Proceedings of the 14th International Conference of the Learning Sciences (ICLS 2020) (Vol. 3, pp. 1389–1396). https://doi.org/10.22318/icls2020.1389

Gane, B. D., Israel, M., Elagha, N., Yan, W., Luo, F., & Pellegrino, J. W. (2021). Design and validation of learning trajectory-based assessments for computational thinking in upper elementary grades. *Computer Science Education*, *31*(2), 141-168. https://doi.org/10.1080/08993408.2021.1874221

Google. (2020). *Introduction to computational thinking*. https://sites.google.com/isabc.ca/computationalthinking/home

Grover, S. (2018). *The 5th "C" of 21st century skills? Try computational thinking (not coding).* EdSurge. https://www.edsurge.com/news/2018-02-25-the-5th-c-of-21st-century-skills-try-computational-thinking-not-coding

Gunhan, B. C., Yılmaz, S. & Turgut, M. (2009). An investigation of knowledge level about spatial ability. *Education Sciences, 4*(2), 317-326.

Hijón-Neira, R., Pizarro, C., French, J., Palacios-Alonso, D., & Çoban, E. (2024). Computational thinking measurement of CS university students. *Applied Sciences, 14(12),* 5261. https://doi.org/10.3390/app14125261

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. https://doi.org/10.1016/j.compedu.2018.07.004

Huda, N., & Rohaeti, E. (2024). Computational thinking skill level of senior high school students majoring in natural science. *International Journal of Learning, Teaching and Educational Research*, *23*(1), 339-359. *https://doi.org/10.26803/ijlter.23.1.17*

Kaleem, M., Hassan, M. A., & Khurshid, S. K. (2024). A machine learning-based adaptive feedback system to enhance programming skill using computational thinking. *IEEE Access, 12,* 59431–59440. https://doi.org/10.1109/ACCESS.2024.3391873

Kalelioğlu, F., Gülbahar, Y., & Kukul, V. (2016). A framework for computational thinking based on a systematic research review. *Baltic Journal of Modern Computing, 4(3),* 583–596. https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/4_3_15_Kalelioglu.pdf

Kallia, M., van Borkulo, S. P., Drijvers, P., Barendsen, E., & Tolboom, J. (2021). Characterising computational thinking in mathematics education: A literature-informed Delphi study. *Research in Mathematics Education*, *23*(2), 159–187. https://doi.org/10.1080/14794802.2020.1852104

Kitchenham, B. (2004). *Procedures for performing systematic reviews.* Keele University.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering.* Keele University, Software Engineering Group, School of Computer Science and Mathematics.

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. https://doi.org/10.1016/j.chb.2023.107749

Li, Y., Xu, S., & Liu, J. (2021). Development and validation of computational thinking assessment of Chinese elementary school students. *Journal of Pacific Rim Psychology*, *15*. https://doi.org/10.1177/18344909211010240

Liu, Z., Gearty, Z., Richard, E., Orrill, C. H., Kayumova, S., & Balasubramanian, R. (2024). Bringing computational thinking into classrooms: A systematic review on supporting teachers in integrating computational thinking into K-12 classrooms. *International Journal of STEM Education, 11(1), 51.* https://doi.org/10.1186/s40594-024-00510-6

Lockwood, J., & Mooney, A. (2018). *Developing a computational thinking test using Bebras problems*. https://mural.maynoothuniversity.ie/id/eprint/10316/1/AM-Developing-2017.pdf

Lu, C., Macdonald, R., Odell, B., Kokhan, V., Demmans Epp, C., & Cutumisu, M. (2022). A scoping review of computational thinking assessments in higher education. *Journal of Computing in Higher Education*, *34*(2), 416-461. https://doi.org/10.1007/s12528-021-09305-y

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior, 41,* 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Manfra, M. M., Hammond, T. C., & Coven, R. M. (2022). Assessing computational thinking in the social studies. *Theory & Research in Social Education*, *50*(2), 255-296. https://doi.org/10.1080/00933104.2021.2003276

National Research Council. (2011). *Report of a workshop on the pedagogical aspects of computational thinking.* National Academies Press.

Ogegbo, A. A., & Ramnarain, U. (2022). A systematic review of computational thinking in science classrooms. *Studies in Science Education*, *58*(2), 203-230. https://doi.org/10.1080/03057267.2021.1963580

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* n71. https://doi.org/10.1136/bmj.n71

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

Pirzado, F. A., Ahmed, A., Hussain, S., Ibarra-Vázquez, G., & Terashima-Marin, H. (2025). Assessing computational thinking in engineering and computer science students: A multi-method approach. *Education Sciences, 15(3),* 344. https://doi.org/10.3390/educsci15030344

Posamentier, A. S., & Krulik, S. (2016). *Problem solving in mathematics, grades 3–6: Powerful strategies to deepen understanding.* Corwin Press.

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. https://doi.org/10.1016/j.chb.2016.08.047

Romero, M., Lepage, A., & Lille, B. (2017). Computational thinking development through creative programming in higher education. *International Journal of Educational Technology in Higher Education*, *14*(1), 42. https://doi.org/10.1186/s41239-017-0080-z

Semiawan, T. (2019). *User interface design analysis pertaining to computational thinking framework.* In Proceedings of the 8th International Conference on Informatics, Environment, Energy and Applications (pp. 238–242). https://doi.org/10.1145/3323716.3323741

Shen, L., Mirakhur, Z., & LaCour, S. (2025). Investigating the psychometric features of a locally designed computational thinking assessment for elementary students. *Computer Science Education*, *35*(2), 414-433. https://doi.org/10.1080/08993408.2024.2344400

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488. https://doi.org/10.3102/00028312033002455

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2(1),* 50–80. https://doi.org/10.1080/1380361960020103

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School, 3*(4), 268-275. https://doi.org/10.5951/MTMS.3.4.0268

Tariq, R., Aponte Babines, B. M., Ramirez, J., Alvarez-Icaza, I., & Naseer, F. (2025). Computational thinking in STEM education: Current state-of-the-art and future research directions. *Frontiers in Computer Science, 6,* 1480404. https://doi.org/10.3389/fcomp.2024.1480404

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, 103798. https://doi.org/10.1016/j.compedu.2019.103798

Tosik-Gün, E., & Güyer, T. (2019). Bilgi işlemsel düşünme becerisinin değerlendirilmesine ilişkin sistematik alanyazın taraması [A systematic literature review on assessing computational thinking skills]. *Ahmet Keleşoğlu Education Faculty Journal, 1*(2), 99–120. https://doi.org/10.38151/akef.597505

Valenzuela, J. (2020). *How to develop computational thinkers*. ISTE. https://www.iste.org/explore/how-develop-computational-thinkers

Wang, C., Shen, J., & Chao, J. (2022). Integrating computational thinking in STEM education: A literature review. *International Journal of Science and Mathematics Education*, *20*(8), 1949-1972. https://doi.org/10.1007/s10763-021-10227-5

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. https://doi.org/10.1007/s10956-015-9581-5

Weng, X., Ye, H., Dai, Y., & Ng, O. L. (2024). Integrating artificial intelligence and computational thinking in educational contexts: A systematic review of instructional design and student learning outcomes. *Journal of Educational Computing Research*, *62*(6), 1420-1450. https://doi.org/10.1177/07356331241248686

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49(3),* 33-35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2011). Research notebook: Computational thinking-What and why? *The Link Magazine, 6,* 20–23. https://people.cs.vt.edu/~kafura/CS6604/Papers/CT-What-And-Why.pdf

Wing, J. M. (2014). *Computational thinking benefits society*. 40th Anniversary *Blog of Social Issues in Computing.*

Xiang, S., Li, J. W., & Yang, W. (2025). Developing a robot-based computational thinking assessment for young children. *Education and Information Technologies*, 1-23. https://doi.org/10.1007/s10639-025-13377-z

Yeni, S. (2020). Bilgi işlemsel düşünme becerisi nasıl değerlendirilir? [How is computational thinking skill assessed?]. In Y. Gülbahar (Ed.), *Bilgi işlemsel düşünmeden programlamaya [From computational thinking to programming]* (pp. 360–394). Pegem Academy.

Yeni, S., Grgurina, N., Saeli, M., Hermans, F., Tolboom, J., & Barendsen, E. (2024). Interdisciplinary integration of computational thinking in K-12 education: A systematic review. *Informatics in Education, 23(1),* 223–278. https://doi.org/10.15388/infedu.2024.08

Yıldız, M. (2021). *Bilgi işlemsel düşünme becerisinin süreç temelli ölçülmesi ve değerlendirilmesi [Process-based measurement and assessment of computational thinking skill]*. [Unpublished Doctoral Dissertation]. Graduate School of Education, Trabzon University.

Yin, S. X., Hoe-Lian Goh, D., & Quek, C. L. (2024). Collaborative learning in K-12 computational thinking education: A systematic review. *Journal of Educational Computing Research*, *62*(6), 1220-1254. https://doi.org/10.1177/07356331241249956

Yun, M., & Crippen, K. J. (2025). Computational thinking integration into pre-service science teacher education: A systematic review. *Journal of Science Teacher Education*, *36*(2), 225-254. https://doi.org/10.1080/1046560X.2024.2390758

Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*, *141*, 103607. https://doi.org/10.1016/j.compedu.2019.103607

Zhang, S., & Wong, G. K. (2023). Development and validation of a computational thinking test for lower primary school students. *Educational Technology Research and Development, 71(4),* 1595-1630. https://doi.org/10.1007/s11423-023-10231-2

Zhong, B., Wang, Q., Chen, J., & Li, Y. (2016). An exploration of three-dimensional integrated assessment for computational thinking. *Journal of Educational Computing Research*, *53*(4), 562-590. https://doi.org/10.1177/0735633115608444

Appendix

Table S1. *Examined Studies in Scoping Review Phase*

Authors and Year	Study Focus	Journal/Source	Country	Indexing
Yeni et al. (2024)	Integration of CT Across Disciplines	Informatics in Education	Netherlands	ESCI
Tosik-Gün & Guyer (2019)	Assessment of CT Skills	Ahmet Kelesoglu Journal of Faculty of Education	Turkey	TR Index
Kalelioglu et al. (2016)	Developing a Framework for CT	Baltic Journal of Modern Computing	Turkey	ESCI
Tang et al. (2020)	Assessment of CT Skills	Computers & Education	USA	SCIE
Ezeamuzie & Leung (2022)	Nature of Computational Thinking	Journal of Educational Computing Research	China	SSCI
Åkerfeldt et al. (2024)	CT and Programming in Education	Technology, Pedagogy and Education	Sweden	SSCI
Ogegbo & Ramnarain (2022)	CT in Science Education	Studies in Science Education	South Africa	SSCI
Zhang & Nouri (2019)	Teaching Computational Thinking	Computers & Education	Sweden	SCIE
Lu et al. (2022)	Assessment of CT	Journal of Computing in Higher Education	Canada	SSCI
Wang et al. (2022)	Integration of CT into STEM Education	International Journal of Science and Mathematics Education	Taiwan	SSCI
Hsu et al. (2018)	Learning and Teaching CT	Computers & Education	Taiwan	SCIE
Lockwood & Mooney (2018)	Positioning CT in Middle School	International Journal of Computer Science Education in Schools	Ireland	ERIC
de Araujo et al. (2016)	Assessment of CT Skills	IEEE Frontiers in Education Conference (FIE)	Brazil	IEEE Xplore
Kallia et al. (2021)	Nature of CT in Maths Education	Research in Mathematics Education	Netherlands	ESCI
Lye & Koh (2014)	CT Education through Programming	Computers in Human Behaviour	Singapore	SSCI
Tariq et al. (2025)	Current State of CT in STEM	Frontiers in Computer Science	Mexico	ESCI
Yun & Crippen (2025)	Integration of CT into Science Teacher Education	Journal of Science Teacher Education	USA	ESCI
Yin et al. (2024)	Collaborative Learning in CT	Journal of Educational Computing Research	Singapore	SSCI
Liu et al. (2024)	Integration of CT in Classrooms	International Journal of STEM Education	USA	SSCI
Weng et al. (2024)	AI and CT in Educational Contexts	Journal of Educational Computing Research	China	SSCI

Table S2. *Examined Studies in Main Systematic Literature Review Phase*

Writers and Year	Title	Publication	Country	Discipline
Altanis & Retalis (2019)	A multifaceted students' performance assessment framework for motion-based game-making projects with Scratch	Educational Media International	Greece	Digital Systems
Arastoopour-Irgens et al. (2020)	Modelling and measuring high school students' computational thinking practices in science	Journal of Science Education and Technology	USA	Science Education
Aristizábal Zapata et al. (2024)	Design and Validation of a Computational Thinking Test for Children in the First Grades of Elementary Education	Multimodal Technologies and Interaction	Colombi a	Mathematics Education
Atmatzidou & Demetriadis (2016)	Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences	Robotics and Autonomous Systems	Greece	Computer Science Education
Bhatt et al. (2024)	A Method for Developing Process-Based Assessments for Computational Thinking Tasks	Journal of Learning Analytics	Belgium	Educational Sciences
Bubica & Boljat (2022)	Assessment of Computational Thinking, A Croatian Evidence-Centred Design Model	Informatics in Education	Croatia	Computer Science and Mathematics
Csernoch et al. (2015)	Testing Algorithmic and Application Skills (TAaAS) of First Year Students	Journal of Applied Computing and Informatics	Hungary	Computer Science Education
Cutumisu et al. (2022)	Using structural equation modelling to examine the relationship between preservice teachers' computational thinking attitudes and skills	IEEE Transactions on Education	Canada	Educational Sciences
Mendoza-Diaz et al. (2023)	Development and Validation of the Engineering Computational Thinking Diagnostic for Undergraduate Students	IEEE Access	USA	Engineering Education
El-Hamamsy et al. (2022)	Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt)	Frontiers in Psychology	Switzerla nd	Computer Science Education
El-Hamamsy et al. (2025)	The competent Computational Thinking test (cCTt): A valid, reliable and gender-fair test for longitudinal CT studies in grades 3-6	Technology, Knowledge and Learning	Switzerla nd	Computer Science Education
Eloy et al. (2022)	A data-driven approach to assess computational thinking concepts based on learners' artifacts	Informatics in Education	Brazil	Computer Science Education
Fagerlund et al. (2020)	Assessing 4th grade students' computational thinking through Scratch programming projects	Informatics in Education	Finland	Computer Science Education
Gane et al. (2021)	Design and validation of learning trajectory-based assessments for	Computer Science Education	USA	Educational Technology

	computational thinking in upper elementary grades			
Geng et al. (2025)	Validating a measure of computational thinking skills in Chinese kindergartners	Education and Information Technologies	China	Educational Sciences
Hijón-Neira et al. (2024)	Computational Thinking Measurement of CS University Students	Applied Sciences	Spain and Ireland	Computer Science Education
Hogenboom et al. (2022)	Computerized adaptive assessment of understanding of programming concepts in primary school children	Computer Science Education	Netherla nds	Computer Science Education
Hsu et al. (2023)	A validity and reliability study of the formative model for the indicators of STEAM education creations	Education and Information Technologies	Taiwan	Educational Technology
Huda & Rohaeti (2024)	Computational thinking skill level of senior high school students majoring in natural science	International Journal of Learning, Teaching and Educational Research	Indonesi a	Educational Sciences
Jiménez et al. (2024)	Computational Concepts and their Assessment in Preschool Students: An Empirical Study	Journal of Science Education and Technology	Spain	Educational Sciences
Kaleem et al. (2024)	A Machine Learning-Based Adaptive Feedback System to Enhance Programming Skill Using Computational Thinking	IEEE Access	Pakistan	Computer Science Education
Kanaki & Kalogiannakis (2022)	Assessing Algorithmic Thinking Skills in Relation to Age in Early Childhood STEM Education	Education Sciences	Greece	Educational Sciences
Kang et al. (2023)	Developing College Students' Computational Thinking Multidimensional Test Based on Life Story Situations	Education and Information Technologies	China	Educational Technology
Kong & Wang (2023)	Monitoring Cognitive Development through the Assessment of Computational Thinking Practices: A Longitudinal Intervention on Primary School Students	Computers in Human Behaviour	China	Mathematics and Information Technology
Lai (2021)	Beyond Programming: A Computer-Based Assessment of Computational Thinking Competency	ACM Transactions on Computing Education	United Kingdom	Computer Science Education
Lai & Ellefson (2023)	How Multidimensional is Computational Thinking Competency? A Bi-Factor Model of the Computational Thinking Challenge	Journal of Educational Computing Research	United Kingdom	Computer Science Education
Li et al. (2021)	Development and Validation of Computational Thinking Assessment of Chinese Elementary School Students	Journal of Pacific Rim Psychology	China	Educational Sciences
Luo et al. (2022)	Elementary Computational Thinking Instruction and Assessment: A Learning Trajectory Perspective	ACM Transactions on Computing Education	USA	Mathematics Education

Manfra et al. (2022)	Assessing Computational Thinking in the Social Studies	Theory & Research in Social Education	USA	Social Studies Education
Mukasheva & Omirzakova (2021)	Computational Thinking Assessment at Primary School in the Context of Learning Programming	World Journal on Educational Technology: Current Issues	Kazakhstan	Computer Science Education
Musaeus & Musaeus (2024)	Computational Thinking and Modelling: A Quasi-Experimental Study of Learning Transfer	Education Sciences	Denmark	Computer Science Education
Pirzado et al. (2025)	Assessing Computational Thinking in Engineering and Computer Science Students: A Multi-Method Approach	Education Sciences	Mexico	Computer Science Education
Relkin et al. (2020)	TechCheck: Development and validation of an unplugged assessment of computational thinking in early childhood education	Journal of Science Education and Technology	USA	Educational Technology
Román-González et al. (2017)	Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test	Computers in Human Behaviour	Spain	Computer Science Education
Romero et al. (2017)	Computational thinking development through creative programming in higher education	International Journal of Educational Technology in Higher Education	France	Educational Technology
Santos et al. (2024)	Computational Thinking in Primary School Using Educational Robots: Construction and Validation of an Assessment Tool	International Journal of Information and Education Technology	Portugal	Computer Science Education
Shen et al. (2025)	Investigating the Psychometric Features of a Locally Designed Computational Thinking Assessment for Elementary Students	Computer Science Education	USA	Educational Sciences
Su et al. (2024)	Constructing a Computational Thinking Evaluation Framework for Pupils	IEEE Transactions on Education	Taiwan	Computer Science Education
Sung (2022)	Assessing young Korean children's computational thinking: A validation study of two measurements	Education and Information Technologies	South Korea	Educational Sciences
Teng & Chung (2025)	Measuring Children's Computational Thinking and Problem-Solving in a Block-Based Programming Game	Education Sciences	USA	Computer Science Education
Tsai et al. (2022)	Development and Validation of the Computational Thinking Test for Elementary School Students (CTT-ES): Correlate CT Competency with CT Disposition	Journal of Educational Computing Research	Taiwan	Computer Science Education
Tsarava et al. (2022)	A Cognitive Definition of Computational Thinking in Primary Education	Computers & Education	Germany	Educational Sciences

Xiang et al. (2025)	Developing a Robot-Based Computational Thinking Assessment for Young Children	Education and Information Technologies	China	Educational Technology
Zakwandi et al. (2024)	A Two-Tier Computerized Adaptive Test to Measure Student Computational Thinking Skills	Education and Information Technologies	Indonesia	Physics Education
Zhang et al. (2024)	Towards an assessment model of college students' computational thinking with text-based programming	Education and Information Technologies	China	Educational Technology
Zhang & Wong (2023)	Development and validation of a computational thinking test for lower primary school students	Educational Technology Research and Development	China	Educational Technology
Zhong et al. (2016)	An exploration of three-dimensional integrated assessment for computational thinking	Journal of Educational Computing Research	China	Educational Technology

Table S3. *Assessment Tools and Types in Computational Thinking*

Study	Assessment Tool	Tool Type
Altanis & Retalis (2019)	Dr. Scratch; Scrape; Quality Hound; CT Journal; storyboard; GDD; rubric; observation; questionnaire	Automated analysis tools; rubric; qualitative tool; observation; questionnaire
Arastoopour-Irgens et al. (2020)	Pre-posttests; embedded tasks; CT-STEM rubric; NetLogo logs; coding scheme	Test; task-based assessment; rubric; log analysis; content analysis
Aristizábal-Zapata et al. (2024)	Computational Thinking Test for Children (CTTC)	Paper-and-pencil test
Atmatzidou & Demetriadis (2016)	Written tests; oral questioning; project rubric	Performance-based assessment; test; rubric
Bhatt et al. (2024)	CT test; process-based assessment; trace analytics	Test; process-based assessment
Bubica & Boljat (2022)	MCQ and open-ended CT tests; questionnaire	Test; questionnaire
Csernoch et al. (2015)	TAaAS test; Excel and CAAD tasks	Test; performance task
Cutumisu et al. (2022)	CCTt; attitude scale; CT skills test; problem scenarios	Likert scale; test; performance task
Mendoza-Diaz et al. (2023)	Engineering CT Diagnostic (ECTD)	Multiple-choice test
El-Hamamsy et al. (2022)	BCTt; cCTt	Multiple-choice test
El-Hamamsy et al. (2025)	cCTt	Paper-and-pencil multiple-choice test
Eloy et al. (2022)	Scratch projects; automated analysis system	Product-based assessment; automated analysis
Fagerlund et al. (2020)	Scratch projects; evaluation framework	Product-based performance assessment
Gane et al. (2021)	Learning-trajectory CT tasks	Structured performance tasks
Geng et al. (2025)	TechCheck-K	Paper-and-pencil multiple-choice test
Hijón-Neira et al. (2024)	CTScore; CTProg	Digital multiple-choice tests
Hogenboom et al. (2022)	CAPCT	Computerized adaptive test
Hsu et al. (2023)	STEAM Creation Indicator Assessment	Likert-type questionnaire
Huda & Rohaeti (2024)	CTS	Likert-type questionnaire

Jiménez et al. (2024)	BCTt-SF	Paper-and-pencil performance test
Kaleem et al. (2024)	Programming assignments; CT tasks	Performance tasks; rubric
Kanaki & Kalogiannakis (2022)	Researcher-developed paper test	Performance-based paper tasks
Kang et al. (2023)	Scenario-based CT test	Multiple-choice test
Kong & Wang (2023)	CT Practices Assessment Test	Multiple-choice test
Lai (2021)	Computational Thinking Challenge	Digital task-based test
Lai & Ellefson (2023)	Computational Thinking Challenge	Digital performance assessment
Li et al. (2021)	CTA-CES	Paper-and-pencil multiple-choice test
Luo et al. (2022)	Cognitive interview protocol	Interview-based assessment
Manfra et al. (2022)	Student artefacts; observation; journals; interviews	Qualitative tools
Mukasheva & Omirzakova (2021)	Guided questions; Scratch tasks	Mixed test and performance tasks
Musaeus & Musaeus (2024)	Adapted Bebras; NetLogo tasks	Test; performance tasks
Pirzado et al. (2025)	CTT; CTS; open-ended questions	Test; Likert scale; qualitative responses
Relkin et al. (2020)	TechCheck; TACTIC-KIBO	Multiple-choice test; paper-based assessment
Román-González et al. (2017)	CTt; PMA; RP30	Multiple-choice test; cognitive tests
Romero et al. (2017)	Dr. Scratch; expert review	Automated analysis; qualitative assessment
Santos et al. (2024)	Researcher-developed CT tasks	Performance-based tasks
Shen et al. (2025)	CTAES	Multiple-choice paper test
Su et al. (2024)	CTEF	Likert-type scale
Sung (2022)	Bebras Cards; TACTIC-KIBO	Performance tasks; visual assessment
Teng & Chung (2025)	codeSpark; TechCheck	Game-based assessment; test
Tsai et al. (2022)	CTT-ES; CTS	Multiple-choice test; Likert scale
Tsarava et al. (2022)	Adapted CTt; reasoning tests	Multiple-choice and short-answer tests
Xiang et al. (2025)	CTOS	Observation scale
Zakwandi et al. (2024)	Two-tier CAT	Online adaptive test
Zhang et al. (2024)	Three-layer CT model	Code-based automated assessment
Zhang & Wong (2023)	CTtLP	Diagnostic multiple-choice test
Zhong et al. (2016)	Structured tasks; reflection reports	Performance tasks; rubric; self-report

Table S4. *Assessed Components in Computational Thinking*

Component	f	Study
Algorithmic Thinking	45	Altanis and Retalis (2019); Aristizábal-Zapata et al. (2024); Atmatzidou and Demetriadis (2016); Bhatt et al. (2024); Bubica and Boljat (2022); Csernoch et al. (2015); Cutumisu et al. (2022); Mendoza-Diaz et al. (2023); El-Hamamsy et al. (2022); El-Hamamsy et al. (2025); Eloy et al. (2022); Fagerlund et al. (2020); Gane et al. (2021); Geng et al. (2025); Hijón-Neira et al. (2024); Hogenboom et al. (2022); Hsu et al. (2023); Huda and Rohaeti (2024); Jiménez et al. (2024); Kaleem et al. (2024); Kanaki and Kalogiannakis (2022); Kang et al. (2023); Kong and Wang (2023); Lai and Ellefson (2023); Lai (2021); Li et al. (2021); Luo et al. (2022); Mukasheva and Omirzakova (2021); Musaeus and Musaeus (2024); Pirzado et al. (2025); Relkin et al. (2020); Roman-Gonzalez et al. (2017); Romero et al. (2017); Shen et al. (2025); Santos et al. (2024); Su et al. (2024); Sung (2022); Teng and Chung (2025); Tsai et al. (2022); Tsarava et al. (2022); Xiang et al. (2025); Zakwandi et al. (2024); Zhang et al. (2024); Zhang and Wong (2023); Zhong et al. (2016)

Component	f	Study
Decomposition	27	Altanis and Retalis (2019); Aristizábal-Zapata et al. (2024); Atmatzidou and Demetriadis (2016); Bhatt et al. (2024); Bubica and Boljat (2022); Csernoch et al. (2015); El-Hamamsy et al. (2022); Fagerlund et al. (2020); Gane et al. (2021); Geng et al. (2025); Hijón-Neira et al. (2024); Hsu et al. (2023); Kaleem et al. (2024); Kang et al. (2023); Kong and Wang (2023); Lai and Ellefson (2023); Lai (2021); Li et al. (2021); Luo et al. (2022); Musaeus and Musaeus (2024); Pirzado et al. (2025); Santos et al. (2024); Su et al. (2024); Teng and Chung (2025); Tsai et al. (2022); Zakwandi et al. (2024); Zhang et al. (2024)
Abstraction	25	Altanis and Retalis (2019); Aristizábal-Zapata et al. (2024); Bhatt et al. (2024); Bubica and Boljat (2022); Fagerlund et al. (2020); Hijón-Neira et al. (2024); Hsu et al. (2023); Kaleem et al. (2024); Kang et al. (2023); Kong and Wang (2023); Lai and Ellefson (2023); Lai (2021); Li et al. (2021); Mukasheva and Omirzakova (2021); Musaeus and Musaeus (2024); Pirzado et al. (2025); Romero et al. (2017); Shen et al. (2025); Santos et al. (2024); Su et al. (2024); Teng and Chung (2025); Tsai et al. (2022); Xiang et al. (2025); Zakwandi et al. (2024); Zhang et al. (2024)
Pattern Recognition	19	Altanis and Retalis (2019); Aristizábal-Zapata et al. (2024); Bhatt et al. (2024); Csernoch et al. (2015); Cutumisu et al. (2022); Mendoza-Diaz et al. (2023); El-Hamamsy et al. (2022); El-Hamamsy et al. (2025); Hijón-Neira et al. (2024); Hsu et al. (2023); Kaleem et al. (2024); Li et al. (2021); Manfra et al. (2022); Musaeus and Musaeus (2024); Pirzado et al. (2025); Santos et al. (2024); Teng and Chung (2025); Tsai et al. (2022); Zakwandi et al. (2024)
Data Representation	15	Altanis and Retalis (2019); Arastoopour-Irgens et al. (2020); Mendoza-Diaz et al. (2023); Eloy et al. (2022); Fagerlund et al. (2020); Gane et al. (2021); Geng et al. (2025); Manfra et al. (2022); Mukasheva and Omirzakova (2021); Relkin et al. (2020); Roman-Gonzalez et al. (2017); Romero et al. (2017); Shen et al. (2025); Sung (2022); Zhang et al. (2024)
Generalization	13	Arastoopour-Irgens et al. (2020); Atmatzidou and Demetriadis (2016); Hogenboom et al. (2022); Kang et al. (2023); Kong and Wang (2023); Lai and Ellefson (2023); Lai (2021); Manfra et al. (2022); Mukasheva and Omirzakova (2021); Musaeus and Musaeus (2024); Su et al. (2024); Teng and Chung (2025); Tsai et al. (2022)
Evaluation	12	Atmatzidou and Demetriadis (2016); Mendoza-Diaz et al. (2023); El-Hamamsy et al. (2022); El-Hamamsy et al. (2025); Kang et al. (2023); Kong and Wang (2023); Li et al. (2021); Musaeus and Musaeus (2024); Su et al. (2024); Teng and Chung (2025); Tsai et al. (2022); Zhong et al. (2016)
Debugging	12	Atmatzidou and Demetriadis (2016); Cutumisu et al. (2022); Geng et al. (2025); Hogenboom et al. (2022); Kong and Wang (2023); Lai and Ellefson (2023); Lai (2021); Relkin et al. (2020); Shen et al. (2025); Sung (2022); Xiang et al. (2025); Zhong et al. (2016)
Modelling	7	Arastoopour-Irgens et al. (2020); Mukasheva and Omirzakova (2021); Relkin et al. (2020); Sung (2022); Teng and Chung (2025); Zhang et al. (2024); Zhong et al. (2016)
Logical Reasoning	5	Arastoopour-Irgens et al. (2020); Mendoza-Diaz et al. (2023); Fagerlund et al. (2020); Mukasheva and Omirzakova (2021); Romero et al. (2017)
Hardware / Software Awareness	4	Sung (2022); Relkin et al. (2020); Geng et al. (2025); Xiang et al. (2025)
Operators	2	Fagerlund et al. (2020); Roman-Gonzalez et al. (2017)
User Interactivity	2	Fagerlund et al. (2020); Romero et al. (2017)