International Journal of Computer Science Education in Schools, November 2024, Vol. 6, No. 4
ISSN 2513-8359

Computational Thinking Assessment in Primary and Secondary Education: A Meta-Synthesis of Tools, Methods And Pedagogical Approaches

Sanna Forsström[1]
Melissa Bond[2]

[1]University of Stavanger, Norway
[2]University College London, UK

DOI: 10.21585/ijcses.v6i4.208

Abstract

Despite the widespread adoption of computational thinking (CT) across educational levels, challenges persist in its assessment due to diverse definitions, frameworks, and practical applications in classroom settings. This meta-synthesis investigates the assessment of computational thinking (CT) in primary and secondary education, synthesising evidence from 12 reviews across five international databases, focusing on tools, methods, and pedagogical practices employed in assessing CT, with the aim to outline practical approaches for evaluating CT components. The review delves into the primary focuses of these syntheses, the CT skills and components assessed, and the methods and tools utilised, identifying gaps in current practices. The findings highlight a prevalent focus on programming skills, with less emphasis on cognitive processes and collaborative aspects of CT. The synthesis also points to the need for developing assessment tools and methods that encompass the broader spectrum of CT skills, suggesting avenues for future research and practical application in educational settings.

Keywords: Assessment, Computational Thinking, Review of reviews, Meta-synthesis

1. Introduction

The shift towards digital education is reshaping the educational field, introducing new teaching methodologies and broadening the horizons of learning. As the educational sector navigates through advancements in artificial intelligence, data management, cloud computing, and green technologies, teachers encounter various obstacles. These include managing the classroom, assessing student learning outcomes, dealing with ethical issues, and the effective use of digital tools (González-Pérez & Ramírez-Montoya, 2022). Central to overcoming these challenges is enhancing students' computational thinking (CT) skills, which are essential for promoting teamwork, critical analysis, and ethical decision-making (González-Pérez & Ramírez-Montoya, 2022; Ye et al., 2022).

The widespread adoption of computational thinking in primary and secondary education highlights its significance (Balanskat & Engelhardt, 2015). Nonetheless, there exists a diversity in how CT is understood among teachers, policymakers, and the media. This ranges from focusing on the core principles of computer science to its application across various disciplines (Lodi & Martini, 2021). The evolution of CT spans from Papert's (1980) vision of empowering individuals and enriching their understanding of complex topics through computational strategies to Wing's (2006) perspective. Wing viewed CT as not merely a set of technical skills but a comprehensive methodology for problem-solving, system design, and understanding human behaviour through computational principles, such as the ways individuals think, act, and interact. This underscores the necessity of integrating CT into education as a foundational skill alongside reading, writing, and arithmetic.

Following Wing's (2006) foundational work, numerous researchers have aimed to clearly define computational thinking (CT) and outline its key components (Lodi, 2020). Despite the creation of various frameworks to categorise CT into multidimensional constructs, the task of assessing

CT in primary and secondary education remains challenging. The diverse definitions and classifications, particularly those focusing on cognitive processes, such as logical reasoning or creativity, present difficulties for teachers in both understanding and evaluating CT.

Previous systematic reviews have examined the assessment of CT from different angles. Much of this research has focused on the theoretical aspects of CT assessment as part of research methodology (e.g., Araujo et al., 2016; Liu et al., 2021; Pan et al., 2023), with less attention given to practical implementation in the classroom. When studies do explore CT assessment within educational contexts, they often emphasise programming tasks using specific tools (e.g., Varghese & Renumol, 2023; Tan et al., 2023; da Cruz Alves et al., 2019) or concentrate on particular education levels (e.g., Fagerlund, 2021).

This meta-synthesis seeks to clarify these complexities by synthesising evidence syntheses that approaches the assessment of computational thinking (CT) in primary and secondary education with a variety of tools, methods, and definitions. This includes a focus on both research methodologies and methods used directly in schools. By interpreting these findings, our goal is to aggregate and thematically synthesise insights to outline practical assessment tools, methods, and pedagogical practices.

The central research question of this meta-synthesis is:

> *How can we understand the assessment of computational thinking in primary and secondary education through the synthesis of evidence syntheses?*

This inquiry also involves examining the connections between different methods and tools with specific CT components. To support this exploration, the following sub-questions guided the review:

1. What is the primary focus of these evidence syntheses?
2. Which CT skills and components are most frequently assessed?

3. What are the less commonly assessed CT components?
4. What methods and tools are used for assessing various CT skills/components?
5. How can different methods and tools be connected to specific CT components based on our interpretation of the results?
6. Which pedagogical practices for the formative assessment of CT are identified?
7. What limitations regarding assessment methods, tools, and pedagogical practices are reported?

2. Methodology

To address the research questions, a 'review of reviews' methodology was adopted, focusing on the selection, data extraction, and synthesis of evidence syntheses (Booth et al., 2022). This approach, recognised as a tertiary review (Kitchenham et al., 2009), aggregates findings from systematic reviews and follows documentation standards in accordance with PRISMA protocols (Page et al., 2021), ensuring rigour and transparency, with all search information available to be downloaded from the OSF[1].

2.1 Search Strategy and Selection of Studies

The evidence syntheses included in this study were identified through a wider scoping meta-review of programming, robotics and computational thinking studies, the methodological framework of which was inspired by prior tertiary analyses (Bond et al., 2024; Buntins et al., 2023).

Development of the Search String

The formulation of the search string (see Fig. 1) was adapted from earlier tertiary reviews (Bond et al., 2024; Buntins et al., 2023) and focused on programming, computational thinking, and

[1] https://osf.io/m8v3r/?view_only=b3c9360dfc2641a8b11c8d2d9924db50

robotics within the K-12 educational framework, alongside various evidence synthesis methodologies (Sutton et al., 2019). Unlike some reviews that specialise on a single secondary research type, such as meta-analyses (Higgins et al., 2012), this study aimed to encompass the entire spectrum of evidence synthesis techniques. This inclusive approach was chosen to fully map the domain without constraining the review to specific secondary research methodologies.

Programming	programming OR "computational thinking" OR coding OR robotics OR robots
AND	
Education sector	school OR "K-12" OR "primary education" OR "secondary education"
AND	
Evidence synthesis	"systematic review" OR "scoping review" OR "narrative review" OR "meta-analysis" OR "evidence synthesis" OR "meta-review" OR "evidence map" OR "rapid review" OR "umbrella review" OR "qualitative synthesis" OR "configurative review" OR "aggregative review" OR "thematic synthesis" OR "framework synthesis" OR "mapping review" OR "meta-synthesis" OR "qualitative evidence synthesis" OR "critical review" OR "integrative review" OR "integrative synthesis" OR "narrative summary" OR "state of the art review" OR "rapid evidence assessment" OR "qualitative research synthesis" OR "qualitative meta-summary" OR "meta-ethnography" OR "meta-narrative review" OR "mixed methods synthesis" OR "scoping study" OR "systematic map"

Figure 1. Meta-synthesis search string

Inclusion/Exclusion Criteria and Screening Process

The initial search was executed in April 2023, with follow-up searches up to January 17, 2024, ensuring comprehensive literature coverage. The platforms and databases searched were the Web of Science, Scopus, EBSCOHost (including ERIC), and Progress, as these have been found suitable and comprehensive for conducting evidence synthesis in the wider social sciences (Gusenbauer & Haddaway, 2020). The choice was made not to search the ACM Digital Library as well, as a recent tertiary review of AI in Education (Bond et al., 2024) included only one extra study from that database that was not found through other methods. Instead, the OpenAlex platform (Priem et al., 2022) was also searched via evidence synthesis software EPPI Reviewer

(Thomas et al., 2023), which indexes approximately 209 million publications. Forward and backward snowballing was also undertaken using OpenAlex.

The search yielded 4,369 records, which were imported into EPPI Reviewer as text or RIS files (see Fig. 2). An initial screening removed 485 duplicates. Two reviewers screened the same 200 titles and abstracts against predefined inclusion/exclusion criteria (see Table 1) to ensure inter-rater consistency. After reaching full agreement, the reviewers assessed the titles and abstracts of the remaining 3,684 records.

The inclusion criteria focused on K-12 programming or computational thinking evidence syntheses, published in English after 2010, identifying 195 articles to screen on full text. To confirm inter-rater reliability, ten additional articles were reviewed by both, achieving unanimous agreement. Ultimately, 120 evidence syntheses were selected for detailed analysis and synthesis in EPPI Reviewer. From these, 12 studies that focused solely on the assessment of computational thinking were identified for data extraction and synthesis.

Table 1. Inclusion/exclusion criteria

Inclusion Criteria	Exclusion Criteria
Published 2010 - 2023	Published before January 2010
Formal K-12 settings	Higher education, informal learning
A form of secondary research	Primary research or a review without a method section
English language	Published in a language other than English
Journal article or conference paper	Bibliometric reviews, editorials, book chapters,
Applications of computational thinking assessment	dissertations, posters, abstracts, workshop papers

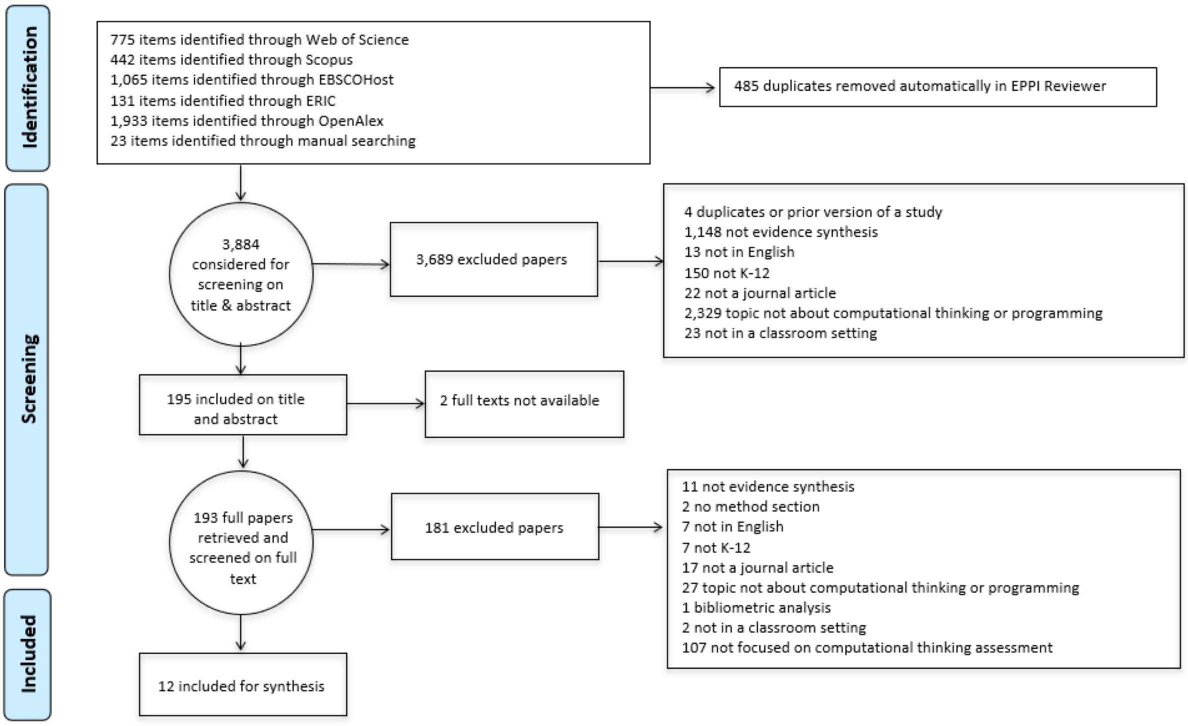

Figure 2. PRISMA diagram

2.2 Data Extraction

A data extraction coding tool was slightly modified from that of Bond et al. (2024)[2] and was developed in EPPI Reviewer. Codes included publication information (e.g., publication name and year published), authorship characteristics (e.g., country affiliation), review type (as self-identified by the authors), specific review focus if present (geographical or subject area), methodological characteristics (e.g., number of studies included), and benefits and challenges, which were inductively coded. Following the foundational work of Wing (2006), diverse definitions and categorisations of computational thinking (CT) have emerged (Lodi, 2020; Lodi & Martini, 2021). An initial review of the included evidence syntheses revealed a variety of definitions and categorisations. Consequently, we opted for inductive coding of the CT components used across these studies, developing our categories for a more coherent analysis.

[2] See https://osf.io/m8v3r/?view_only=b3c9360dfc2641a8b11c8d2d9924db50 for the full coding tool.

To thoroughly explore our research questions, we inductively coded all CT components/skills mentioned in the evidence syntheses. Additionally, we identified and coded the methods and tools utilised in CT assessment, along with any discussions on pedagogical practices and noted limitations. These coded elements formed the basis of a table that facilitated our synthesis process. Connecting methods and tools to assess CT components/skills involved analysing data from studies that explicitly detail these relationships. This included examining various tables and diagrams from sources such as Fagerlund et al. (2021), Babazadeh & Negrini (2022), da Cruz Alves et al. (2019), and Varghese and Renumol (2023), which illustrate the links between CT components and the assessment methods used. These elements were meticulously read, coded, and interpreted for relevance.

2.3 Synthesis

The findings were synthesised narratively (Petticrew & Roberts, 2006), and a tabulation of included studies was developed (see Appendix 1). Additional tables are presented within the text or as appendices, created using Word and Excel, supplemented by narrative descriptions. Regarding RQ2, given the wide array of definitions and descriptors for CT skills, each component was listed individually, merging similar ones and grouping the most frequently mentioned components into themes. RQ 5 was answered by identifying, examining and interpreting results presented in the studies analysing connections between CT components and assessment tools and methods. This process was complex due to the variability in how studies defined components and categorised their findings. Our interpretation of these results, aligned with our categorisation, aimed to clarify the practical application and utility of these methods in the CT assessment landscape. To map the results visually and to provide an openly accessible database to practitioners and researchers, a web database was created using the EPPI Visualiser app[3]. This allows users to explore the data by creating frequency and cross-tabulations, view all raw coding, and directly export and save metadata.

2.4 Limitations

Whilst every attempt was made to conduct this meta-synthesis as transparently and rigorously as possible, methodological limitations must be acknowledged. A protocol was not pre-registered; however, all metadata and coding are openly accessible via the OSF[4] and the EPPI Visualiser database. Five international databases were searched, along with snowballing. However, it is possible that potential includes were missed, given the more Western focus of the research indexed in these platforms (Mertala et al., 2022). Likewise, this review was limited to the period 2010-2023 and English language-only publications due to project resources. However, in the future, research in languages other than English should be included to help mitigate potential bias and improve generalisability (Bahji et al., 2022; Stern & Kleijnen, 2020). The choice to exclude evidence syntheses without a method section might also have led to the exclusion of more conceptual reviews, although this choice was made to include reviews that attested to being conducted to a particular standard of rigour (Bond et al., 2024).

3. Findings

3.1 General publication characteristics

The 12 studies included in this meta-review (see Appendix 1) were undertaken by the first authors from Europe (Babazadeh & Negrini, 2022; Fagerlund et al., 2021; Tikva & Tambouris, 2021), North America (Liu et al., 2021; Pan et al., 2023; Tan et al., 2023; Tang et al., 2020), South & Central America (Araujo et al., 2016; da Cruz Alves et al., 2019; Muñoz et al., 2023), and Asia (Haseski & Ilic, 2019; Varghese & Renumol, 2023), demonstrating a widespread

[3] https://eppi.ioe.ac.uk/eppi-vis/login/open?webdbid=597
[4] https://osf.io/m8v3r/?view_only=b3c9360dfc2641a8b11c8d2d9924db50

interest in and appreciation of the importance of understanding how CT research is being undertaken in order to best inform practice. However, less than half (42%, n = 5) are available open access, which limits the extent to which the results of these studies can be applied in practice or used to inform policy. The majority of reviews (n = 10) included both primary and high school students, five of which also included higher education (e.g., Muñoz et al., 2023). Two evidence syntheses explored CT in primary schools only; Fagerlund et al. (2021) concentrated on students using Scratch, while Liu et al. (2021) explored research methodologies for assessing CTs.

3.2 RQ1: What is the primary focus of the evidence syntheses?

Although the topic of each included evidence synthesis is the assessment of computational thinking in schools, they all have a slightly different focus (see Appendix 1). Some syntheses evaluated CT skills through programming activities, with Da Cruz Alves et al. (2019) investigating block-based languages and Fagerlund et al. (2021) focusing on Scratch, whilst other studies focused on specific methods or tools for assessing CT. Haseski and Ilic (2019) investigated the efficacy of paper-and-pencil tests, Varghese and Renumol (2023) assessed the use of video games, and Pan et al. (2022) examined the utility of think-aloud interviews for understanding student thought processes. Tan et al. (2023) explored the application of machine learning in assessment processes, while Varghese and Renumol (2023) evaluated digital games for comprehensive CT assessment, highlighting the increasing importance of digital media in education. Babazadeh and Negrini (2022) uniquely focused on the European context, emphasising geographical specificity in their review.

These syntheses vary in their emphasis, ranging from the exploration of research methodologies in CT assessment to the provision of actionable tools for teachers. This range underscores the complexity of adapting intricate research methodologies for practical use in the classroom, revealing a disconnect between theoretical research and practical teaching needs. The detailed nature of some research methods presents challenges for straightforward application by teachers, indicating a need to bridge the divide between academic research and classroom practice to make CT assessment both meaningful and feasible.

3.3 RQ2: Which CT skills and components are most frequently assessed?

Based on our thematic synthesis we divided the assessed CT skills and components into six main categories: 1) Core CT Components, 2) Programming Concepts, 3) Cognitive Processes, 4) Problem-Solving Strategies, 5) Collaborative and Communicative Skills, and 6) Dispositions and Attitudes (see Table 2).

The evidence syntheses in this corpus focused on four core CT components; abstraction, algorithmic thinking, decomposition and pattern recognition (see Table 2). Abstraction was the most prevalent, which relates to identifying and extracting relevant information while ignoring irrelevant details, followed by algorithmic thinking (developing a step-by-step solution to a problem), decomposition (breaking down a complex problem into smaller, more manageable sub-problems), and pattern recognition (identifying similarities, differences and patterns within and across problems).

Table 2. Skills and components reported within evidence syntheses

Core CT components	***n***	**% of reviews**
Abstraction	7	58%
Algorithmic thinking	6	50%
Decomposition	6	50%
Pattern recognition	4	33%
Programming concepts		

Sequencing	7	58%
Conditionals	6	50%
Parallelism	6	50%
Loops/iteration	5	42%
Variables & Data Representation	5	42%
Modularity	4	33%
Events & Synchronization	3	25%
Cognitive processes		
Logic & Reasoning	5	42%
Creativity and Innovation	3	25%
Critical Thinking	1	8%
Problem-solving strategies		
Problem-solving	5	42%
Debugging	3	25%
Testing	3	25%
Efficiency	2	16%
Collaborative & communication skills		
Collaboration & cooperation	3	25%
Communication	1	8%
Dispositions & attitudes		
Interest & engagement	1	8%

Seven specific programming concepts were identified (see Table 2), with sequencing (arranging steps in a logical order) the most frequent, followed by conditionals (making decisions based on specific conditions), parallelism (executing tasks simultaneously to increase efficiency), loops/iteration (repeating a set of instructions until a specific condition is met) and variables and data representation (storing, retrieving and manipulating data). Modularity (dividing a program into smaller, reusable parts) and events and synchronization (coordinating and synchronizing actions and events) were less considered.

Three cognitive processes were identified (see Table 2), with logic and reasoning (applying logical thinking to solve problems) being the most mentioned. This was followed by creativity and innovation (e.g., Fagerlund et al., 2021), and critical thinking, which was only found in one review (Tikva & Tambouris, 2021). Problem-solving strategies were divided into general developing and applying strategies to solve problems (e.g., Varghese & Renumol, 2023), and specific task-oriented strategies: debugging, testing and efficiency (optimizing a solution for better performance). Three studies (Fagerlund et al., 2021; Haseski & Ilic, 2019; Tikva & Tambouris, 2021) focused on working with others to achieve a common goal, one study (Tikva & Tambouris, 2021) explored expressing ideas and solutions effectively, and one study (Tan et al., 2023) investigated studies that related to being motivated to learn and apply CT skills.

3.4 RQ3: What are the less commonly assessed CT components?

Several evidence syntheses highlight gaps in assessing various CT components, overemphasizing tangible programming skills at the expense of broader, complex CT skills and affective variables. Fagerlund et al. (2021) point to a predominant focus on Scratch projects, which primarily assess programming skills but overlook broader thinking skills. Liu et al. (2021) discussed the underassessment of cognitive processes, particularly visual behaviors and verbalizations in CT problem-solving, whilst Haseski and Ilic (2019) noted the lack of studies measuring CT through affective variables, such as self-efficacy and attitude.

Muñoz et al. (2023) identified debugging, simulation, and decomposition as seldom assessed CT components in educational settings, highlighting the challenge for teachers in crafting activities that accurately measure these competencies. Varghese and Renumol (2023) and

Araujo et al. (2016) revealed that competencies such as conditional logic, iteration, modularity, modeling, and parallelization are minimally explored. Da Cruz Alves et al. (2019) implied an emphasis on quantifiable programming aspects rather than abstract CT concepts, a sentiment echoed by Tan et al. (2023), who points to the infrequent assessment of skills such as creativity and collaboration due to the difficulties in quantification. Similarly, Tang et al. (2020) observed a preference for assessing tangible programming skills over abstract CT components.

In summary, the literature indicates that higher-level thinking skills, affective variables, and complex CT competencies such as debugging are infrequently assessed. This is attributed to various factors, including the lack of standardized assessment models, the complexity of these skills, and the difficulty in quantifying them, especially for creative and collaborative skills.

3.5 RQ4: What methods and tools are used for assessing various CT skills/components?

Assessment methods

A variety of methods are employed in the assessment of computational thinking (CT) across the included evidence syntheses (see Appendix 1). Based on the thematic synthesis, assessment methods were categorized as direct, indirect, and innovative to capture the full spectrum of Computational Thinking (CT) strategies, aligning with their focus on either tangible outputs, cognitive processes, or the application of emerging technologies.

Four direct assessment methods were coded, with *standard tests* being administered to measure specific CT skills the most used, including multiple-choice or open-ended questions (n = 6, 50%; e.g., Babazadeh & Negrini, 2022), followed by *artifact analysis* in four reviews (33%), which involves evaluating the products of CT activities, such as code or digital artifacts, to determine the level of computational understanding (e.g., Pan et al., 2023). *Code analysis* was identified in two reviews (Muñoz et al., 2023; da Cruz Alves et al., 2019), which is when manual or automated code analysis is used to evaluate students' programming projects and can involve tools that specifically assess block-based programming languages like Scratch. Students undertaking *self-evaluation* of their own work or performance was found in one review (Fagerlund et al., 2021).

Four indirect assessment methods were identified. Nine reviews (75%) reported the use of *interviews and questionnaires* to gauge students' understanding and thought processes related to CT concepts and practices (e.g., Tang et al., 2020), with four reviews mentioning the use of *observations* (e.g., Tikva & Tambouris, 2021), where either researchers or educators observe students during CT activities to assess skill application in real-time. Three reviews identified think-aloud protocols, where students verbalize their thought processes while engaging in CT tasks, providing insights into their problem-solving strategies (e.g., Liu et al., 2021), and two reviews (Fagerlund et al., 2021; Pan et al., 2023) reported the use of *log data and error analysis*, where log data from digital tools or error patterns in code are analysed to understand students' learning processes and misconceptions.

Four innovative and emerging methods were mentioned in regard to CT assessment methods. Two reviews (Muñoz et al., 2023; Tikva & Tambouris, 2021) explored evidence-centered design, which is a systematic approach to developing assessments that align with the targeted competencies and skills. Additionally, two reviews (Tikva & Tambouris, 2021; Varghese & Renumol, 2023) considered the use of data mining and machine learning, where computational techniques are employed to analyze large datasets, such as responses or interaction patterns, to identify CT skills. Furthermore, one review mentioned eye-tracking (Liu et al., 2021), which offers deeper insights by tracking where and how students focus their attention during CT tasks. Another review (Tan et al., 2023) identified gamified assessments, which integrate assessment into the learning experience using game-based environments to capture data on students' CT skills.

Many studies also combined various methods to provide a more comprehensive assessment. For instance, think-aloud interviews may be used alongside programming assignments, which is a combination seen in Pan et al. (2023), or surveys might be combined with machine learning techniques to assess broader cognitive strategies and validate CT skill levels, as in Tikva & Tambouris (2021) and Varghese and Renumol (2023).

Assessment tools

In the exploration of tools used for assessing Computational Thinking (CT), the referenced evidence syntheses provide insights into a range of instruments tailored to capture the varied dimensions of CT skill development. These tools, classified into several thematic categories, serve distinct functions within the assessment process and are detailed as follows.

Programming and Development Environments: Tools such as Dr. Scratch, Scratch, Ninja Code Village, and App Inventor are frequently used for artifact analysis and development of CT skills (Fagerlund et al., 2021; Liu et al., 2021; Araujo et al., 2016; Tang et al., 2020). These environments allow for both the creation of digital artifacts and the assessment of the coding process itself.

Standardized Assessments and Testing Platforms: Various tests and tasks, including Bebras tasks, Visual Blocks Creative Computing Test, and multiple-choice or open-ended paper-and-pencil tests, are applied to evaluate specific CT competencies across a broad educational spectrum (Haseski & Ilic, 2019; Babazadeh & Negrini, 2022; Muñoz et al., 2023).

Questionnaires and Surveys: Self-efficacy scales, ability scales, and custom online questionnaires are implemented to gauge students' perceptions and self-assessed proficiency in CT, providing a subjective measure of cognitive and affective aspects of CT learning (Babazadeh & Negrini, 2022; Varghese & Renumol, 2023).

Interactive and Observational Tools: Think-aloud protocols, interviews, and observation strategies enable the assessment of students' thought processes and problem-solving strategies in real-time, offering qualitative insights into their CT approach (Fagerlund et al., 2021; Pan, 2023; Varghese & Renumol, 2023).

Robotic and Hands-on Tools: The use of robotic kits, microcontrollers, and web tools encourages the practical application of CT concepts, with students demonstrating their understanding through tangible, interactive projects (Muñoz et al., 2023; Tikva & Tambouris, 2021).

Data Analytics and Feedback Instruments: Log data, error analysis, and student response analysis through gamified assessments and evidence-centered design offer quantitative insights into students' learning processes, highlighting areas for growth and development in CT skills (Tan, 2023; Pan et al., 2023).

Emerging Technologies: Eye-tracking, game-based learning environments, and machine learning techniques represent the cutting edge of CT assessment, capturing nuanced data on how students interact with CT tasks and engage with problem-solving (Liu et al., 2021; Varghese & Renumol, 2023; Tikva & Tambouris, 2021).

3.6 RQ5: How can different methods and tools be connected to specific CT components, based on our interpretation of the results?

Only Fagerlund et al. (2021), Babazadeh & Negrini (2022), Da Cruz Alves et al. (2019), and Varghese and Renumol (2023), provided information about connections between different CT components and tools and methods used in the assessment. The connections between CT components and skills were explicitly outlined only in Fagerlund et al. (2020). We have conducted the Figure 3 based on the results in Fagerlund et al. (2020), which demonstrates the relationship between the content of Scratch programming projects evaluated in Fagerlund et al.

(2020) and the methodologies employed in their assessment. This figure indicates that artifact analysis and tests are predominantly utilized for evaluating code constructs. Here, Fagerlund et al. (2020) specifically refer to the logical structures in conducted Scratch programs, such as sequences of blocks. Coding patterns, which denote combinations of code constructs functioning as broader programming units, are primarily assessed through artifact analysis. Furthermore, the assessment of programming activities in the studies included by Fagerlund et al. (2020) is mainly conducted via observations, discourse analysis, and interviews.

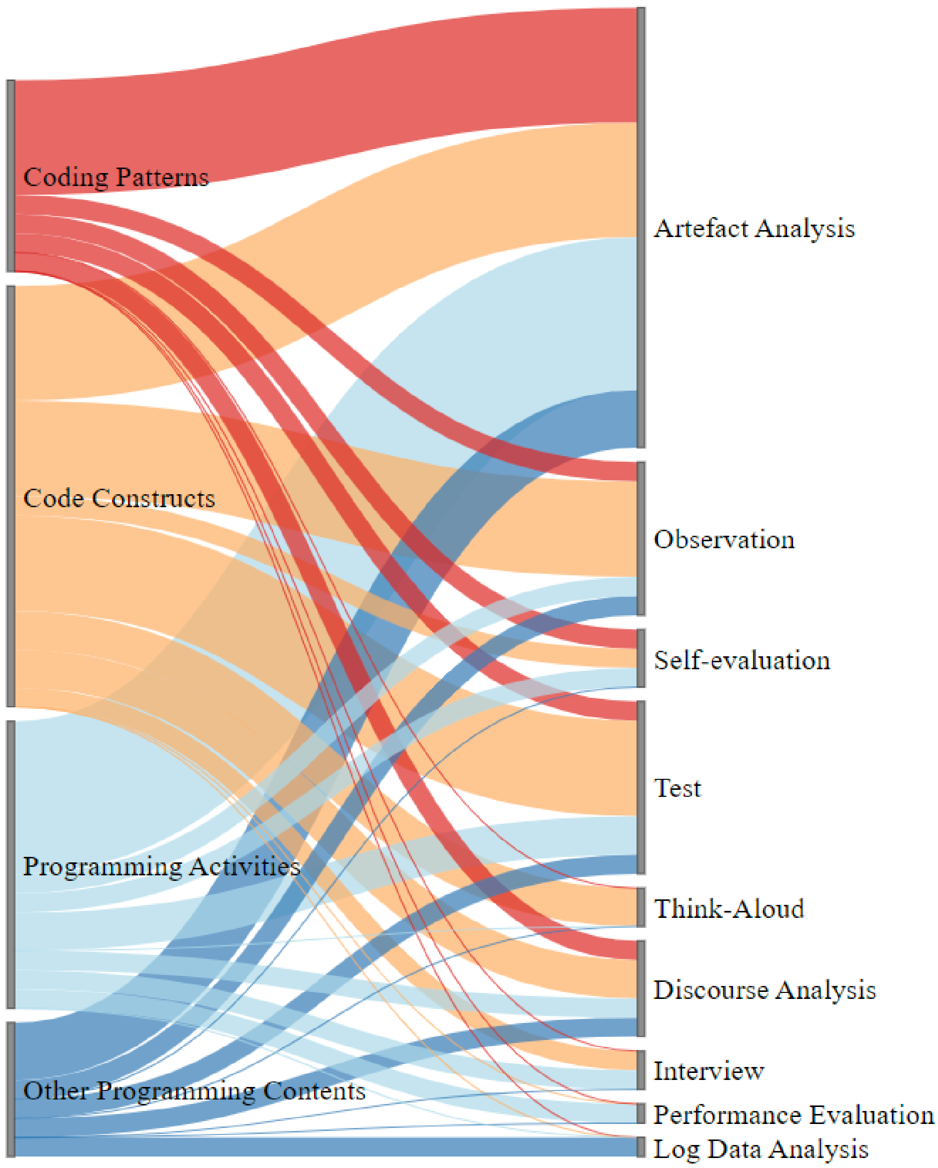

Figure 3. Our interpretation of the results in Fagerlund et al. (2020) about connections between assessed CT components and different methods.

Da Cruz Alves et al. (2019), Varghese and Renumol (2023), and Babazadeh and Negrini (2022) provide insights into the connections between assessed computational thinking (CT) components and the tools used in their assessment. Da Cruz Alves et al. (2019) focus on code analysis, thus the discussion is limited to tools appropriate for assessing codes. Varghese and Renumol (2023) specifically concentrate on video games as a tool for assessing CT, whereas Babazadeh and Negrini (2022) consider assessments solely in a European context but offer a broader range of assessment tools discussed.

The tools discussed in da Cruz Alves et al. (2019), including Hairball, Dr.Scratch, Ninja Code Village and Quizly, Fairy Assessment, were linked to CT components related to Core CT Components and Programming Concepts. These tools can quite straightforwardly be used to measure core CT components and programming concepts, and they can be categorised under the tool category of Programming and Development Environments. This was expected, given that da Cruz Alves et al. (2019) concentrated solely on code analysis.

Varghese and Renumol (2023) discuss what CT skills were assessed by researchers with videogames. Based on our interpretations on the results in Varghese and Renumol (2023), most of the assessed skills handled Core CT Components (Abstraction, Algorithmic Thinking, Decomposition, Pattern Recognition), Programming Concepts (Conditionals, Iteration, Modularity, Parallelism, Synchronization) or Problem-Solving Strategies (Debugging, Problem-solving, Efficiency). There were no assessed components that matched directly with Cognitive Processes, Collaborative and Communicative Skills, or Dispositions and Attitudes. To assess these CT components, researchers in the included studies in Varghese and Renumol (2023) used Interviews, Think-aloud protocols, and self-reported feedback surveys as assessment methods.

In Figure 4, we illustrate our interpretation of how different assessment tools are linked with various CT components, interpreted to align with our categories, in Babazadeh and Negrini (2022). It is apparent that Scratch, Alice, Dr. Scratch, and CT tests are frequently used to evaluate programming concepts. Core CT components are predominantly assessed through Java tasks, Bebras tasks, Robotics tasks, and Scratch tasks. The CT self-efficacy scale is also employed to evaluate core CT components. Cognitive processes are not that widely assessed, but the assessment is made with the CT ability scale, Scratch, and the CT self-efficacy scale. There are not many tools used to assess students' dispositions and attitudes or their collaboration and communication, based on our interpretation of the data in Babazadeh and Negrini (2022). Many of the tools used can be categorised as Programming and Development Environments or Standardized Assessments and Testing Platforms.

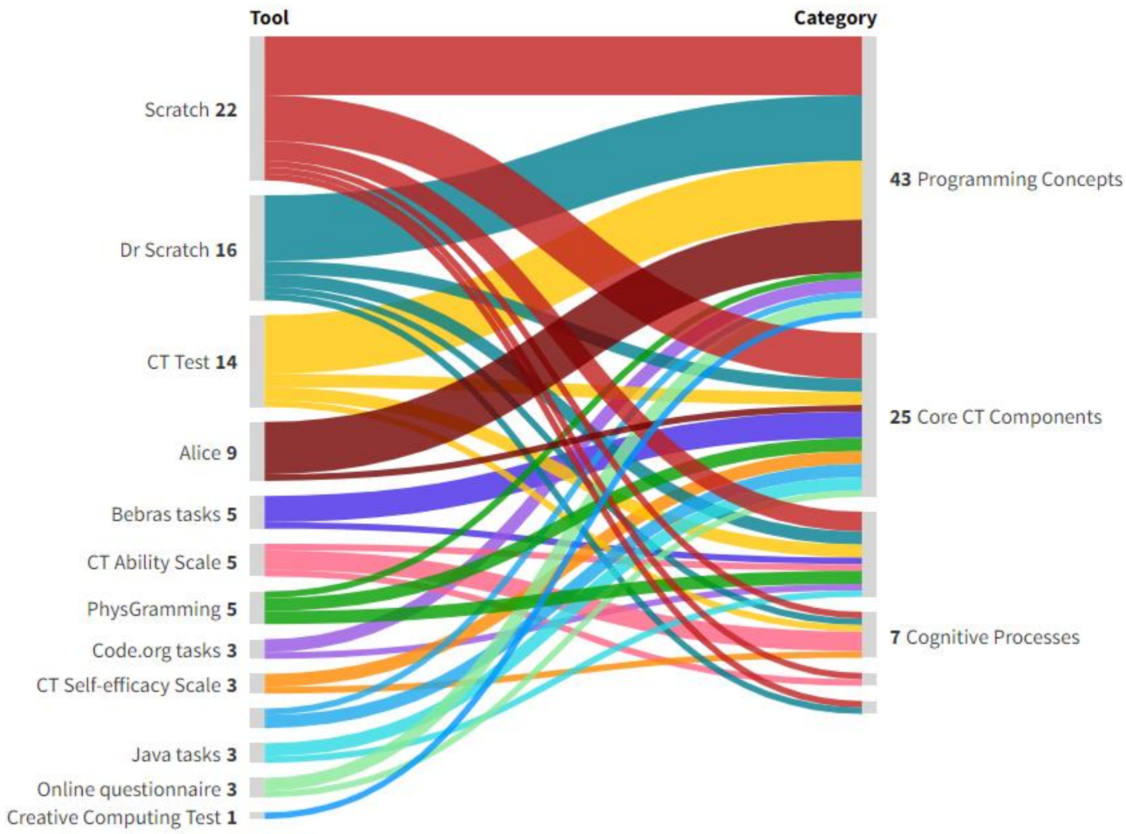

Figure 4. Our interpretation of the results in Babazadeh and Negrini (2022) connections between different assessment tools and assessed CT components

3.7 RQ6: Which pedagogical practices for the formative assessment of CT are identified?

The pedagogical practices in the formative assessment of Computational Thinking (CT) were not that widely discussed in the included systematic reviews. However, the discussions were still varied and reflective of the complexity inherent in CT itself.

Many of the studies mentioned the focus on *learning processes* instead of learning results in assessing CT. Pan et al. (2023) suggested that think-aloud protocols are effective in identifying students' learning processes, allowing teachers to diagnose and subsequently address gaps in CT comprehension and application. This method fosters an understanding of the individual student's thought process, providing insights into their problem-solving strategies. Similarly, Fagerlund et al. (2021) focus on project evaluations that examine students' coding projects in Scratch, assessing not just the final product but also the developmental processes of CT skills.

Further, the studies discussed the possibilities to give feedback to the students with the help of automated assessment tools. The use of immediate and *automated feedback* is exemplified by da Cruz Alves et al. (2019), who described how tools like Dr. Scratch can offer instant feedback on code quality and complexity. Tan et al.(2023) extend this concept with adaptive feedback informed by machine learning algorithms, delivering a personalised learning experience that evolves with the student's performance in CT tasks. Fagerlund et al. (2021) demonstrate the use of *scaffolding* through structured rubrics in Scratch projects, which guide students in developing their CT competencies. This strategic approach provides a clear pathway for students to follow, marking progressions in their skill development and understanding of CT principles.

Many assessment methods discussed in the studies are mainly used for research purposes. However, many of these methods can be used in school activities. One of these methods is a portfolio-driven approach that is discussed by Tang et al.(2020), which is a purposeful, systematic process of collecting and assessing different types of student products. The material can also include student observations and notes based on their work. Student observation was also mentioned in other studies (Fagerlund et al., 2020). However, many of these research methods were only mentioned briefly, with authors neglecting to include wider discussions about the pedagogical use of these methods in the classroom.

The studies also emphasised a *holistic approach* and *differentiated instruction* in CT assessment. The holistic approach, enhanced, for instance, by portfolios, assesses various CT components collectively through multiple methods (Fagerlund et al., 2020; Tang et al., 2020). While many studies suggested integrating these diverse methods to support differentiated instruction—tailoring assessments to meet students' unique learning needs (Liu et al., 2021)—they often stop short of discussing how to effectively combine these methods in practice, leaving a gap in the pedagogical application.

Finally, the included reviews mentioned approaches that can be referred to as *collaborative learning*. Tikva and Tambouris (2021) implied the use of collaborative learning strategies through the inclusion of game design and project-based learning, although the study does not detail specific formative assessment practices. Regarding collaborative learning, Fagerlund et al. (2020) also mentioned peer-to-peer assessment, referring to interactions between students during assessment processes. This was, however, only briefly mentioned and not discussed in other studies.

3.8 RQ7: What limitations regarding assessment methods, tools and pedagogical practices are reported?

The studies reported various challenges in assessment processes regarding pedagogy, tools used and methods, which are listed below.

Pedagogical Challenges:

- Complexity and Time Constraints: Tools designated for research are often impractical for classroom use due to their complexity and the extensive time required for implementation (Fagerlund et al., 2021).
- Insufficient Technological and Pedagogical Resources: Challenges such as lack of technological infrastructure, time for planning and material preparation, and limited instructional time (Tikva & Tambouris, 2021).
- Teachers' CT Content Knowledge: Uncertainties about appropriate CT content for different student age groups and the need for enhanced teacher knowledge and proficiency in teaching CT (Tikva & Tambouris, 2021).

Challenges with Tools:

- Narrow Assessment Tools: A focus on programming and coding, particularly with block-based languages like Scratch, while neglecting broader CT skills and other programming languages (Araujo et al., 2016; da Cruz Alves et al., 2019).
- Underexplored Data Analysis Techniques: The potential of data mining and machine learning is not fully harnessed in existing tools (Varghese & Renumol, 2023).

Challenges with Methods:

- Single-Method Limitations: The use of single-method approaches can fail to capture the full spectrum of CT, necessitating the development of multifaceted approaches (Muñoz et al., 2023).
- Variability and Lack of Standardization: Variability in how assessments are conducted and a lack of standardised procedures challenge the methodological rigour and the consistency of assessments across different settings (Pan et al., 2023).

- Conceptual Consistency and Validation: A lack of consensus on the definition of CT leads to confusion in assessment constructs, and there are challenges in validating the reliability of assessment instruments (Haseski & Ilic, 2019; Tang et al., 2020).

4. Discussion

The landscape of assessing computational thinking (CT) in educational settings appears to be both rich and complex, revealing a 'jungle' that teachers must navigate. This review has undertaken an inductive approach to synthesise what is known in the field and identify the gaps that persist. Such an exploration is vital, as it not only highlights the diversity in assessment methods and tools but also underscores the challenges and opportunities facing teachers in implementing effective CT assessments.

The results of this review indicate a prevalent focus on programming and core CT components across various studies. These components are often assessed using direct methods such as artefact analysis and testing, employing tools from Programming and Development Environments or Standardized Assessments and Testing Platforms (e.g., Scratch, Dr. Scratch, Bebras tasks). However, a noticeable gap exists in assessing the broader spectrum of CT skills, particularly cognitive processes, dispositions, attitudes, and collaborative aspects. These are crucial elements of CT as defined by pioneers like Papert (1980) and Wing (2006), which transcend mere programming skills.

While much of the current research has concentrated on theoretical frameworks and methodologies for assessing CT (Araujo et al., 2016; Liu et al., 2021; Pan et al., 2023), there has been comparatively less emphasis on the practical implementation of these assessments in classroom settings. When CT assessment is explored within educational contexts, studies often highlight programming tasks using specific tools (Varghese & Renumol, 2023; Tan et al., 2023; da Cruz Alves et al., 2019) or focus on specific education levels (Fagerlund, 2021). Our findings underscore this trend and point to the need for a more holistic approach that not only includes a broader range of CT skills but also emphasises their practical application in diverse classroom environments.

Interestingly, more indirect methods such as observation and interviews, typically utilised in research, offer insights into these less commonly assessed components. However, their application in classroom settings poses significant challenges due to their time-consuming nature and the specialised expertise required for conducting and interpreting such assessments. This discrepancy highlights a gap between research methodologies and practical pedagogical tools available for teachers.

4.1 Pedagogical Implementations and Limitations

This review suggests that while many tools and methods are designed for research purposes, their direct translation into pedagogical practice remains limited. The need for methodologies that are feasibly integrated into classroom activities is evident. Specifically, the use of peer assessment (Fagerlund et al., 2020) as a pedagogical method in the form of a formative assessment of CT requires further exploration. Such strategies could potentially address the gap in assessing soft skills like collaboration, creativity, and dispositions within CT education.

Furthermore, despite the recognition of CT as encompassing more than programming skills, the majority of assessment methods and tools remain focused on coding aspects. This imbalance points to a critical need for developing and validating tools that can effectively measure the full range of CT components. For teachers, this means exploring ways to operationalize research methods such as observations and interviews into their assessment practices, potentially through the development of observation sheets, interview guides, and other resources specifically designed for educational contexts.

4.2 Bridging Research and Practice

The transition from research to practice necessitates a deeper understanding and transformation of research methodologies into practical, pedagogical tools. There is a compelling need for resources that explicitly connect CT components with assessable elements, tailored for the teacher's use. Creating such resources, including detailed observation sheets and structured interview guides, could facilitate a more holistic assessment of CT skills in classroom settings. Moreover, integrating these tools with other assessment methods, such as self-assessment and peer-to-peer assessment, could enrich the formative assessment process, offering a more nuanced and comprehensive view of students' CT capabilities.

4.3 Implications for Practice, Future Studies, and Limitations

By synthesizing existing evidence and identifying practical assessment tools and methods, we have offered insights that can be useful in future research and development to create tools and practices for classroom use. Based on the results and methodology of this meta-synthesis, we can highlight some key implications for practice and future research.

Implications for Practice:

1. *Need for Practical Tools for Educators*: There is a need to develop user-friendly observation sheets and structured interview guides aligned with CT components identified in research, making them feasible for classroom use.
2. *Need for Integrated Assessment Methods*: It is necessary to use a combination of direct and indirect assessment methods, including self-assessment and peer-to-peer assessment, to capture a broader range of CT skills beyond programming tasks.
3. *Need for Professional Development for Teachers*: Professional development programs are needed to equip teachers with the skills and knowledge for effective CT assessment. These programs should focus on the practical application of research methodologies in

the classroom, ensuring teachers are well-prepared to implement the tools and strategies identified in our review.

Need for Future Studies:

4. *Practical Classroom Applications*: There is a need to translate theoretical frameworks into practical classroom applications, developing and validating tools that can effectively measure the full range of CT components in real-world educational settings.
5. *Systematic Reviews in Different Subjects*: New systematic reviews are needed that focus on the detailed connections between assessment methods and their practical implementation. These reviews should be conducted in the context of different subjects to explore how CT can be integrated across various disciplines.

Limitations of the Study:

6. *Review of Reviews Methodology*: The wide scope of the review of reviews methodology may have overlooked specific details about the connections between methods and their operationalization. Future studies should focus on these detailed connections to provide more targeted insights.
7. *Evolving Nature of CT*: The field of CT is continuously evolving, and new methodologies and tools are constantly being developed. This tertiary review may not capture the latest primary studies in this field but is able to provide a broad overview of the current approaches.

5. Conclusion

The discussion in this tertiary review underscores a critical intersection between theoretical research and practical teaching needs in the assessment of computational thinking. While the field has advanced in identifying and employing a variety of assessment methods and tools, significant gaps remain in translating these into accessible and effective pedagogical practices.

Addressing these gaps requires not only a re-evaluation of the tools and methods themselves but also a concerted effort to align them with pedagogical goals, ensuring that they are both meaningful and feasible for teachers to implement.

References

* Refers to a study included in the review corpus

* Araujo, A. L. S. O. de, Andrade, W. L., & Guerrero, D. D. S. (2016). A systematic mapping study on assessing computational thinking abilities. https://doi.org/10.1109/FIE.2016.7757419

* Babazadeh, M., & Negrini, L. (2022). How is computational thinking assessed in European K-12 education? A systematic review. *International Journal of Computer Science Education in Schools*, *5*(4). https://doi.org/10.21585/ijcses.v5i4.138

Bahji, A., Acion, L., Laslett, A.-M., & Adinoff, B. (2023). Exclusion of the non-English-speaking world from the scientific literature: Recommendations for change for addiction journals and publishers. *Nordic Studies on Alcohol and Drugs*, *40*(1), 6–13. https://doi.org/10.1177/14550725221102227

Balanskat, A., & Engelhardt, K. (2015). *Computing our future: Computer programming and coding - priorities, school curricula and initiatives across Europe.* Brussel European Schoolnet.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics collaboration and rigour. *International Journal of Educational Technology in Higher Education, 21*(4). https://doi.org/10.1186/s41239-023-00436-z

Booth, A., Sutton A., Clowes, M., Martyn-St James, M. (2022). *Systematic Approaches to a Successful Literature Review.* SAGE

Buntins, K., Bedenlier, S., Marín, V., Händel, M., & Bond, M. (2023). Methodological approaches to evidence synthesis in educational technology: A tertiary systematic mapping review. MedienPädagogik, 54, 167–191. https://doi.org/10.21240/mpaed/54/2023.12.20.X

* da Cruz Alves, N., Gresse von Wangenheim, C., & Hauck, J. C. R. (2019). Approaches to Assess Computational Thinking Competences Based on Code Analysis in K-12 Education: A Systematic Mapping Study. *Informatics in Education, 18*(1), 17-39. https://doi.org/10.15388/infedu.2019.02

* Fagerlund, J., Häkkinen, P., Vesisenaho, M., & Viiri, J. (2021). Computational thinking in programming with Scratch in primary schools: A systematic review. *Computer Applications in Engineering Education, 29*(1), 12–28. https://doi.org/10.1002/cae.22255

González-Pérez, L. I. & Ramírez-Montoya, M, S. (2022). Components of Education 4.0 in 21st Century Skills Frameworks: Systematic Review. *Sustainability 14* (3). https://doi.org/10.3390/su14031493

* Haseski, H. İ., & Ilic, U. (2019). An Investigation of the Data Collection Instruments Developed to Measure Computational Thinking. *Informatics in Education, 18*(2), 297–319. https://doi.org/10.15388/infedu.2019.14

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). The impact of digital technology on learning: A summary for the Education Endowment Foundation. Education Endowment Foundation. https://eric.ed.gov/?id=ED612174

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—A systematic literature review. Information and Software Technology, 51(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

* Liu, R., Luo, F., & Israel, M. (2021). What Do We Know about Assessing Computational Thinking? A New Methodological Perspective from the Literature. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2021), June 26-July 1*, Virtual Event, Germany. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3430665.3456380

Lodi, M. (2020). Informatical Thinking. *Olympiads in Informatics, 14*, 113–132. https://doi.org/10.15388/ioi.2020.09

Lodi, M., & Martini, S. (2021). Computational thinking between Papert and Wing. *Science & Education, 30*, 883–908. https://doi.org/10.1007/s11191-021-00202-5

Mertala, P., Moens, E. & Teräs, M. (2022). Highly cited journal articles on educational technology: A descriptive and critical analysis. *Learning, Media and Technology*, 1-14. https://doi.org/10.1080/17439884.2022.2141253

* Muñoz, R. F. Z., Hurtado Alegría, J. A., & Robles, G. (2023). Assessment of Computational Thinking Skills: A Systematic Review of the Literature. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 18*(4), 319-330. https://doi.org/10.1109/RITA.2023.3323762

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hofmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaf, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ (clinical Research Ed.), 372, n71. https://doi.org/10.1136/bmj.n71

* Pan, Z., Cui, Y., Leighton, J. P., & Cutumisu, M. (2023). Insights into computational thinking from think-aloud interviews: A systematic review. *Applied Cognitive Psychology, 37*(1), 71–95. https://doi.org/10.1002/acp.4029

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

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv. https://arxiv.org/abs/2205.01833

Stern, C. & Kleijnen, J. (2020). Linguistic biases in systematic reviews: You only get out what you put in. *JBI Evidence Synthesis*, 18(9), 1818-1819. https://doi.org/10.11124/JBIES-20-00361

Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: Exploring review types and associated information retrieval requirements. Health Information and Libraries Journal, 36(3), 202–222. https://doi.org/10. 1111/hir.12276

* Tan, B., Jin, H.-Y., & Cutumisu, M. (2023). The applications of machine learning in computational thinking assessments: A scoping review. *Computer Science Education*. https://doi.org/10.1080/08993408.2023.2245687

* 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

Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., Bond, M., & Koryakina, A. (2023). EPPI Reviewer: Advanced software for systematic reviews, maps and evidence synthesis [Computer software]. EPPI Centre Software. UCL Social Research Institute. London. https://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4

* Tikva, C., & Tambouris, E. (2021). Mapping computational thinking through programming in K-12 education: A conceptual model based on a systematic literature review. *Computers & Education, 162*, 104083. https://doi.org/10.1016/j.compedu.2020.104083

Varghese, V. V., & Renumol, V. G. (2023). Video games for assessing computational thinking: A systematic literature review. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-023-00284-w

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

Ye, J., Lai, X., & Wong, G. K.-W. (2022). The transfer effects of computational thinking: A systematic review with meta-analysis and qualitative synthesis. *Journal of Computer Assisted Learning*, *38*(6), 1620–1638. https://doi.org/10.1111/jcal.12723

Appendix 1: K-12 Computational thinking assessment reviews (n = 12)

Research Question 2: Which skills and components are most frequently assessed?

Author	Year	OA[1]	Type	# Studies	Years	Author Countries	Primary Focus	Core CT components				Programming Concepts							Cognitive Processes			Problem-solving				Collaborative & communicative skills		Dispositions & Attitudes
								Abstraction	Algorithmic thinking	Decomposition	Pattern recognition	Sequencing	Loops/Iteration	Conditionals	Variables & Data Representation	Modularity	Parallelism	Events & Synchronization	Logic & reasoning	Creativity & innovation	Critical thinking	Problem solving	Debugging	Testing	Efficiency	Collaboration & Cooperation	Communication	Interest & Engagement
Araujo et al.	2016	-	MR	27	2009 - 2016	Brazil	Identify, classify & review approaches & methodologies for assessing CT.	X	X	X		X	X	X	X	X	X		X			X	X					
Babazadeh & Negrini	2022		SR	26	2016 - 2020	Italy	Analyse how CT is assessed in European K-12 education, identifying tools used for assessment & CT dimensions evaluated.	X	X	X	X	X	X	X					X									
da Cruz Alves et al.	2019		MR	23	1997 – 2018	Brazil	Evaluate existing approaches for assessing CT through code analysis in K-12.					X		X	X	X				X				X				
Fagerlund et al.	2021		SR	30	2007 – 2019	Finland	Investigate the assessment of CT through Scratch in primary education.	X		X							X			X					X	X		
Haseki & Ilic	2019		SR	64	2010 - 2018	Turkey	Investigate the properties of paper-and-pencil data collection instruments to measure CT.	X	X	X	X	X					X					X				X		
Liu et al.	2021	-	IR	28	2010 - ?	USA	Analyse research methodologies for assessing CT in primary education.																	X				

[1] Cropped from the Open Access Logo image at https://creativecommons.org/about/program-areas/open-access/open-access-logo/

Note: SR = Systematic Review, MR = Mapping Review, MS = Meta-synthesis, IR = Integrative review, ScR = Scoping review, ML = Machine learning

ISSN 2513-8359

Appendix 1: K-12 Computational thinking assessment reviews (n = 12)

Research Question 2: Which skills and components are most frequently assessed?

Author	Year	OA[2]	Type	# Studies	Years	Author Countries	Primary Focus	Core CT components				Programming Concepts							Cognitive Processes			Problem-solving				Collaborative & communicative skills		Dispositions & Attitudes
								Abstraction	Algorithmic thinking	Decomposition	Pattern recognition	Sequencing	Loops/Iteration	Conditionals	Variables & Data Representation	Modularity	Parallelism	Events & Synchronization	Logic & reasoning	Creativity & innovation	Critical thinking	Problem solving	Debugging	Testing	Efficiency	Collaboration & Cooperation	Communication	Interest & Engagement
Muñoz et al.	2023	-	SR	65	2012 – 2022	Colombia, Spain	Identify methods of assessing CT in K-12 & CT skills.	X				X		X	X		X	X	X									
Pan et al.	2023	-	SR	35	2011 – 2021	Canada	Examine the use of think-aloud interviews in CT & the methodology for understanding cognitive processes.															X						
Tan et al.	2023	-	ScR	20	2014 - 2021	Canada	Examine scope of ML approaches to assess CT (educational context, data, algorithms, aspects of CT assessed).																					X
Tang et al.	2020	-	SR	96	2011 – 2019	USA	Review how CT has been assessed (tools, educational context, subjects).	X	X	X	X	X	X	X	X		X		X									
Tikva & Tambouris	2021	-	SR	101	2006 – 2019	Greece	Develop a conceptual model based on K-12 CT programming studies.	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	
Varghese & Renumol	2023		SR	11	2010 – 2021	India	Examine the effectiveness of using video games for assessing CT skills.		X		X		X			X		X				X	X	X				

[2] Cropped from the Open Access Logo image at https://creativecommons.org/about/program-areas/open-access/open-access-logo/

Note: SR = Systematic Review, MR = Mapping Review, MS = Meta-synthesis, IR = Integrative review, ScR = Scoping review, ML = Machine learning

Appendix 1: K-12 Computational thinking assessment reviews (n = 12)

Research Question 4: What methods and tools are used for assessing various CT skills/components?

Author	Year	OA[3]	Type	# Studies	Years	Author Countries	Primary Focus	Direct assessment methods				Indirect assessment methods				Innovative & emerging methods			
								Self-/performance evaluation	Artifact analysis	Testing	Code analysis	Interviews & questionnaires	Observation	Think-aloud protocols	Log data & error analysis	Eye-tracking	Gamified assessments	Evidence-centred design	Data mining & machine learning
Araujo et al.	2016	-	MR	27	2009 - 2016	Brazil	Identify, classify & review approaches & methodologies for assessing CT.			X		X							
Babazadeh & Negrini	2022		SR	26	2016 - 2020	Italy	Analyse how CT is assessed in European K-12 education, identifying tools used for assessment & CT dimensions evaluated.			X		X	X						
da Cruz Alves et al.	2019		MR	23	1997 – 2018	Brazil	Evaluate existing approaches for assessing CT through code analysis in K-12.				X								
Fagerlund et al.	2021		SR	30	2007 – 2019	Finland	Investigate the assessment of CT through Scratch in primary education.	X	X			X	X	X	X				
Haseki & Ilic	2019		SR	64	2010 - 2018	Turkey	Investigate the properties of paper-and-pencil data collection instruments to measure CT.												
Liu et al.	2021	-	IR	28	2010 - ?	USA	Analyse research methodologies for assessing CT in primary education.		X	X		X		X		X			

[3] Cropped from the Open Access Logo image at https://creativecommons.org/about/program-areas/open-access/open-access-logo/

Note: SR = Systematic Review, MR = Mapping Review, MS = Meta-synthesis, IR = Integrative review, ScR = Scoping review, ML = Machine learning

Appendix 1: K-12 Computational thinking assessment reviews (n = 12)

Research Question 4: What methods and tools are used for assessing various CT skills/components?

Author	Year	OA[4]	Type	# Studies	Years	Author Countries	Primary Focus	Direct assessment methods				Indirect assessment methods				Innovative & emerging methods			
								Self-/performance evaluation	Artifact analysis	Testing	Code analysis	Interviews & questionnaires	Observation	Think-aloud protocols	Log data & error analysis	Eye-tracking	Gamified assessments	Evidence-centred design	Data mining & machine learning
Muñoz et al.	2023	-	SR	65	2012 – 2022	Colombia, Spain	Identify methods of assessing CT in K-12 & CT skills.				X	X						X	
Pan et al.	2023	-	SR	35	2011 – 2021	Canada	Examine the use of think-aloud interviews in CT & the methodology for understanding cognitive processes.		X			X	X	X	X				
Tan et al.	2023	-	ScR	20	2014 - 2021	Canada	Examine scope of ML approaches to assess CT (educational context, data, algorithms, aspects of CT assessed).			X		X					X		
Tang et al.	2020	-	SR	96	2011 – 2019	USA	Review how CT has been assessed (tools, educational context, subjects).			X		X							
Tikva & Tambouris	2021	-	SR	101	2006 – 2019	Greece	Develop a conceptual model based on K-12 CT programming studies.		X			X	X					X	X
Varghese & Renumol	2023		SR	11	2010 – 2021	India	Examine the effectiveness of using video games for assessing CT skills.			X									X

[4] Cropped from the Open Access Logo image at https://creativecommons.org/about/program-areas/open-access/open-access-logo/

Note: SR = Systematic Review, MR = Mapping Review, MS = Meta-synthesis, IR = Integrative review, ScR = Scoping review, ML = Machine learning