ISSN Online 2527-5933
MATH-EDU: JurnalIlmu Pendidikan Matematika
Volume 11, Nomor 2, Agustus 2026, pp. 180-197

Assessment Methods for Computational Thinking Skill in Mathematics Education: A Systematic Literature Review

Zahra As Syifa[1]*, Putri Nur Malasari,
[1,2]Mathematics Education Department, Universitas Islam Negeri Sunan Kudus, Central Java, Indonesia
*Correspondence: zasyifa860@gmail.com

Article Information

Received:
26th May 2026

Accepted:
30th August 2026

Publish:
31st August 2026

Keywords
Assessment, computational thinking, mathematics education, systematic literature review

Abstract
This study aims to identify CT assessment methods in mathematics education and analyze their advantages and limitations through a Systematic Literature Review (SLR) approach. Using the PRISMA guidelines, this study analyzed 13 articles published between 2018 to 2025 from national and international databases. The results show that CT assessment methods in Indonesia include the Bebras Task instrument, CT indicator-based essay tests, a combination of written tests and interviews, and technology-based digital instruments. The Bebras Task is able to measure CT components comprehensively but requires high abstraction skills. Descriptive tests can assess deep thinking processes but require long analysis times and are susceptible to the influence of students' basic mathematical abilities. Combination methods produce richer data but are difficult to standardize with limited generalizability. Digital instruments offer ease of access but depend on the availability of technological infrastructure. Findings indicate that no single assessment method is most effective, so instrument selection must be tailored to measurement objectives, learning context, and resource availability.

How to Cite: Syifa, Z. A. & Malasari, P. N. (2026). Assessment Methods for Computational Thinking Skill in Mathematics Education: A Systematic Literature Review. *Math-Edu: Jurnal Ilmu Pendidikan Matematika*, 11 (2), 180-197.

Introduction

Technological advances in the era of the Fourth Industrial Revolution and Society 5.0 have brought about major changes in education, including mathematics learning (Budiyanto et al., 2025). In this context, computational thinking (CT) is seen as an important new 21st-century literacy that all students should possess, alongside reading, writing, and arithmetic skills. CT emphasizes systematic thinking in formulating problems and developing solutions that can be processed by both humans and computers (Wing, 2006).

The link between CT and mathematics learning is getting stronger because both emphasize systematic problem-solving processes (Kallia et al., 2021). Various studies show that applying CT can not only deepen understanding of mathematical concepts, but mathematics learning designed based on CT can also strengthen students' CT abilities (Pei et al., 2018). In Indonesia, the need for CT integration is becoming increasingly urgent given the low performance of students in the Programme for International Student Assessment (PISA), one of which is related to students' weaknesses in applying CT in the context of mathematics (Prasetyawan et al., 2025).

Although the importance of CT has been widely recognized, there are several challenges in its implementation, particularly in relation to appropriate assessment methods. CT assessment is considered one of the main challenges in integrating CT into compulsory education (Allsop, 2019). Teachers report that CT assessment is very challenging because of its broad scope, which is not limited to coding alone. Varied definitions of CT and unclear curriculum objectives make it difficult for teachers to conduct effective assessments. This situation makes the need for comprehensive assessment guidelines and professional development programs for teachers even more urgent. Various instruments have been developed to measure CT, ranging from the Bebras Challenge, Computational Thinking Test (CTt), to programming-based instruments such as Scratch (Román-González et al., 2019). In Indonesia, several CT instruments have also been developed, but there has been no systematic review summarizing and evaluating these assessment methods (Fitriyah et al., 2024).

Research findings on Indonesian students' CT abilities in mathematics still show variation. Research Rianto et al. (2024) shows that 60% of junior high school students can use CT to solve geometry problems, with achievable CT components including abstraction, decomposition, and algorithms. However, students' CT abilities are still limited to pattern recognition, with deficiencies in abstract and algorithmic thinking skills. These findings indicate the need to develop more effective and comprehensive CT assessment methods in mathematics education.

However, the integration of CT in the Indonesian curriculum is still limited. CT is often only taught in computer science classes, while teacher skills, technological support, and existing assessment systems are not yet capable of optimally supporting CT development. In addition, unequal access to technology, especially in rural or underdeveloped areas, is a significant obstacle to CT education (Yuliana et al., 2021). The trend of research on CT in mathematics education has continued to increase in recent years. The focus of these studies includes teacher and student perceptions, assessment instrument development, CT integration in mathematics learning, CT-based learning models, and bibliometric analysis (Majid, 2025). However, there is still a scarcity of studies that specifically discuss the relationship or interaction between CT and K-12 mathematics thinking and learning (Shumway et al., 2021).

Given the significant role of CT in mathematics learning and the complexity of the assessment process, a systematic review is needed to summarize the various CT assessment methods that have been developed and used in education. This review is expected to map existing assessment methods, identify their strengths and weaknesses, and provide recommendations for the development of CT assessment instruments that are more relevant to the context of mathematics learning at various levels of education.

Methods

This study uses a Systematic Literature Review (SLR) approach. This approach uses data from articles searched in national and international databases such as Google Scholar, ResearchGate and Crossref with the keywords "assessment method", "computational thinking", and "mathematics education". The articles used were publications in international and national journals from 2018 to 2025 with the research subject of students. The SLR approach aims to examine, review, and synthesize the results of previous studies relevant to a specific focus of study, based on systematic research questions and predetermined criteria (Latifah & Ritonga, 2020). In research based on Systematic Literature Review (SLR), determining the research question is a fundamental stage that serves as the main guide in the entire review process. This research question is formulated to direct literature searches, article selection, and the data analysis and synthesis process. The research questions in this study are as follows:

RQ1 : What assessment methods are used to measure computational thinking skills in mathematics education?

RQ2 : What are the advantages and limitations of the assessment methods used to evaluate computational thinking skills in mathematics learning?

The article review process was carried out using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, as described by Akande et al. (2020), to minimize bias in the literature identification and selection process, thereby ensuring that the results of the study were more objective and comprehensive. The phases that must be carried out in this approach include determining the topic and focus of the study, searching for relevant studies, analyzing and selecting documents according to criteria, and assessing the quality of articles (through analysis, synthesis, and description) (Khurshid et al., 2020). The following are the steps taken by the author based on the PRISMA guidelines:

a. Identification

   There are 226 articles identified related to methods for measuring computational thinking skills in mathematics education using Publish or Perish 8 from 2018 to 2025.

b. Screening

   Of the 226 articles, 41 were not relevant to the context of mathematics education, 13 articles did not have elementary, junior high, or high school students as their subjects, 77 articles were not open access, and 16 data were not articles. In this stage, a total of 147 articles were removed.

c. Eligibility

   In this stage, 49 articles could not answer the research question and were therefore removed.

d. Included

   There were 13 articles that met the criteria for measuring computational thinking skills in mathematics education.

The article selection process above can be seen in the inclusion and exclusion criteria table below.

Table 1. Inclusion Exclusion Criteria

Aspect	Inclusion Criteria	Exclusion Criteria
Article duplicate	No duplicates	Duplicate articles
Method	Uses qualitative, quantitative, and developmental research methods	Uses SLR and Meta-Analysis research methods
Articles access	Full text articles that can be accessed	Articles that are not full text and cannot be accessed
Main topic	Articles discuss computational thinking and assessment methods	Articles do not discuss computational thinking and assessment methods
Context	Articles in the context of mathematics education	Outside articles in the context of mathematics learning
Research subject	Elementary, junior high, and high school students	Not elementary, junior high, or high school students
Research question	Articles can answer the research question	Articles cannot answer the research question

Based on the table above, the results of the article selection can be visualized in Figure 1 below.

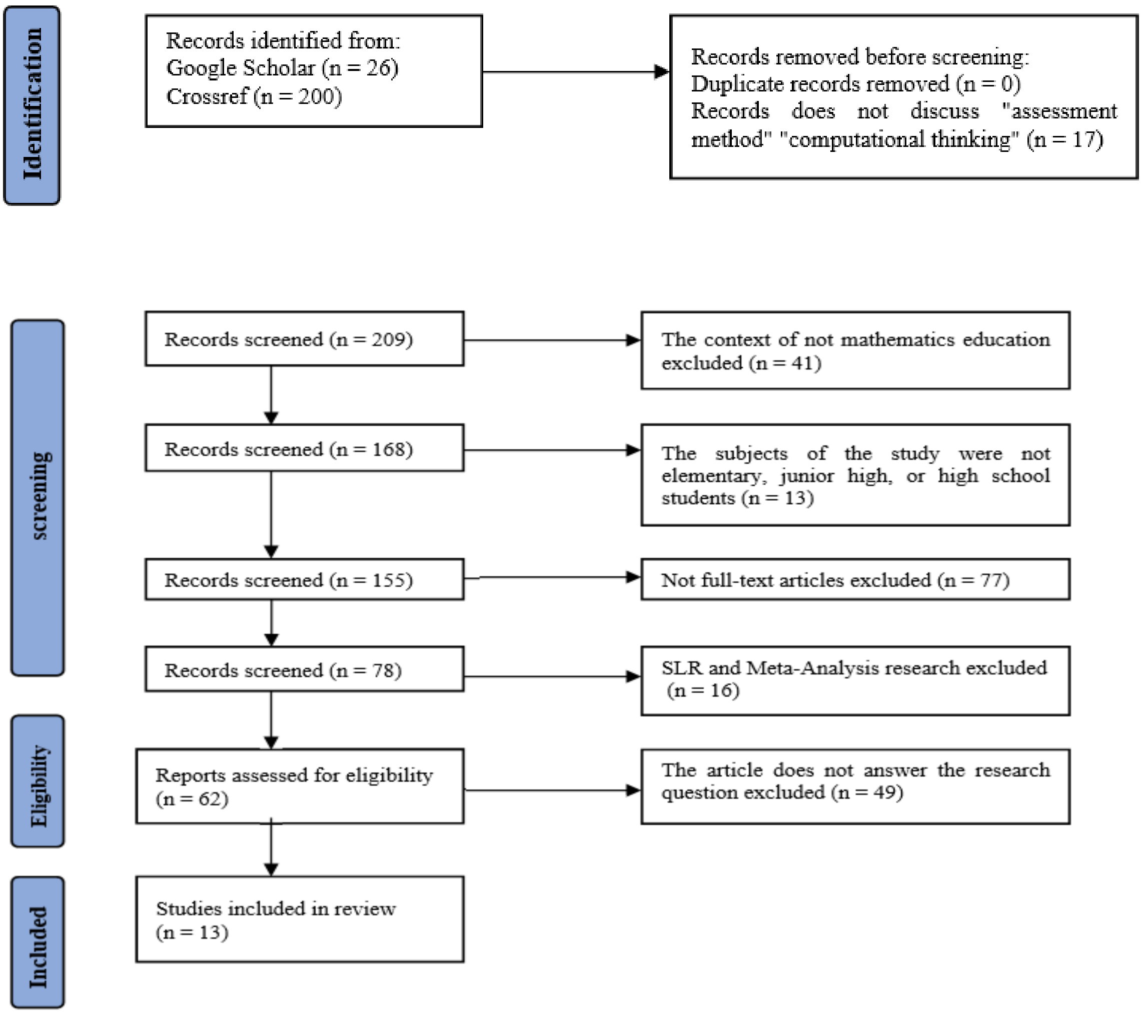

Figure 1. Prisma Diagram Of Computational Thinking

Result and Discussion

RQ1 : What assessment methods are used to measure computational thinking skills in mathematics education?

Computational thinking (CT) skills in mathematics learning require appropriate assessment instruments to measure the extent to which students can apply computational thinking skills in problem solving. To answer the first research question, an analysis was conducted on 14 studies that used various CT assessment methods in the context of mathematics learning in Indonesia. Table 2 below summarizes the assessment methods used by each researcher along with the characteristics of the instruments developed.

Table 2. Assessment Methods for Computational Thinking

Author	Result
(Mustaqimah, 2024)	Researchers used three Bebras Task questions from the Collector category (junior high school) entitled “Right and Left,” “City Full of Turns,” and “Four Jobs,” which were selected from the 2016–2018 period and modified as necessary.
(Mania, 2021)	This study used a performance-based written test that was systematically developed and validated using the Borg & Gall model to measure students' computational thinking skills.
(Hartawan., et al 2024)	This study used the Bebras Task adaptation essay test (performance-based written test) with CT indicators (decomposition, patterns & generalization, abstraction, algorithmic) and supporting interviews as assessment methods to measure the computational thinking abilities of junior high school students.
(Mukhibin., et al 2024)	The CT assessment method used is an indicator-based computational thinking test (decomposition, patterns, abstraction, algorithms) supplemented by qualitative interviews to assess students' thinking processes. The assessment focuses on analyzing the quality and depth of students' computational thinking in the context of solving mathematical problems (geometry).
(Anwar et al., 2025)	Researchers assessed computational thinking skills through direct observation of teacher and student activities, pre-tests and post-tests of programming comprehension, and CT assessment rubrics used during Scratch project presentation sessions. The assessment focused primarily on performance-based assessment as teachers and students programmed, solved problems, and implemented computational logic in the context of mathematics learning.
(Hasanah & Masduki, 2024)	The computational thinking (CT) assessment method in this study was conducted through a CT skills test covering four main indicators, namely decomposition, pattern recognition, abstraction, and algorithmic thinking. The test instruments were given to students with field independent (FI) and field dependent (FD) cognitive styles to explore differences in CT abilities based on cognitive characteristics.

(Susanti et al., 2025)	Researchers used a linear programming-based problem-solving test worksheet, think-aloud exercises, and semi-structured interviews. The test given to 98 students contained problem-solving problems that were analyzed using four computational thinking indicators.
(Fitri et al., 2024)	Researchers measured students' computational thinking abilities using test instruments administered by teachers in five classes. The scores for each indicator were then processed using the Tsukamoto Fuzzy algorithm to determine students' CT ability levels, which were grouped into low, medium, and high categories.
(Amelia et al., 2024)	In this study, students' computational thinking skills were measured using an Android-based mathematics assessment tool containing 15 essay questions specifically developed based on four CT indicators. This tool was contextualized with the Tabot Bengkulu culture and flat-sided spatial construction material, enabling students to analyze problems, recognize patterns, and systematically develop solution steps.
(Ahmadi et al., 2025)	In this study, students' computational thinking skills were measured using a descriptive test consisting of three math questions on number patterns. In addition to the written test, the researchers also used structured interviews as a supporting instrument to confirm students' thinking processes in three categories (high, medium, low), so that the mapping of CT skills would be more accurate.
(Adiyastuti, 2024)	Researchers measured students' computational thinking skills using AKM numeracy essay tests consisting of three essay questions based on personal and social contexts. In addition to written tests, researchers also used unstructured interviews to explore students' thinking processes and triangulated data through answer sheet documentation.
(Hauda,. et al 2024)	Researchers measured students' computational thinking skills using a written test consisting of two essay questions on exponential functions that contained four CT indicators, namely decomposition, pattern recognition, abstraction, and algorithms. The test was accompanied by interview guidelines to explore students' thought processes in greater depth and confirm their answers.
(Febriantoro et al., 2025)	Students' computational thinking skills were measured using an ISTE indicator-based computational thinking test, which covered four main components, namely problem formulation, data organization, automation, and generalization of solutions. In addition, the study also used a Bloom's Taxonomy-based cognitive test at levels C4–C6 to assess the aspects of analysis, evaluation, and creation, which are considered part of computational thinking skills.

Based on Tabel 2, it was found that there are various approaches to measuring students' CT abilities. Several studies use the Bebras Task as an assessment instrument. Mustaqimah (2024) used the Bebras Task test in the Penggalang (junior high school) category with three questions selected from the

2016–2018 period and modified as needed. A similar approach was also applied by Hartawan. et al (2024), who used an adapted Bebras Task essay test with clear CT indicators, namely decomposition, patterns and generalization, abstraction, and algorithmics.

The majority of studies developed essay test instruments designed based on computational thinking indicators. Research conducted by Mania (2021) used a performance-based written test that was systematically developed and validated using the Borg & Gall model. Mukhibin et al. (2024) also developed a descriptive test based on CT indicators with a focus on analyzing the quality and depth of students' computational thinking, particularly in the context of solving geometric mathematics problems. Hasanah & Masduki (2024) used a CT skills test covering four main indicators, namely decomposition, pattern recognition, abstraction, and algorithmic thinking. Meanwhile, Ahmadi et al. (2025) developed an essay test consisting of three mathematics questions on number patterns, and Hauda et al. (2024) used two essay questions on exponential functions with the same four CT indicators.

Contemporary studies show a trend toward using mixed assessment methods that combine written tests with interviews. Hartawan et al. (2024) supplemented the Bebras Task adaptation essay test with supporting interviews. Mukhibin. et al (2024) integrated qualitative interviews to assess students' thinking processes in greater depth. Susanti et al. (2025) used a combination of linear programming-based problem-solving tests, think-aloud exercises, and semi-structured interviews. Adiyastuti (2024) used AKM numeracy descriptive tests with unstructured interviews and data triangulation through answer sheet documentation, while Hauda et al. (2024) also supplemented written tests with interview guidelines to explore students' thinking processes.

Anwar et al. (2025) applied a performance-based assessment approach through direct observation of teacher and student activities, pre-tests and post-tests of programming comprehension, and CT assessment rubrics during Scratch project presentation sessions. Amelia et al. (2024) developed an Android-based mathematics evaluation instrument containing 15 essay questions contextualized with the Tabot Bengkulu culture and flat-sided spatial figures. Fitri et al. (2024) applied an innovative approach using the Tsukamoto Fuzzy algorithm to process the values of each CT indicator and determine the level of student ability, which was grouped into low, medium, and high categories. Febriantoro et al. (2025) developed a computational thinking test based on ISTE indicators covering four components, namely problem formulation, data organization, automation, and generalization of solutions, supplemented with a Bloom Taxonomy-based cognitive test at levels C4–C6 to assess the aspects of analysis, evaluation, and creation.

An analysis of 13 studies shows that computational thinking assessment methods in mathematics education in Indonesia have evolved from simple task-based approaches to more comprehensive and integrated assessment systems. The results of the analysis show a diversity of approaches in measuring students' computational thinking (CT) abilities in the context of mathematics learning in Indonesia. These findings are in line with global trends which show that there is no single consensus on the most effective CT assessment instruments. The diversity of assessment methods found in this study reflects the complexity and multidimensional nature of CT abilities themselves.

The findings show that several studies utilize the Bebras Task as an assessment instrument in line with the global trend of using Bebras as a tool to promote CT among children and adolescents. This study, validated by Dagiene (2022), shows that the Bebras Task is specifically designed to assess CT components such as abstraction, algorithmic thinking, decomposition, evaluation, and generalization. However, research by Zapata et al. (2024) in Education and Information Technologies emphasizes the importance of rigorous psychometric validation when adapting the Bebras Task to local contexts. They point out that although the Bebras Task has great potential as a CT assessment tool, adjustments need to be made taking into account the cognitive development level of students and the local cultural context. In the Indonesian context, research adapting the Bebras Task needs to report a more comprehensive validation process to ensure that the instrument is appropriate for the characteristics of Indonesian students.

One important finding from this study is the trend of using mixed assessment methods that combine written tests with interviews. A similar study by El-hamamsy et al. (2021) published in the Journal of Educational Computing Research also recommends the use of mixed methods in CT assessment, especially to overcome the limitations of single-method assessment. They found that interviews can help confirm and clarify students' responses on written tests, thus providing a more accurate picture of students' CT abilities. The practices carried out by Indonesian researchers demonstrate methodological maturity and awareness of the complexity of measuring CT abilities.

The use of programming platforms such as Scratch for CT assessment is in line with global research. Ye et al. (2023) in the International Journal of STEM Education found that programming-based project assessments can effectively measure CT while providing authentic learning experiences for students. However, they also warned that there is a risk of confusing CT with coding skills, so caution is needed in designing and interpreting programming-based assessment results.

A more unique approach is the use of fuzzy logic to classify students' CT ability levels. This finding is interesting because it contrasts with the conventional approach of using a fixed cut-off score. Research by Filiz (2025) published in Scientific Reports supports the use of adaptive neuro-fuzzy inference systems (ANFIS) to predict students' CT abilities. They found that the fuzzy logic approach can handle uncertainty and ambiguity in CT assessment better than traditional statistical methods. This shows that the innovations made by Indonesian researchers are at the forefront of global CT assessment methodology development.

One of the notable innovations from Indonesian research is the effort to integrate cultural context into CT assessment. This approach demonstrates an awareness of the importance of cultural relevance in the development of assessment instruments. This finding is in line with the arguments put forward by international researchers about the importance of cultural relevance in CT education. Leung

(2020) in the International Journal of STEM Education discusses the concept of "boundary crossing pedagogy," which emphasizes the importance of integrating local and cultural contexts into STEM learning, including CT. They argue that meaningful assessments must consider the social and cultural contexts of students in order to measure their abilities more accurately.

RQ2 : What are the advantages and limitations of the assessment methods used to evaluate computational thinking skills in mathematics learning?

Based on an analysis of various studies related to instruments for measuring computational thinking (CT) skills in mathematics learning, several important findings were discovered that can be categorized based on the type of instrument, strengths, limitations, and implications for measuring CT. The results of this study are presented based on the category of instrument used.

Bebras Task-Based Instrument

The Bebras Task instrument is one of the most widely adapted instruments in CT research in Indonesia. Mustaqimah (2024) found that the Bebras Task has a major advantage because it is specifically designed to comprehensively cover CT indicators, including decomposition, pattern recognition, abstraction, and algorithmics. This instrument does not require computer science knowledge, making it easy to apply in mathematics learning. However, the study also identified significant limitations, particularly in questions that require a high level of abstraction and algorithmic skills, causing many students to have difficulty understanding the context and determining the solution strategy.

Similar findings were also revealed by Hartawan et al. (2024) who adapted the Bebras Task in their research. This instrument is able to comprehensively measure all components of CT, and is reinforced by the use of in-depth interviews that help provide a clearer picture of students' difficulties. However, this study found that the Bebras instrument sometimes requires strong basic mathematical skills, such as understanding inequality signs or arithmetic operations, so that CT measurement results can be influenced by students' mathematical abilities, not purely their computational abilities. Other limitations include the use of only five questions, resulting in relatively limited measurement coverage, a low number of subjects (only 5 students), and a paper-and-text-based instrument, meaning that literacy skills could influence students' understanding of the question context.

The findings of the study on the Bebras Task instrument are in line with the results of a study Santaengracia et al. (2025) which found that Bebras tasks were able to comprehensively measure computational thinking (CT) abilities in 7th grade students with a sample of 1,513 students from 86 schools in Spain. The study confirmed that the Bebras Task did not show significant gender differences in overall performance, supporting the potential of Bebras as a fair assessment tool.

In line with the study Masiulionytė-dagienė & Jevsikova (2025), it provides a deeper perspective on the limitations of Bebras. Their findings show that the biggest challenge in Bebras is not solely computational logic, but rather students' ability to understand the context and interpret the task comprehensively. This confirms the findings of this study that Bebras questions that require high abstraction and algorithmic abilities often challenge students due to the complexity of understanding the context, not solely because of weaknesses in computational thinking.

Essay Test Instruments

Essay test instruments are a popular choice because of their ability to assess students' thinking processes in depth. Mania (2021) developed instruments designed in accordance with the curriculum, containing CT indicators (decomposition, pattern recognition, abstraction, algorithms) and equipped with attractive visual displays that make it easier for students to understand the questions. The use of the Borg & Gall development model makes the instrument more systematic and tested. However, this study found that students had difficulty with the pattern recognition and abstraction components due to the high demands of reasoning, so that the measurement results could be influenced by students' uneven thinking abilities. Another limitation is that the paper-based test format makes the scoring process less efficient than digital instruments.

Mukhibin et al. (2024) used a combination of descriptive tests, self-efficacy questionnaires, and interviews in their research. The descriptive test instrument contained four CT indicators so that it could assess students' abilities in greater depth, while the questionnaire helped classify students based on their level of self-efficacy, and the interviews strengthened the validity of the findings by clarifying the students' thinking processes. However, the main limitation of this study is the very small number of subjects, only three students, so the results cannot be generalized. The essay test is also potentially influenced by mathematical ability and understanding of lines and angles, so the CT measurement is not entirely pure. The use of manual instruments such as interviews and essays makes the analysis time-consuming and potentially subjective.

These findings indicate that essay-based instruments remain the dominant choice for measuring computational thinking (CT) abilities in mathematics learning. This dominance is in line with the trend in research in recent years, as reported by Atmatzidou & Demetriadis (2016), which states that essay instruments allow researchers to observe students' thinking processes more deeply than objective instruments. Essay tests are also considered capable of assessing CT indicators more comprehensively, especially decomposition, pattern recognition, and algorithms, as emphasized in a study (Sırakaya, 2020).

However, previous studies also show that essay tests have limitations, particularly in terms of scoring objectivity and analysis time. This is consistent with the trend identified by Weintrop et al. (2021), which emphasizes that traditional instruments require a lengthy correction process and are prone to subjectivity if not supported by a strong analytical rubric. In line with this, several recent studies, such as Tikva & Tambouris (2023) recommend the use of digital instruments or automated scoring to improve efficiency and reliability in CT assessment.

The difficulties students encounter in pattern recognition and abstraction indicators found in the reviewed studies are also consistent with global findings. Studies by Shute et al. (2021) show that abstraction is the most challenging component of CT because it requires high-level mental representation skills. Therefore, description-based instruments often capture large variations in ability among students, which affects score stability.

In addition, the limited sample size in several of the studies reviewed shows that many CT studies in the context of mathematics are still exploratory and qualitative in nature. This condition is in line with the international literature mapping by Chen et al. (2023), which found that CT assessment research in mathematics is still dominated by small-scale studies focusing on student thinking processes rather than the generalization of results.

Combined Test and Interview Instruments

Several studies use a combination of instruments to strengthen the validity and depth of data. Hasanah & Masduki (2024) used computational thinking tests, GEFT cognitive style tests, and interviews that were able to provide a more in-depth and triangulated picture of students' CT. The CT test adapted from Asesmenpedia has been validated by three experts and tested on students, so the instrument has good content validity. The use of GEFT is also an added value because it allows researchers to analyze CT abilities based on differences in field independent (FI) and field dependent (FD) cognitive styles. However, this instrument has limitations in that there are only three CT questions, so the coverage of CT indicators is relatively narrow. The selection of only five subjects for in-depth analysis limits generalization, and the essay test is greatly influenced by students' basic mathematical abilities.

Ahmadi et al. (2025) used an essay test based on CT indicators and interviews that were able to explore students' thinking processes in depth, reveal differences in abilities between high, medium, and low categories, and provide a detailed picture of how students understood the questions, found information, and drew conclusions. The combination of tests and interviews made the data richer and more valid for assessing CT abilities qualitatively. However, this method has limitations in the form of a limited sample size, the potential for students' answers to not be entirely authentic due to limited supervision during the test, and dependence on the quality of essay questions, which are sometimes difficult for students to understand.

Limitations related to the small number of subjects and the potential for subjectivity are in line with research arguments stating that portfolio and interview assessment formats are indeed difficult to standardize. As a result, it is difficult to provide evidence of validity and reliability. Furthermore, it is difficult to apply on a large scale and cannot be used in pre-post assessments. This contradicts the need for instruments that can be used widely and efficiently (Han, 2023).

Adiyastuti (2024) used AKM numeracy essay tests and interviews that were able to assess computational thinking skills in depth, especially in seeing how students describe information, recognize patterns, and compile solution steps. Interviews help confirm students' reasoning and thought processes so that the data is more accurate. However, this instrument has limitations in the form of questions that are quite complex, causing many students to fail to identify important information, as well as the long data collection time because the analysis of essay answers and interviews requires detailed interpretation.

Hauda et al (2024) also used a combination of written tests and interviews that were able to capture students' thinking processes in depth, especially on CT indicators such as decomposition, pattern recognition, and abstraction, which were clearly seen from the students' essay answers. The combination of tests and interviews strengthened the validity of the data because researchers could confirm the reasoning behind the students' solution steps. However, this instrument shows low student accuracy on algorithm indicators, indicating that the questions and interview process are not yet fully capable of helping students model exponential problems systematically. The use of a focus sample of only three students makes the instrument results less representative and difficult to generalize.

Susanti et al. (2025) used a more complex method by combining problem-solving tests, think-alouds, and interviews. This method provides the advantage of very rich data and allows researchers to assess the CT process in depth at each stage (decomposition, pattern recognition, abstraction, and algorithms). The combination of these three instruments improves the accuracy of interpretation because it allows for data triangulation. However, this method has significant limitations, namely that it is very time-consuming, the analysis is complex, and the results are prone to subjectivity because they depend on the researcher's interpretation. The small sample size of only four students limits generalizability, and the test instrument, which focuses on only one context (linear programming), results in a narrow scope of CT measurement.

The use of combined instruments, such as essay tests integrated with interviews, think-aloud protocols, or cognitive instruments e.g., GEFT, supports the view that assessing computational thinking requires a multimethod approach**.** Recent studies indicate that employing multiple assessment methods enables researchers to capture both the products of thinking (written responses) and the processes of thinking**,** including verbal reasoning and problem-solving strategies, which cannot be fully revealed through a single instrument (Pirzado et al., 2025). This triangulation strengthens the depth and validity of CT assessment by allowing a more comprehensive interpretation of students' cognitive processes (Liu & Li, 2020).

Despite providing depth, the combination of instruments also poses methodological challenges that are consistently reported in the literature. Several studies highlight limitations in generalization due to small samples and heavy analytical burdens (Scherer et al., 2019). Findings related to sample size limitations and subjectivity of assessment in the reviewed studies are consistent with this criticism; although data become more in-depth, evidence of reliability and generalizability is often weak. Therefore, CT research relying on qualitative designs needs to complement procedures with rigorous analysis rubrics and documented triangulation procedures to improve transparency and reproducibility (Curzon et al., 2020).

Instrument validity is also a major concern. The literature emphasizes the need for construct and criterion validation, as well as adequate testing, before instruments are used to draw broader conclusions (Román-González et al., 2018). Findings related to narrow item coverage, for example, only three CT items and complex questions, indicate the risk of insufficient CT indicator samples and ambiguity in the interpretation of answers. This is in line with academic recommendations that CT instruments cover adequate indicator domains and be systematically tested for validity so that the interpretation of scores truly reflects CT ability, not just mastery of mathematical content.

Furthermore, research on think-aloud protocols and interviews confirms the added value of both in revealing cognitive processes, but also highlights their vulnerability to reactivity bias and researcher interpretation (Grover & Pea, 2018). Therefore, when using think-aloud protocols or interviews, it is important to implement informant training procedures, transcription guidelines, and inter-rater triangulation techniques to reduce subjectivity. The literature also recommends combining qualitative approaches with standardized quantitative instruments when research objectives involve generalization or large-scale program evaluation (Ericsson & Simon, 1993; Creswell & Plano Clark, 2018).

The direction of development recommended by the literature is the integration of technology to improve the efficiency, objectivity, and scalability of CT assessment (Curzon et al., 2020). Automated assessment tools and digital platforms enable the collection of process traces (data logs) and automated assessment of certain aspects, thereby balancing the need for process depth and pragmatic needs in large classes. Thus, while a combination of qualitative instruments remains relevant for in-depth studies, incorporating digital approaches may be a middle ground for improving reliability and generalizability in the future.

Digital and Technology-Based Tools

Technological developments have encouraged the development of digital tools to measure CT. Amelia et al. (2024) developed an Android-based mathematics evaluation tool with the context of Tabot Bengkulu, which has been proven valid in terms of content, construct, and language, and is capable of measuring CT indicators such as decomposition, patterns, abstraction, and algorithms. The integration of cultural context and Android display makes this tool more attractive, accessible, and helps students understand problems more meaningfully. However, this tool has limitations in the form of several questions that are still difficult for students to understand due to unclear images and complex calculations, as well as the long development time required and dependence on the availability of digital devices in schools.

Anwar et al. (2025) used a different approach by applying Scratch as a learning medium that was evaluated through pre-post assessments, classroom observations, and evaluation questionnaires. This method was able to provide a comprehensive picture of the improvement in computational thinking in both teachers and students, with results showing an increase in teacher understanding of up to 90% and an increase in student engagement of up to 85%. The combination of needs surveys, practical workshops, and classroom implementation produced rich and in-depth data. However, this method has limitations because the results are highly dependent on technical conditions such as internet connection, computer equipment, and teacher readiness, so the findings do not fully reflect individual CT abilities. The measurement tools also measure aspects of engagement, technological understanding, and pedagogical competence rather than directly measuring students' CT abilities using specific tests. This study was only conducted in one MGMP, so the generalization of the results is still limited.

Febriantoro et al. (2025) used a cognitive test developed based on computational thinking indicators to assess students' ability to solve programming problems. This tool is capable of directly measuring aspects of computational thinking through problem-solving questions that are in line with the competencies required of students in project-based learning, as well as providing a quantitative picture of individual students' abilities. However, this tool has limitations because the assessment only focuses on the final results of the questions, so it does not comprehensively capture students' computational thinking processes. This tool also does not evaluate the thinking strategies or problem-solving steps used by students, so the representation of CT abilities is not fully comprehensive.

The findings of this study regarding the limitations of digital tools related to infrastructure availability and technical readiness are in line with a comprehensive systematic review study by Oliveira et al. (2025), which identified recurring challenges in the implementation of digital tools for CT development, including inadequate infrastructure, difficulties in using the tools, teacher training, adaptation of pedagogical practices, and measurement of students' CT skills. They emphasize that inequalities in technology access and teacher training affect the implementation and effectiveness of CT education, especially in resource-limited areas. They note that disagreement over the most effective CT teaching methods and assessment of CT skills creates significant challenges. Differences in digital infrastructure levels, cultural differences in educational priorities, and the need for local curriculum adaptation further complicate the global landscape of CT education. These findings reinforce the limitations identified in this study that digital tools require a long development time and depend on the availability of digital devices in schools.

Conclusion

Based on the results of an analysis of 13 studies, it was found that computational thinking (CT) assessment methods in mathematics education in Indonesia have developed into a variety of approaches, including the Bebras Task instrument, CT indicator-based descriptive tests, a combination of written tests and interviews, and technology-based digital instruments such as Android and Scratch applications. Each method has different advantages and limitations. The Bebras Task is able to comprehensively measure CT components without requiring computer science knowledge, but it often requires high levels of abstraction and algorithmic skills that are difficult for students. Essay tests can assess deep thinking processes but require longer analysis times and are susceptible to the influence of students' basic mathematical abilities. The combination of tests and interviews produces richer and more valid data but is difficult to standardize and has limitations in generalization due to the limited number of subjects, while digital instruments offer ease of access and cultural contextualization but are highly dependent on the availability of technological infrastructure and require a long development time. These findings indicate that there is no single most effective assessment method, so the selection of instruments must be tailored to the measurement objectives, learning context, and resource availability.

Recommendation

Future research is recommended to develop more comprehensive and standardized computational thinking (CT) assessment instruments that are specifically designed for mathematics education contexts. Further studies should focus on integrating qualitative and quantitative approaches in order to capture both students' problem-solving processes and learning outcomes more accurately. In addition, larger and more diverse research samples are needed so that the findings can be generalized across different educational levels and learning environments. The development of digital-based assessment tools that are adaptive, efficient, and culturally contextualized is also important to support the implementation of CT assessment in the era of digital learning. Future researchers are also encouraged to conduct rigorous validity and reliability testing to ensure that CT assessment instruments truly measure computational thinking skills rather than merely mathematical achievement or literacy skills.

Acknowledgments

The authors would like to express sincere gratitude to the lecturers of the Mathematics Education Department, Universitas Islam Negeri Sunan Kudus, for their guidance and support throughout the completion of this study. Appreciation is also extended to all researchers whose works contributed to this systematic literature review. The authors are grateful to colleagues and reviewers who provided valuable suggestions and feedback during the manuscript preparation process. Finally, the authors thank all parties who supported the completion of this research and writing process.

References

Adiyastuti, N. N., & Surakarta, U. M. (2024). Computational thinking analysis in solving elementary school AKM numeracy problems. *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran*, 121–137. https://doi.org/10.25273/pe.v14i2.21496

Ahmadi, D. L. A., Astuti, & Zulfah. (2025). Analisis Kemampuan Computational Thinking (CT) Peserta Didik Kelas VIII SMP Negeri 1 Bangkinang Kota. *Jurnal PEKA (Pendidikan Matematika)*, *8*(2), 246–255. https://doi.org/10.37150/jp.v8i2.3298

Akande, A., Cabral, P., & Casteleyn, S. (2020). Understanding the sharing economy and its implication on sustainability in smart cities. *Journal of Cleaner Production*, 277, 124077.

Allsop, Y. (2019). Assessing computational thinking process using a multiple evaluation approach. *International Journal of Child-Computer Interaction*. https://www.sciencedirect.com/science/article/pii/S2212868918300588

Amelia, L., Mulyono, B., Matematika, M. P., & Sriwijaya, U. (2024). Pengembangan Instrumen Evaluasi Matematika Berbasis Android Konteks Tabot Bengkulu untuk Mengukur

Computational Thinking Pengembangan Instrumen Evaluasi Matematika Berbasis Android Konteks Tabot Bengkulu untuk Mengukur Computational Thinking. *JRIP: Jurnal Riset dan Inovasi Pembelajaran 4*(3), 1893–1910. https://doi.org/10.51574/jrip.v4i3.2498

Anwar, K., Kamid, K., Kurniawan, W., Sofnidar, S., & Alrizal, A. (2025). Peningkatan Ketrampilan Berfikir Computational Thinking Dalam Pembelajaran Berbasis Programming-Matematika Menggunakan Menggunakan Scratch. *Jurnal JUPEMA*, *4*(1), 50–60. https://doi.org/10.22437/jupema.v4i1.38084

Astuti, A., Suryawati, E., Suanto, E., Yuanita, P., & Noviana, E. (2025). Charting a course : Exploring computational thinking skills in statistics content among junior high school students. *Journal of Pedagogical Research 9*(1), 182–202. https://doi.org/10.33902/JPR.202531653

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

Budiyanto, C. W., Wahyudi, M. N. A., Widiastuti, I., & Latifah, R. (2025). Teaching Computational Thinking in Indonesia: Trend in Conception, Methods, Technology, and Evaluation. *Educational Process: International Journal*, *17*, e2025308. https://doi.org/10.22521/edupij.2025.17.308

Chen, H. E., Sun, D., Hsu, T. C., Yang, Y., & Sun, J. (2023). Visualising trends in computational thinking research from 2012 to 2021: A bibliometric analysis. *Thinking Skills and Creativity*, *47*, 101224. https://doi.org/10.1016/j.tsc.2022.101224

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Curzon, P., Bell, T., Waite, J., & Dorling, M. (2019). Computational thinking. In S. A. Fincher & A. V. Robins (Eds.), The Cambridge Handbook of Computing Education Research (pp. 513-546). Cambridge University Press. https://doi.org/10.1017/9781108654555.018

Dagiene, V.; Dolgopolovas, V. Short Tasks for Scaffolding Computational Thinking by the Global Bebras Challenge. *Mathematics* 2022, *10*, 3194. https://doi.org/ 10.3390/math10173194

El-hamamsy, L., Zapata-cáceres, M., Barroso, E. M., Mondada, F., Zufferey, D., & Bruno, B. (2021). *The competent Computational Thinking test (cCTt): Development and validation of an unplugged Computational Thinking test for upper primary school*. *435*, 1–32. https://doi.org/10.1177/ToBeAssigned

Febriantoro, S., Fatharani, A., & Dewi, N. C. (2025). Assessing the Efficacy of Coding with Scratch and AI Interaction Using ChatGPT on 5th Graders ' Math Performance and Computational Thinking. *Reforma: Jurnal Pendidikan dan Pembelajaran 15*(1), 78–99. https://doi.org/10.30736/reforma.v15i1.1212

Filiz, A. (n.d.). *OPEN A neuro-fuzzy model for evaluating and predicting computational thinking skills of students*. *2025*, 1–13. https://doi.org/10.1038/s41598-025-20089-4 10

Fitri, Z., Safriana, Meiyanti, R., Faisal, M., & Mamat, R. 2024. Analysis of gender-based computational thinking skills through project base learning in programming using fuzzy method. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 12(1):138-154. https://dx.doi.org/10.24815/jpsi.v12i1.34798

Fitriyah, Y., Nurhayati, H., & Febrianti, T. S. (2024). Indonesian Students ' Computational Thinking Performance Based on Level and Gender. *International Journal of Pedagogy and Teacher Education 8*(1), 50–62. https://doi.org/10.20961/ijpte.v8i1.89464

Grover, S., & Pea, R. (2018). *Computational thinking: A competency whose time has come*. In S. Sentance, E. Barendsen, & C. Schulte (Eds.), *Computer science education: Perspectives on teaching and learning in school* (pp. 20–38). Bloomsbury Academic.

Han, J. (2023). A Systematic Review of Computational Thinking Assessment in the Context of 21 st Century Skills (Issue 2011). *Atlantis Press SARL*. https://doi.org/10.2991/978-2-38476-068-8

Hasanah, M. U., Masduki. (2024). Exploration Of Computational Thinking Skills Of Junior High School Student In Term Of FI And FD Cognitive Style. *Mathline: Jurnal Matematika dan Pendidikan Matematika*. 9(4), 1119-1138. http://doi.org/10.31943/mathline.v9i4.695

Hauda, Nabilah., Mulyono, B,Hapizah. (2024). Kemampuan Computational Thinking Materi Fungsi Eksponensial Menggunakan Problem Based Learning. *Jurnal Derivat*: *Jurnal Matematika dan Pendidikan Matematika 11*(1). https://doi.org/10.31316/jderivat.v11i1.6129

Kallia, M., Borkulo, S. P. Van, Drijvers, P., & Barendsen, E. (2021). Characterising computational thinking in mathematics education : a literature-informed Delphi study. *Research in Mathematics Education*, *0*(0), 1–29. https://doi.org/10.1080/14794802.2020.1852104

Khurshid, M. M., Zakaria, N. H., Rashid, A., Ahmad, M. N., Arfeen, M. I., & Faisal Shehzad, H. M. (2020). Modeling of open government data for public sector organizations using the potential theories and determinants a systematic review. *Informatics*, 7(3), 24.

Latifah, L., & Ritonga, I. (2020). Systematic Literature Review (SLR): Kompetensi Sumber Daya Insani Bagi Perkembangan Perbankan Syariah Di Indonesia. *Al Maal: Journal of Islamic Economics and Banking*, 2(1), 63-80.

Leung, A. (2020). *Boundary crossing pedagogy in STEM education*. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-020-00212-9

Liu, J., & Li, S. (2020). *Assessing students' computational thinking: A multimethod approach*. ***Journal of Educational Computing Research,*** ***58***(6), 1227–1254. https://doi.org/10.1177/0735633120905931

Majid. (2025). Tren Computational Thinking in Mathematics Education : A Global Bibliometric Analysis 2016-2025. *EDUKASIA: Jurnal Pendidikan dan Pembelajaran 6*, 303–314. https://doi.org/10.62775/edukasia.v6i1.1424

Mania, S. (2021). Pengembangan instrumen tes untuk mengukur kemampuan berpikir komputasi siswa. *JPMI – Jurnal Pembelajaran Matematika Inovatif 4*(1), 17–26. https://doi.org/10.22460/jpmi.v4i1.17-26

Masiulionytė-dagienė, V., & Jevsikova, T. (2025). Towards a process-oriented assessment of computational thinking : behavioural data and machine learning approach. *Interactive Learning Environments*, *4820*, 1–15. https://doi.org/10.1080/10494820.2025.2567601

Mukhibin, A., Herman, T., A, E. C. M., Arianti, D., & Utomo, S. (2024). Kemampuan Computational Thinking Siswa Pada Materi Garis Dan Sudut Ditinjau Dari Self-Efficacy. *JPMI – Jurnal Pembelajaran Matematika Inovatif*, *7*(1), 143–152. https://doi.org/10.22460/jpmi.v7i1.21239

Mustaqimah, U. P. S., & Ni'mah, K. (2024). Profil kemampuan berpikir komputasi siswa smp pada soal tantangan bebras. *JPMI – Jurnal Pembelajaran Matematika Inovatif 7*(2), 297–308. https://doi.org/10.22460/jpmi.v7i2.21590

Nyoman, I. G., Hartawan, Y., Putri, L. H., & Mahayukti, G. A. (2024). Junior High School Student ' s Computational Thinking Ability in Solving Mathematical Problems. *Jurnal Pedagogi dan Pembelajaran*, *7*(1), 124–133. https://doi.org/10.23887/jp2.v7i1.78001

Oliveira, R. M. D. E., Garbin, M. C., & Azevedo, R. (n.d.). *Global Overview of Computational Thinking and Digital Tools for Teaching*. 1–36.

Pei, C. (Yu), Weintrop, D., & Wilensky, U. (2018). Cultivating Computational Thinking Practices and Mathematical Habits of Mind in Lattice Land. *Mathematical Thinking and Learning*, *20*(1), 75–89. https://doi.org/10.1080/10986065.2018.1403543

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

Prasetyawan, E., Kusumah, Y. S., & Prabawanto, S. (2025). Students ' learning obstacles in systems of linear equations in two variables related to mathematical computational thinking skills. *Jurnal Gantang*, 10(1), 79 – 90. https://doi.org/10.31629/jg.v10i1.7168

Rianto, Budi., Kusnandi., Jupri, A. (2024). Computational Thinking of Indonesian Junior High School Students in Solving Geometry Problems. *Jurnal Pendidikan MIPA*. *25*, 300–312. http://dx.doi.org/10.23960/jpmipa/v25i1.pp300-312

Román-González, M., Pérez-González, J.-C., Moreno-León, J., & Robles, G. (2018). Can computational talent be detected? Predictive validity of the Computational Thinking Test. International Journal of Child-Computer Interaction, 18, 47-58. https://doi.org/10.1016/j.ijcci.2018.06.004

Román-González, M., Moreno-León, J., Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. *Computational ThinkingEducation*. https://doi.org/10.1007/978-981-13-6528-7_6
Santaengracia, J. J., Palop, B., García, T., Pérez, C. R., & Rodríguez-muñiz, L. J. (2025). Bebras-Based Assessment for Computational Thinking : Performance and Gender Analysis. *Education Sciences*, 15(7), 899. https://doi.org/10.3390/ educsci15070899
Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2019). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. Journal of Educational Psychology, 111(5), 764-792. https://doi.org/10.1037/edu0000314
Shumway, J. F., Welch, L. E., Kozlowski, J. S., Clarke-midura, J., Lee, V. R., Shumway, J. F., Welch, L. E., Kozlowski, J. S., Clarke-, J., & Lee, V. R. (2021). Kindergarten students ' mathematics knowledge at work : the mathematics for programming robot toys mathematics for programming robot toys. *Mathematical Thinking and Learning*. https://doi.org/10.1080/10986065.2021.1982666
Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, *22*, 142-158. https://doi.org/10.1016/j.edr.2017.09.003
Sirakaya, D.A. (2020). Investigating computational thinking skills based on different variables and determining the predictor variables. *Participatory Educational Research*. http://dx.doi.org/10.17275/per.20.22.7.2
Sunendar, Aep., Santika, S., Supratman., Nurkamilah, M. (2020). The Analysis of Mathematics Students ' Computational Thinking Ability at The Analysis of Mathematics Students ' Computational Thinking Ability at Universitas Siliwangi. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1477/4/042022
Susanti., E., Supiarmo, M.G., Turmudi., Harini, S. (2025). TRANSFORMATION OF THE COMPUTATIONAL THINKING PROCESS OF STUDENTS TO SOLVE MATHEMATICAL PROBLEMS THROUGH REFLECTION. *Matematika dan Pembelajaran 13*(1), 129–154. http://dx.doi.org/10.33477/mp.v13i1.8210
Tikva, C., & Tambouris, E. (2023). The effect of scaffolding programming games and attitudes towards programming on the development of Computational Thinking. *Education and Information Technologies*, *28*(6), 6845-6867. https://doi.org/10.1007/s10639-022-11465-y
Vaida Masiulionytė-Dagienė., Tatjana Jevsikova. (2025). Towards a process-oriented assessment of computational thinking: behavioural data and machine learning approach. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2025.2567601
Weintrop, D., Rutstein, D., Bienkowski, M., & McGee, S. (2021). Assessing computational thinking: An overview of the field. *Computer Science Education*, *31*(2), 113-116. https://doi.org/10.1080/08993408.2021.1918380
Wing, J. M. (2006). *Computational Thinking*. *49*(3), 33–35.
Ye, H., Liang, B., Ng, O. L., & Chai, C. S. (2023). Integration of computational thinking in K - 12 mathematics education : a systematic review on CT - based mathematics instruction and student learning. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-023-00396-w
Zapata, M., Pedro, C., Laila, M., & Hamamsy, E. (2024). A Bebras Computational Thinking (ABC - Thinking) program for primary school : Evaluation using the competent computational thinking test. *Education and Information Technologies*, *29*(12), 14969–14998. https://doi.org/10.1007/s10639-023-12441-w