Informatics in Education, 2025, Vol. 24, No. 1, 1–44

DOI: 10.15388/infedu.2025.02

Critical Thinking Assessment in K-12 Computing Education: A Systematic Mapping

Deise Monquelate ARNDT[1,2], Ramon Mayor MARTINS[2],
Jean Carlo Rossa HAUCK[1]
[1]*Graduate Program in Computer Science, Department of Informatics and Statistics, Federal University of Santa Catarina, Florianópolis/SC, Brazil*
[2]*Federal Institute of Santa Catarina, São José/SC, Brazil*
e-mail: {deise.arndt, ramon.mayor}@ifsc.edu.br, jean.hauck@ufsc.br

Received: June 2024

Abstract. Critical thinking is a fundamental skill for 21st-century citizens, and it should be promoted from elementary school and developed in computing education. However, assessing the development of critical thinking in educational contexts presents unique challenges. In this study, a systematic mapping was carried out to investigate how to assess the development of critical thinking, or some of its skills, in K-12 computing teaching. The results indicate that primary studies on the development of critical thinking in K-12 computing education are concentrated in Asian countries, mainly focusing on teaching concepts such as algorithms and programming. Moreover, the studies do not present a fixed set of critical thinking skills assessed, and the skills are selected according to specific teaching and research needs. Most of the studies adopted student self-assessment using instruments that are well-known in the literature for assessing critical thinking. Many studies measured the quality of instruments for their research, obtaining favorable results and demonstrating consistency. However, the research points to a need for more diversity in assessment methods beyond student self-assessment. The findings suggest a need for more comprehensive and diverse critical thinking assessments in K-12 computing education, covering different educational stages and computing education concepts. This research aims to guide educators and researchers in developing more effective critical thinking assessments for K-12 computing education.

Keywords: Computing Education, Critical Thinking, Assessment, K-12.

1. Introduction

Computing is fundamental in shaping our technology-driven future, thus it is essential to teach computing to students (U.S. Department of Education, 2019). Learning computing also helps students to develop computational thinking, including critical thinking, creativity, problem-solving, and collaboration (Lin & Chen, 2020). Within the broad spectrum of essential 21st-century competencies, critical thinking is fundamental in computing, across knowledge domains, and everyday life (Sari *et al.*, 2022; World Economic Forum, 2020). Although there is no consensus, Facione (1990) defines critical thinking as the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from or generated by observation, experience, reflection, reasoning, or communication, serving as a guide for beliefs and actions.

Recognizing the importance of critical thinking, it has also become a goal of K-12 education to develop critical thinkers (UNICEF, 2023; OECD, 2019), helping students to develop higher-order thinking skills (e.g., to analyze, evaluate, and solve complex problems) to enable them to think effectively and rationally (Spector & Ma, 2019; Saadé *et al.*, 2012). This aim extends beyond mastering essential subject matter, as it seeks to shape citizens who can reason ethically and act for the public good (Elder & Paul, 2006;) and apply learned skills to real-life problems (Shafiyeva, 2021). Furthermore, proficiency in critical thinking, linked to reflective thinking and skillful judgment, is acknowledged as a key to success in higher education and is considered a key skill for future leaders (OECD, 2019; Hussein *et al.*, 2019).

Some initiatives aim to promote and develop critical thinking skills in K-12 education, each with a unique approach. “The Foundation for Critical Thinking” customizes webinars and courses, focusing on the disciplined process of conceptualizing, applying, analyzing, synthesizing, and evaluating information (CriticalThinking.org, 2019). The “Insight Assessment” company provides research-based tools for assessing critical thinking and reasoning skills, which are used globally by employers and educators to develop these fundamental skills (Insight Assessment, 2023). The “Instituto Ayrton Senna” guides educators in Brazil to foster creativity and critical thinking, focusing on holistic human development and creating evidence-based educational policies and practices (InstitutoAyrtonSenna.org, 2022).

As an alternative, critical thinking skills can also be developed as part of computing education (Huang and Qiao, 2022; Voskoglou and Buckley, 2012), enabling students to understand and navigate the challenges and opportunities presented by rapidly advancing technology and its applications in various fields, such as Artificial Intelligence (AI) (Lee *et al.*, 2023; Ten Haken, 2017). Furthermore, in a society where social media is a prevalent source of information and fake news is a growing concern, the acquisition of critical thinking skills becomes an essential competency, to discern the reliability of information, thereby equipping young people to navigate the digital landscape and make informed decisions effectively (Cortazar *et al.*, 2021). And, especially when interacting with artificial intelligence (AI) technologies, developing critical thinking skills is essential for understanding and analyzing AI outputs, assessing the technology’s ethical, biases and privacy implications, guiding them in making responsible and informed decisions about its use, its role in society, and its potential impact on their lives (Lee *et al.*, 2023; UNICEF, 2023a; 2023).

Critical thinking is recognized as a fundamental skill in the contemporary educational landscape as part of computing education (UNICEF, 2023a; OECD, 2019). Various frameworks guide the integration of computing and critical thinking into K-12 curricula globally. In the United States, the "K-12 Computer Science Framework" (K12CS) suggests that all students should be capable of learning basic computer science concepts and that understanding these fundamentals is key to developing critical thinking skills (K12CS.org, 2016). In Europe, the 'Informatics Reference Framework for School' by Informatics for All provides comprehensive guidance for integrating informatics education across different educational systems (Caspersen *et al.*, 2022). Other frameworks like "OECD Learning Framework 2030" (Vincent-Lancrin *et al.*, 2019), and the "Computational Thinking for Science framework" (CT-S) (Hurt *et al.*, 2023) also emphasize critical thinking development within computing education. Critical thinking is mostly stimulated by adopting active learning methodologies, such as problem-based, project-based, and task-based teaching. These approaches encourage students to engage in authentic, meaningful learning experiences that require them to apply critical thinking skills to solve problems (Mäkiö and Mäkiö, 2023; Rehmat and Hartley, 2020; Anizifa and Djukri, 2017).

Therefore, it is essential that educators help students develop critical thinking, and are also able to assess the development of this skill to guide the student learning process and identify opportunities for improvement (Paul *et al.*, 2023; Cortázar *et al.*, 2021).

Specifically for the assessment of critical thinking, it is necessary to define appropriate assessment methods that are well integrated into existing curricula, in order to provide effective feedback to students and teachers (Cortázar *et al.*, 2021; Saadé *et al.*, 2012). Recognizing the importance of developing critical thinking in K-12, some research has explored teaching this skill in K-12 computing education, but they do not specifically focus on the assessment methods used to evaluate critical thinking skills. Lee and Nuatomue (2022) primarily reviewed how computer science teaching was implemented in schools and its effectiveness in developing computational thinking, including critical thinking. Aktoprak and Hursen (2022) carried out a bibliometric analysis of research on critical thinking in primary education, identifying trends, without specific emphasis on assessment in computing education. Popat and Starkey (2019) reviewed research to analyze the educational outcomes of children learning to program, including critical thinking skills, but did not delve into assessment methods.

While these studies provide important findings, there remains a gap in the literature regarding a comprehensive review of assessment approaches for critical thinking specifically within K-12 computing education.

To address this gap, we conducted a systematic mapping of the literature focused on the research question (RQ): Which studies exist to assess critical thinking, or some of its skills, in K-12 computing education? The main contributions of this research include identifying existing studies on critical thinking assessment in K-12 computing education, analyzing critical thinking definitions and skills assessed, reviewing assessment methods used, and evaluating the quality of these assessment approaches. The results of this systematic mapping are expected to guide educators in applying critical thinking assessments and help researchers create effective critical thinking assessments in K-12 computing education.

2. Background

2.1. *Critical Thinking*

Critical thinking is considered one of the essential skills of the 21st century in the context of Learning & Innovation Skills, forming the “four C’s” along with communication, collaboration, and creativity (P21.org, 2019). Critical thinking can be defined from diverse points of view, such as philosophy, psychology, and education (Spector and Ma, 2019).

From a philosophical point of view, Dewey (1933) defined critical thinking as “active, persistent, and careful consideration of a belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends”. In psychology, critical thinking is typically defined as a higher-order type of reasoning that involves a repertoire of faculties, such as articulation of arguments, evaluation of evidence, and correction of one’s activity and progress towards an established goal (Halpern, 1998). From an educational point of view, critical thinking is commonly considered “the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from or generated by observation, experience, reflection, reasoning, or communication, as a guide to belief and action” (CriticalThinking.org, 2019; Facione, 1990). Within this context, “skills” refer to the learned techniques and methods for performing tasks effectively, such as analyzing arguments or synthesizing information, while “abili-

Table 1
Core skills of critical thinking according to Delphi Report

Skill	Brief explanation
Interpretation	Understanding and expressing the meaning or significance of various forms of information, including experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures, or criteria.
Analysis	Identifying the intended and actual inferential relationships among statements, questions, concepts, descriptions, or other forms of representation intended to express belief, judgment, experiences, reasons, information, or opinions.
Evaluation	Assessing the credibility of statements or descriptions of a person’s experience, judgment, belief, or opinion, and assessing the logical strength of the actual or intended inferential relationships among statements, descriptions, questions, or other forms of representation.
Inference	Drawing reasonable conclusions from information, including predicting the future, hypothesizing about the past, and drawing conclusions from data.
Explanation	Stating the results of one’s reasoning, justifying that reasoning based on evidential, conceptual, methodological, criteriological, and contextual considerations, and presenting one’s reasoning in the form of cogent arguments.
Self-Regulation	Monitoring and evaluating one’s own cognitive activities, the elements used in those activities, and the results obtained, mainly by applying skills in analysis and evaluation to one’s inferential judgments to question, confirm, validate, or correct either one’s reasoning or one’s result.

Table 2
Additional skills of critical thinking (Yeh, 2003)

Additional skills	Brief explanation
Recognition of assumptions	Identifying statements or claims implicit in general premises.
Induction	Inferring the most likely outcome from known facts.
Deduction	Using reason to draw a necessary conclusion from two given premises.

ties" are the innate or acquired capacity to perform these tasks, such as reasoning or problem-solving (CriticalThinking.org, 2019). Together, they encompass the range of competencies that critical thinking entails. Specifically, the Delphi Report (Facione, 1990) presents a consensus set of cognitive skills that constitute a core of critical thinking (Table 1).

In addition to these core critical thinking skills, other skills are also considered, such as the additional skills presented by Yeh (2003) (Table 2).

The critical thinking skills that comprise the core skills (Facione, 1990) and the additional skills (Yeh, 2003) presented in Tables 1 and 2 are complementary but distinct. Although there are conceptual overlaps, each skill has its own nuances. For example, 'Inference' (Facione, 1990) is a broader concept that includes both 'Induction' and 'Deduction' (Yeh, 2003). 'Recognition of assumptions' (Yeh, 2003) can be considered a specific aspect of 'Analysis' (Facione, 1990).

2.2. *Assessment of Critical Thinking*

An important aspect of promoting critical thinking is its evaluation. The assessment aims to provide valuable feedback to the students on developing their critical thinking skills, helping them identify areas of strength and improvement, thereby facilitating their learning process and personal development (Pedrosa-de-Jesus and Guerra, 2018). For educators, understanding students' cognitive abilities, including their capacity to analyze, perceive, and empathize, can guide them to develop and/or adopt teaching methods to suit students' needs better and identify gaps in their understanding (Vincent-Lancrin, 2023; Criticalthinking.org, 2019).

Assessment paradigms. Critical thinking assessments can be broadly categorized into three main paradigms: summative, formative, and self-assessment (Brookhart, 2010; Popham, 2008). Summative assessments evaluate learning outcomes at the end of an instructional unit application. Formative assessments provide ongoing feedback during the learning process. Self-assessment involves students evaluating their own progress.

Assessment methods. Several assessment methods have been used to assess students' learning, including developing critical thinking (Soland *et al.*, 2013). So far, there has yet to be a consensus on the definition of the best method to assess students' learning (Anders *et al.*, 2019). Each method is designed to evaluate different aspects (Soland *et al.*, 2013) using various types of data collection (Hattie & Timperley, 2007) (Table 3).

Table 3
Examples of methods for the assessment of students' learning

Assessment Method	Description	Examples of data collection instruments
Performance-based assessment	A method that requires students to demonstrate or apply their skills by creating a response, product, or performing a task. It assesses students' skills to apply what they have learned in authentic or real-world contexts, thus measuring higher-order thinking skills such as analysis, synthesis, and evaluation (Braun *et al.*, 2020; N.Y. State Education Department, 2024).	• Project documentation • Artifact (e.g., applications, code, machine learning model, documentation)
Student self-assessment	A method that students analyze their own learning progress to understand their perception of learning, attitudes, and beliefs. This encourages them to actively participate in their education (Andrade, 2019). However, due to the characteristics of the self-assessor, it is reported to have limitations in terms of validity and accuracy (Taylor, 2014).	• Self-assessment questionnaire (Multiple-choice, Likert-scale, etc.)
Observation	A method assesses transformations in behavior, performance, interactions, and other aspects by observing the learners (Allen *et al.*, 2011).	• Checklist • Observer annotations
Interview	A method used for feedback, understanding students' learning thought processes and applying critical thinking. Interviews can be structured, semi-structured, or unstructured, each facilitating different levels of insight (Creswell *et al.*, 2018)	• Interview script • Interview notes
Test	A method used to evaluate the student's acquisition of knowledge. Tests are used to measure the progress and impact of educational intervention and performance (Morrison, *et al.*, 2019).	• Test with Multiple-choice, open-ended questions, etc.

Some instruments are widely used in the literature to assess critical thinking, for example, the California Critical Thinking Skills Test (CCTST) (Facione, 1990), Test of Everyday Reasoning (TER) (Facione *et al.*, 2012), Cornell Critical Thinking Test (CCTT) Level X (Ennis *et al.*, 2005) and the Computer Thinking Skill Level - Secondary school (CTLS) (Korkmaz, Çakÿr, Özden, 2015). These instruments are generally available commercially, developed with a focus on reliability, ensuring consistent scores, validity, and accuracy of the assessment (Criteriacorp, 2023; Insight Assessment, 2023; CriticalThinking.org, 2019). Such assessments are utilized in diverse scenarios, including job selection, professional training (Criteriacorp, 2023), school and university admissions (Insight Assessment, 2023), and specifically the assessing of students' critical thinking abilities (Reynders *et al.*, 2020; CriticalThinking.org, 2019). Table 4 summarizes some of the main instruments for assessing critical thinking in the educational context.

Effectiveness of assessments. To ensure the effectiveness of assessments, it is important to evaluate their quality in terms of the reliability and validity of the instruments (Moskal and Leydens, 2000; Morrison *et al.*, 2019) (Table 5).

Table 4

Summary of the main critical thinking assessment instruments

Reference	Critical thinking assessment instruments	General information			Characteristics of the test					Findings	
		Purpose	Educational stage	Typical application context	Idiom	Skills	Items	Scores	Administration	Reliability	Validity
(Facione, 1990)	California Critical Thinking Skills Test (CCTST)	To measure the core reasoning skills necessary for reflective decision-making.	Undergraduate and graduate students, executive-level adults	Accreditation, admissions, advising and retention, curriculum effectiveness, and the documentation of student learning outcomes.	English	Analysis, evaluation, explanation, inference e interpretation	34-item multiple-choice	100-point scoring scale to interpret overall results, accompanied by a corresponding qualitative rating	55 minutes timed administration	A KR20 coefficient between .69-.68, pre- and post-test paw, indicates medium reliability	The CCTST has shown criterion validity through factor analysis and internal consistency. The test scores have also demonstrated predictive validity, explaining a significant proportion of variance in outcome measures*§
(Facione *et al.*, 2012)	Test of Everyday Reasoning (TER)	To measure the core reasoning skills necessary for reflective decision-making.	Late childhood to adulthood	Educational assessments, employee selection, and program evaluations.	English	Analysis, deduction, evaluation, explanation, inference, induction e interpretation	35 item multiple-choice	Dichotomously scored	50 minutes timed administration	A KR-20 coefficient (for instruments with dichotomously scored items) ranges from .71 to .86, indicating good reliability	The TER has demonstrated validity in measuring the critical thinking skills of individuals and groups. The exact validity value may vary depending on the specific version of the TER and the population being tested*§
(Ennis *et al.*, 2005)	Cornell Critical Thinking Test (CCTT) Level X	To measure critical thinking abilities, including analyzing and evaluating information, making logical inferences, and drawing accurate conclusions.	Students in Grades 5-12+	Educational assessment to predict students' performance on state proficiency exams, for honors/AP programs, college admissions, careers, and employment.	English	Induction, deduction	71 item, multiple-choice	The score includes the candidate's raw score, standardized score, and percentage position. The raw score represents the number of correct answers	50 minutes timed administration or untimed	Cronbach's Alpha = .77 for the total scores on the CCTT Level X indicates good reliability	Using Confirmatory Factor Analysis (CFA) the test is valid in measuring the constructs it is intended to measure*§
(Korkmaz, Çakÿr, Özden, 2015)	Computer Thinking Skill Level - Secondary school (CTLS)	To measure critical thinking computational thinking skills, algorithmic thinking, problem-solving, and cooperative learning.	Secondary school students	Educational assessment	Turkish	Analysis, evaluation, inference, explanation, self-regulation	22 items multiple-choice	Score derived from 5-point Likert scale		Cronbach's alpha = .82 for total score indicates good reliability	Confirmatory Factor Analysis ($\chi 2$ =4481.1628 (d=195, N=241, p<.01), RMSEA = .074, SRMR = .078, GFI = .89, AGFI = .84, CFI = .91 and IFI = .90). The values indicate a good fit of the model, with the RMSEA and SRMR within acceptable ranges and the GFI, AGFI, CFI and IFI indices showing a good fit of the model. The confirmatory factor analysis therefore suggests that the instrument has adequate construct validity

*Exact reliability and validity values may vary depending on the specific version of the test and the population being tested.

§Exact values are not informed.

Table 5

Reliability and validity of instruments to assess critical thinking

Evaluation aspects	Definitions	Considerations	Common Analysis Method	Used	Brief explanation
Reliability	Refers to the consistency of scores from an assessment instrument (Moskal and Leydens, 2000).	Homogeneity, stability, and equivalence are considered (Heale and Twycross, 2015).	● Cronbach's alpha coefficient (Cronbach, 1951);	Used as an indicator of the internal consistency of an evaluation instrument.	● Higher values (close to 1.0) indicate greater internal consistency, suggesting that the items in the instrument measure the same concept. Values of α. are considered as a rule of thumb according to Gliem and Gliem (2003): ○ $\alpha \geq .9$ is considered Excellent; $9 > \alpha \geq .8$ is considered Good; $.8 > \alpha \geq .7$ considered Acceptable $.7 > \alpha \geq .6$ considered Questionable; $.6 > \alpha \geq .5$ considered Poor; $.5 \geq \alpha$ considered Unacceptable
			● Cohen's kappa coefficient (Cohen, 1960).	A statistical measure of agreement between two raters, used to assess inter-rater reliability when they rate items in mutually exclusive categories.	● Can range from -1 to +1, where 0 represents the amount of agreement expected from random chance, and 1 represents perfect agreement between the raters.
			● KR-20 (Kuder-Richardson Formula 20) coefficient (Kuder and Richardson, 1937).	Measure the homogeneity of test items for items scored by binary choice ("true" or "false" answers).	● A high value indicates that the items consistently measure the same construct or skill.
			● Omega coefficient (McDonald's omega) (McDonald, 1999).	Used to estimate the proportion of variance in test scores that can be attributed to the overall construct being measured, offering a more nuanced assessment of internal consistency.	● Values close to 1.0 indicate a robust underlying construct with good internal consistency.
Validity	Refers to the process of accumulating evidence that supports the appropriateness of conclusions made based on student responses for specific assessment purposes. The accuracy of the measurement instrument (Moskal and Leydens, 2000).	It relates to content, construct, and/or criterion validity.	● Correlation Matrix (DeVellis, 2017).	Used to examine the relationship between variables.	● Correlation coefficients between variables.
			● Factor Analysis (Glorfeld, 1995) supported by Kaiser-Meyer-Olkin (KMO) (Kaiser and Rice, 1974).	●KMO is a measure of sample adequacy used to assess whether a data set is suitable for factor analysis. ●Factor Analysis is used to identify underlying structures (factors) in a set of variables. It is often used to reduce the dimensionality of data.	● A KMO value higher than .50 indicates that factor analysis can reasonably proceed. ● Factor analysis: A factor loading of more than .30 usually indicates a moderate correlation between the item and the factor. (Tavakol and Wetzel, 2020).
			● Item Response Theory (DeVellis, 2017)	A set of models used to analyze the relationship between individuals' latent abilities or traits and their responses to test items.	● There are three primary parameters: item difficulty (b), item discrimination (a), and the pseudo-guessing parameter (c). The item difficulty parameter (b) represents the latent trait level at which examinees are expected to have a 50% probability of answering a test item correctly, assuming no guessing (DeVellis, 2017).
			● Bartlett's test of sphericity (Snedecor *et al.*, 1989).	Used for correlation matrix suitability.	●Low significance values (*p*-value) (usually below .05) indicate that the correlation matrix variables are sufficiently correlated.

	● Confirmatory Factor Analysis (CFA) (Jöreskog,1969).	Used to verify the factor structure of a set of observed variables. Test the hypothesis that a relationship between observed variables and their underlying latent constructs exists.	Uses metrics are, according with Hu and Bentler (1999): ● Chi-square ($\chi 2$): A low $\chi 2$ value suggests a good fit of the model. High values may indicate a mismatch between the model and the observed data. ○ Cutoff criteria for $\chi 2$: < 3 good; < 5 acceptable ● Comparative Fit Index (CFI): Compares the model's fit to a null model, with values closer to 1 indicating a better fit. ● Goodness of Fit Index (GFI): Reflects the proportion of variation the model explains. ● Adjusted Goodness of Fit Index (AGFI): Adjusts the GFI for the number of parameters in the model, with values closer to 1 indicating a better fit ● Incremental Fit Index (IFI): Also known as Bollen's IFI, similar to CFI but less affected by sample size, with values closer to 1 indicating a better fit. ● Tucker-Lewis Index (TFI): Also known as the non-normed fit index, compares the fit of the model to a null model, adjusting for model complexity ● Normed Fit Index (NFI): Assesses the model by comparing the $\chi 2$ value of the model to the $\chi 2$ of the null model, with values closer to 1 indicating a better fit. ○ Cutoff criteria for fit CFI, GFI, AGFI, IFI, TFI, NFI: .95 excellent; .90 good; .80 acceptable ●Standardized Root Mean Square Residual (SRMR): The standardized difference between the observed and predicted correlations ○ Cutoff criteria for fit SRMR: < .08 good fit ● Root Mean Square Error of Approximation (RMSEA): Measures the discrepancy per degree of freedom ○ Cutoff criteria for fit RMSEA: < .05 excellent; .05-.08 good; .08-.10 moderate; .10 poor

2.3. *Critical Thinking in K-12 Computing Education*

Critical thinking is an essential skill in K-12 computing education, involving the ability to analyze, evaluate, and synthesize information to make decisions. It can help students develop problem-solving capabilities, foster innovation, and facilitate effective decision-making to address technological issues (Huang and Qiao, 2022; Voskoglou and Buckley, 2012). This skill applies to various computing concepts, such as

Algorithms, logic, and programming. Critical thinking helps students to develop efficient and effective algorithms and improve their ability to understand and solve algorithmic problems, logic, and analytical thinking skills (İlic, 2021; Velázquez-Iturbide, 2013; Fagin *et al.*, 2006). Critical thinking skills can also be fostered through programming languages like block-based programming, Scratch and Alice, and text-based programming, Python (Create-Learn, 2023; İlic, 2021; Sontag, 2009).

Information literacy. Critical thinking can be instrumental in discerning the integrity of information found on social media and combating fake news. It involves rational thinking, considering evidence, and seeking additional sources (Cortazar *et al.*, 2021).

STEM integration. Another key aspect of K-12 computing education is promoting interdisciplinary learning and integrating STEM subjects, which, for example, helps with mathematical skills and stimulates problem-solving and critical thinking (Karaahmetoğlu and Korkmaz, 2019).

Robotics. Critical thinking is developed through systematic problem-solving as students analyze robotic systems e.g. sensor data, mechanical systems, and evaluate hardware-software interactions, to make reasoned decisions when programming devices like Arduino or Raspberry Pi to interact with the physical world (Karaahmetoğlu and Korkmaz, 2019).

Artificial Intelligence. As AI technologies become increasingly integrated into daily lives, students need to understand the ethical implications of AI, including issues of fairness, bias, and privacy. This understanding can help students become responsible digital citizens and make informed decisions about AI technologies (UNICEF, 2023; Lee *et al.*, 2023; Martins *et al.*, 2024a; 2024b).

Several frameworks guide the integration of computing into the K-12 curriculum, each emphasizing critical thinking. The “K12C framework” (K12CS.org,2016) aims to make computer science education accessible to all students in the U.S. The ‘Informatics Reference Framework for School’ by Informatics for All provides a European perspective on integrating informatics education (Caspersen *et al.*, 2022). The “OECD Learning Framework 2030” (Vicente-Lancrin, S. *et al.*, 2019) seeks to foster creativity and critical thinking in primary and secondary education globally, while the “CT-S framework” (Hurt *et al.*, 2023) applies computational thinking as both an input and outcome of science learning. All these frameworks incorporate critical thinking by encouraging students to engage innovatively with issues and problems, fostering problem-solving skills and resilience.

Table 6
Overview of common pedagogical approaches to promote critical thinking in K-12 computing education

Pedagogical Approaches	Description
Discussion	This method aims to help students articulate opinions, assess class arguments and evidence, and revise their positions based on discussion insights (Taylor, 2022).
Inquiry-based teaching	A student-centered approach driven by students' questions and their innate curiosity. It engages students in active learning by exploring topics, asking questions, and discovering answers through critical thinking (Gholam, 2019).
Problem-based teaching	This method stimulates problem-solving by requiring the active application of critical thinking skills (Rehmat and Hartley, 2020).
Project-based teaching	A student-centered approach that encourages critical thinking through active exploration of real-world challenges and problems (Anizifa and Djukri, 2017).
Socratic questioning	A method of inquiry using leading questions to stimulate rational thinking and logical responses, promoting critical thinking (Liu, 2019).
Task-based teaching	A method that operates at the module level and is based on the principle of perceptual learning, it stimulates the process of finding solutions to problems, requiring the active application of critical thinking skills (Mäkiö and Mäkiö, 2023).

To promote critical thinking skills in the classroom, various pedagogical approaches are adopted, emphasizing active learning and problem-solving activities, encouraging questioning and reflection, and fostering a supportive learning environment (Insight Assessment, 2023; Taylor, 2022; Rehmat and Hartley, 2020; Liu, 2019; Anazifa and Djukri, 2017) (Table 6).

3. Definition and Execution of the Systematic Mapping

To elicit the state-of-the-art approaches for assessing critical thinking (or any of its skills) in the context of computing education in K-12, a systematic mapping was performed following the procedure defined by Petersen *et al.* (2008). Starting with defining research and analysis questions that adhere to the study's objectives and delineate the research scope, a review protocol was defined, specifying the sources, search strings, and selection criteria. Following the review protocol, searches were executed, and relevant results were selected based on the pre-established inclusion, exclusion, and quality criteria. The eligibility of studies was determined by their adherence to these criteria. After identifying relevant articles, information related to the analysis questions was extracted, following the defined extraction strategy. The softwares Zotero was used to manage the selected articles, while Google Spreadsheet was employed to organize and analyze the extracted data. The extracted data was then analyzed, interpreted, and discussed.

3.1. *Considerations of the Research Scope*

This systematic mapping examines critical thinking assessment approaches and specific methods used in K-12 computing education. The assessments are analyzed through various paradigms, including summative, formative, and self-assessment methods. This review explores diverse assessment methods in the literature and seeks to identify various assessment approaches.

3.2. *Definition of the Review Protocol*

The research question is:

- **RQ**. Which studies exist to assess critical thinking, or some of its skills, in K-12 computing education?

The research question was refined into the following analysis questions:

- **AQ1**. What existing studies include assessing the development of critical thinking in the context of K-12 computing education?
- **AQ2**. How is critical thinking defined in the studies, and what skills are being assessed?
- **AQ3**. How are these critical thinking skills assessed?
- **AQ4.** How has the assessment approach been evaluated?

Data sources. Searches were performed on the main digital libraries and repositories in computing, including the ACM Digital Library, arXiv, ERIC (U.S. Dept. of Education), IEEE Xplore, Scopus, ScienceDirect, SocArXiv, SpringerLink, and Wiley Online Library, accessible via Portal Capes[1]. Searches were also conducted on Google Scholar and Google to ensure a comprehensive search and reduce the risk of omission (Piasecki *et al.*, 2018).

Inclusion and exclusion criteria. As part of this mapping, artifacts that present the application or development of an assessment of critical thinking as part of teaching computing in K-12 were considered following the inclusion/exclusion criteria presented in Table 7.

Quality criteria. Only primary studies that present substantial information regarding the analysis questions were considered. Abstract-only or one-page articles were excluded.

Definition of the search strings. Following the research objective, the search string was defined by identifying core concepts and considering synonyms, as indicated in Table 8.

[1] A web portal for access to scientific knowledge worldwide, managed by the Brazilian Ministry of Education for authorized institutions, including universities, government agencies, and private companies (`www.periodicos.capes.gov.br`).

Table 7

Inclusion and exclusion criteria

	Inclusion	Exclusion
Focus	Assessment of critical thinking in computing education	Assessment of only other skills (e.g., creativity, learning, analytical thinking, problem-solving, etc.)
Context	Computing education	Other areas (e.g., psychology, medicine, etc.)
Content	Application or development of the assessments of critical thinking skills	No application nor development of a critical thinking assessment
Educational stage	K-12	Other educational stages (e.g., higher education) or teacher training
Publication language	English	Other languages, e.g., Chinese, Spanish, Portuguese, etc.
Type of publication	Scientific articles in journals, conferences, well-known online repositories, and academic works (e.g., dissertations, theses, etc.)	Website articles, blogs, videos, and other systematic reviews/mappings

Table 8

Main concepts and synonyms

Main concepts	Synonyms
Critical thinking	
Assessment	measur*, evaluat*, analy*
School	K-12, learn*, teach*, teen*, course
Computing	coding, programming, computer science, computational thinking

The selection of the search string was carefully calibrated through several preliminary searches to reduce the risk of omission of relevant research.

Considering the main concepts, a generic search query was formulated using Boolean operators and wildcard symbols to capture variants of the terms:

```
("critical thinking") AND (assess* OR measur* OR evaluat* OR analy*) AND (k-12 OR school OR learn OR teach OR course OR teen) AND (computing OR coding OR programming OR "computational thinking" OR "computer science").
```

This query was then adapted for the specific syntax of each data repository, as detailed in Table 9.

Table 9

Search string per data source

Source	Search string
ACM Digital Library	[Abstract: "critical thinking"] AND [[Abstract: assess*] OR [Abstract: measur*] OR [Abstract: evaluat*] OR [Abstract: analy*]] AND [[Abstract: k-12] OR [Abstract: school*] OR [Abstract: learn*] OR [Abstract: teach*] OR [Abstract: course] OR [Abstract: teen*]] AND [[Abstract: computing] OR [Abstract: coding] OR [Abstract: programming] OR [Abstract: "computational thinking"] OR [Abstract: "computer science"]]
arXiv	classification: Computer Science (cs); include_cross_list: True; terms: AND abstract=critical thinking; AND abstract=assess* OR measur* OR evaluat* OR analy*; AND abstract=assess* OR measur* OR evaluat* OR analy*; AND abstract=computing OR coding OR programming OR "computational thinking" OR "computer science"
ERIC (U.S. Dept. of Education)	abstract:(("critical thinking") AND (assess* OR measur* OR evaluat* OR analy*) AND ("k-12" OR school* OR learn* OR teach* OR course OR teen*) AND (computing OR coding OR programming OR "computational thinking" OR "computer science"))
Google	*Due to limitations of the Google search engine a reduced search string has been used:* "critical thinking" "learning" "school" "computing" "k-12" "artificial intelligence" "assessment"
Google Scholar	("critical thinking") AND (assess* OR measur* OR evaluat* OR analy*) AND (k-12 OR school* OR learn* OR teach* OR course OR teen*) AND (computing OR coding OR programming OR "computational thinking" OR "computer science")
IEEE Xplore	("Abstract":critical thinking) AND ("Abstract":assess* OR "Abstract":measur* OR "Abstract":evaluat* OR "Abstract":analy*) AND ("Abstract":k-12 OR "Abstract":school* OR "Abstract":learn* OR "Abstract":teach* OR "Abstract":course OR "Abstract":teen*) AND ("Abstract":computing OR "Abstract":coding OR "Abstract":programming OR "Abstract":"computational thinking" OR "Abstract":"computer science")
ScienceDirect (Elsevier)	*Due to limitations of the ScienceDirect search engine a reduced search string has been used:* "critical thinking" AND (assess OR measure OR evaluate OR analyse) AND (school) AND (computing OR "computational thinking" OR "computer science")
Scopus (Elsevier)	(TITLE-ABS-KEY ("critical thinking") AND TITLE-ABS-KEY (assess* OR measur* OR evaluat* OR analy*) AND TITLE-ABS-KEY ("k-12" OR school* OR learn* OR teach* OR course OR teen*) AND TITLE-ABS-KEY (computing OR coding OR programming OR "computational thinking" OR "computer science")) AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (LANGUAGE , "English"))
SocArXiv	("critical thinking") AND (assess* OR measur* OR evaluat* OR analy*) AND ("k-12" OR school* OR learn* OR teach* OR course OR teen*) AND (computing OR coding OR programming OR "computational thinking" OR "computer science")
SpringerLink	"critical thinking" AND ("assess*" OR "measur*" OR "evaluat" OR "analy*") AND ("k-12" OR "school*" OR "learn*" OR "teach*" OR "course" OR "teen*") AND ("computing" OR "coding programming" OR "computational thinking" OR "computer science")
Wiley Online Library	"critical thinking"" in Abstract and "assess* OR measur* OR evaluat* OR analy*" in Abstract and "k-12* OR school* OR teach* OR course OR teen*" in Abstract and "computing OR coding OR programming OR "computational thinking" OR "computer science" in Abstract

Table 10
Number of identified artifacts per repository and selection stage

Source	No. of search results	No. of analyzed search results	No. of potentially relevant results	No. of relevant results (without duplicates)
ACM Digital Library	83	83	16	0
arXiv	39	39	0	0
ERIC	12	12	0	0
Google	131,000	200	5	0
Google Scholar	339,000	200	26	13
IEEE Xplore	222	222	4	2
ScienceDirect (Elsevier)	4,064	200	7	1
Scopus (Elsevier)	551	200	8	1
SocArXiv	0	0	0	0
SpringerLink	2,849	200	7	1
Wiley Online Library	17	17	1	0
Total number of relevant results without duplicates				**18**

3.3. *Search Execution*

The first author realized the search in April 2024 and revised it with the co-authors. The initial search returned 477,837 studies. Analyzing the titles, abstracts, and keywords of the 200 most relevant results from each search with regard to the inclusion/exclusion criteria identified 74 potentially relevant artifacts (Table 10).

Subsequently, the author and co-authors reviewed the full articles and excluded those not meeting the established inclusion and quality criteria. Articles that did not focus on computing were excluded (e.g., Dominguez *et al.*, 2021; Clark *et al.*, 201; Gentile *et al.*, 2019; Hsu *et al.*, 2022; Tasgin and Dilek, 2023). Were also excluded articles on assessments aimed at undergraduate and graduate levels (e.g., Azhar *et al.* 2023; Haghparast *et al.*, 2018; Walden *et al.*, 2013) or in the context of teacher training programs (e.g., Mouta *et al.*, 2019). In addition, applying the quality criteria excluded lightning talks (e.g., Günay *et al.*, 2019), abstracts only (e.g., Fouché and Mangle, 2017), or articles not available in English (e.g., Kim *et al.*, 2019; Bae and Nam, 2010). Articles inaccessible via the Capes Portal were also excluded (e.g., Adams *et al.*, 2019; Chen *et al.*, 2021). Finally, duplicates were excluded, and articles referring to the same assessment approach were unified. As a result, a total of 18 articles were considered relevant for subsequent analysis.

4. Analysis of the Results

This section presents the results for each analysis question based on information extracted from the relevant articles.

4.1. *Considerations on Analysis Procedures*

When information was not explicitly presented within the primary studies, some characteristics were inferred based on the context of the studies, including the analysis of original measurement instruments used, referenced by the studies.

The inference process, following Krippendorff (2023), was conducted only when essential information was not explicitly reported. The lead author made initial inferences based on the context and information from the studies. These inferences were then reviewed and discussed in detail with the co-authors, and were only considered after consensus was reached among all authors.

The extracted information is detailed in Appendix A–D.

4.2. *Results of Analysis Questions*

AQ1. What existing studies include assessing the development of critical thinking in the context of K-12 computing education?

The search identified 18 articles that present studies that include the assessment of critical thinking in the context of K-12 computing education (Table 11).

It was observed that “Critical Thinking” as a topic has been considered in recent studies, mainly from 2018 onwards (Fig. 1). However, given the importance of critical thinking, few studies assess the development of critical thinking in computer science teaching.

Most of the studies (n=9) were conducted in the Asian continent, mainly in China. A notable set of applications (n=6) was observed in Turkey (Fig. 2).

Only a subset (n=6) of these articles specifically investigated the development of critical thinking. The other studies had the general objective of evaluating students’ computational thinking, in which critical thinking is one of the skills assessed. Other assessments related to critical thinking include algorithmic thinking (e.g., Jiang and Li, 2019), creativity (e.g., Sun and Li, 2019), and problem-solving (e.g., Durak, 2020).

The majority of the studies (n=12) took place in the context of extracurricular courses addressing concepts such as algorithms, logic, and programming aimed at students with no prior experience in computing (Fig. 3). Some studies (n=3) reported the use of programming associated with robotics, providing students with a practical and applied experience using Arduino hardware (Durak *et al.*, 2019; Liu *et al.*, 2022; Saritepeci and Durak, 2017).

Three studies have been conducted in interdisciplinary instructional units in STEM education (Duran and Şendağ, 2012; Huang and Qiao, 2024; Yang and Chang, 2013). Yang and Chang (2013) reported the development of a game addressing the knowledge learned in the biology instructional unit. Duran and Şendağ (2012) integrated IT into STEM education through projects. Huang and Qiao (2024) utilized an AI model using machine learning techniques to classify images of dogs and cats.

Table 11
Relevant articles

Reference	Title
(Durak, 2020)	The Effects of Using Different Tools in Programming Teaching of Secondary School Students on Engagement, Computational Thinking, and Reflective Thinking Skills for Problem-Solving
(Durak, *et al.*, 2019)	Computational Thinking, Programming Self-Efficacy, Problem Solving, and Experiences in the Programming Process Conducted with Robotic Activities
(Duran and Şendağ,2012)	A Preliminary Investigation into Critical Thinking Skills of Urban High School Students: Role of an IT/STEM Program
(Huang and Qiao, 2024)	Enhancing Computational Thinking Skills Through Artificial Intelligence Education at a STEAM High School
(Jiang and Li, 2019)	Effect of Scratch on computational thinking skills of Chinese primary school students.
(Jin, *et al.*, 2021)	The impact of different types of scaffolding in project-based learning on girls' computational thinking skills and self-efficacy
(Li *et al.*, 2023a)	A study on the relationship between student learning engagements and higher-order thinking skills in programming learning.
(Li *et al.,* 2023b)	Developing and testing a Design-Based Learning Approach to Enhance Elementary Students' Self-Perceived Computational Thinking
(Liu *et al.*, 2022)	Innovation of Teaching Tools during Robot Programming Learning to Promote Middle School Students' Critical Thinking
(Negoro, *et al.*, 2023)	Scratch-Assisted Waves Teaching Materials: ICT Literacy and Students' Critical Thinking Skills
(Oluk and Korkmaz, 2016)	Comparing Students' Scratch Skills with Their Computational Thinking Skills in Terms of Different Variables
(Qu *et al.*, 2023)	Research on the Application of Gamification Programming Teaching for High School Students' Computational Thinking Development
(Saritepeci and Durak,2017)	Analyzing the effect of block and robotic coding activities on computational thinking in programming education
(Saritepeci, 2020)	Developing Computational Thinking Skills of High School Students: Design-Based Learning Activities and Programming Tasks
(Sun and Li, 2019)	Improving Junior High School Students' Creativity, Critical Thinking and Learning Attitude in Minecraft Programming
(Tonbuloğlu and Tonbuloğlu,2019)	The Effect of Unplugged Coding Activities on Computational Thinking Skills of Middle School Students
(Wong and Cheung, 2020)	Exploring children's perceptions of developing twenty-first-century skills through computational thinking and programming
(Yang and Chang, 2013)	Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement

The instructional units on programming mainly adopted block-based visual programming environments, primarily using Scratch in eight studies (e.g., Durak, 2020; Jin *et al.*, 2021; Li *et al.*, 2023b). Kodu (Wong and Cheung, 2020) and Alice (Durak, 2020) were other environments used.

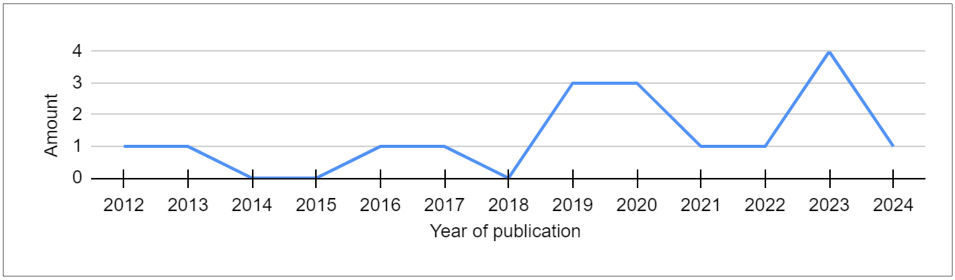


Fig. 1. Publications on the assessment of critical thinking in the context of computing education in K-12 per year.

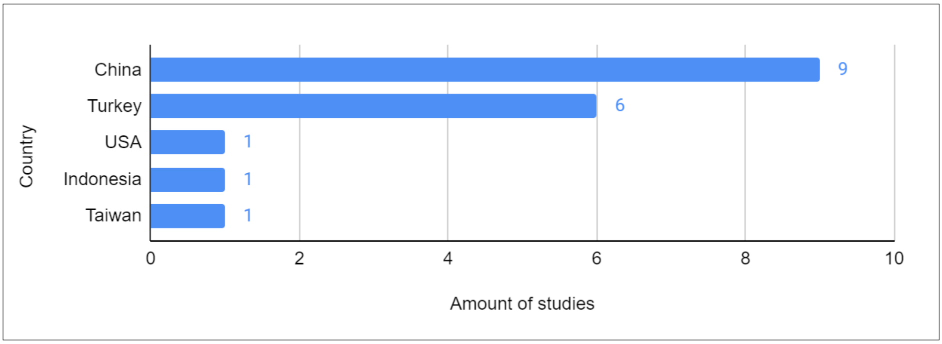


Fig. 2. Distribution of studies per country.

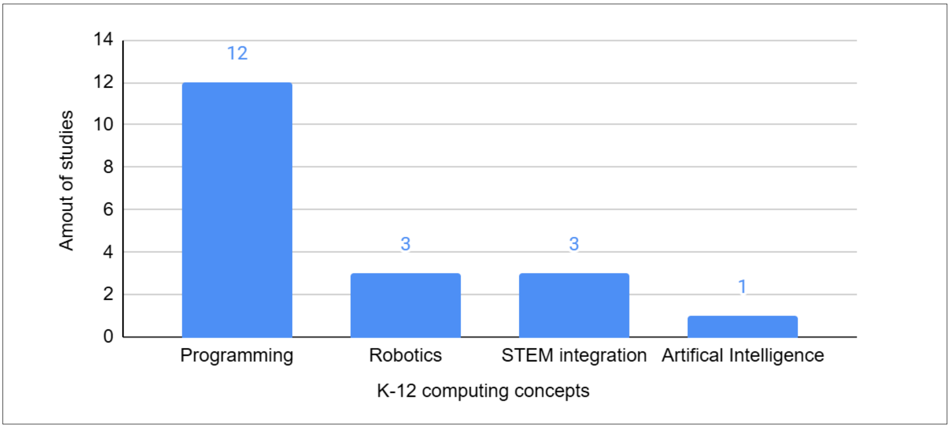


Fig. 3. Distribution of studies per computing concepts taught.

Some authors reported the use of game-based environments, such as Minecraft (Qu *et al.*, 2023 and Sun and Li, 2019) and CodeCombat (Saritepeci, 2020) (Fig. 4). The use of the text-based Python programming language was reported in three studies (Qu

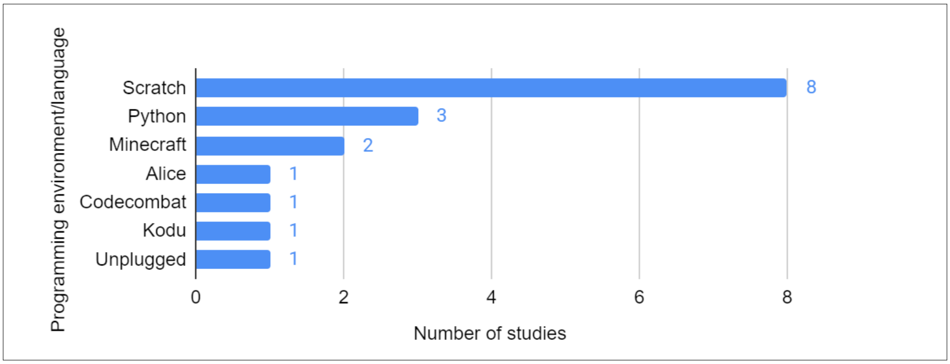


Fig. 4 Distribution of studies per programming environments/languages used in studies

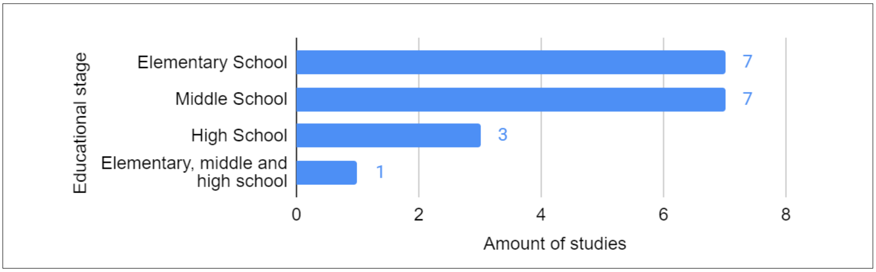


Fig. 5. Distribution of studies per educational stage

et al., 2023; Sun and Li, 2019; and Saritepeci, 2020). Saritepeci (2020) investigated the impact of design-based teaching activities, including the collaborative preparation of documents, images, videos, posters, infographics, and interactive pages, compared to teaching Python programming. In another study, Sun and Li (2019) reported that students were instructed to develop Python code to solve problems encountered in games. Qu *et al.* (2023) utilized game-oriented programming in Python. Only one study reported using an unplugged approach for teaching programming, in which students developed flowcharts and algorithms to solve everyday problems, such as water pollution (Tonbuloğlu and Tonbuloğlu, 2019).

Regarding the educational stage, most studies focused on applications in elementary (n=14) and middle school (n=4). Studies targeting these educational stages range from 3rd grade in elementary school (e.g., Jin *et al.*, 2021) to 9th grade in middle school (Liu *et al.*, 2022).

Three studies were exclusively applied in high school (Huang and Qiao, 2024; Negoro *et al.*, 2023; and Qu *et al.*, 2023). Only one study (Li *et al.*, 2023a) researched all three K-12 educational stages (Fig. 5).

AQ2. How is critical thinking defined in the studies, and what skills are being assessed?

All studies assessed at least one of the main skills of critical thinking acquired by students, within the set of core cognitive skills that constitute critical thinking as reported in the Delphi Report (Facione, 1990). Only five studies explicitly report the assessed skills of critical thinking (Duran and Şendağ, 2012; Li *et al.*, 2023a; Liu *et al.*, 2022; Negoro *et al.*, 2023; Yang and Chang, 2013). In other cases, the assessed skills were inferred in this mapping by authors from the context, following the methodology of Krippendorf (2013) (Fig. 6). The inferences were made based on the context and the original measurement instruments referenced by the studies.

Among the most assessed skills of critical thinking, "Evaluation" was identified in all studies. In these cases, for example, Negoro *et al.* (2023) compared students' evaluation skills of critical thinking before and after implementing a study instruction on the analysis of wave phenomena (in the subject of physics, in K-12), simulated with Scratch, to students who studied this phenomenon without the practical programming intervention. Yang and Chang (2013) investigated students' ability to evaluate the strength of an argument in creating biology-themed games. Huang and Qiao (2024) examined students' evaluation ability in an AI course integrated with STEM education, in which students created an image classification system using a Machine Learning model.

Another widely assessed skill was "Analysis" (n=16); however, three studies explicitly reported investigating this skill (Duran and Sendag, 2012; Li *et al.*, 2023a; Negoro *et al.*, 2023). Duran and Sendag (2012) examined students' ability to "analyze" their IT projects, which, in the context of the Test of Everyday Reasoning (TER) instrument used (Facione, 2012), corresponds to the ability to break down problems in their projects, distinguish relevant information, and recognize the situation. On the other hand, Li *et al.* (2023a) examined students' perception of analysis in creating artifacts with real-world problem solutions using programming. Negoro *et al.* (2023) assessed "Analysis" in the field of students' argumentation, given the experiments in the application.

Several studies (n=15) evaluate the "Inference" skill by investigating students' ability to draw conclusions based on the information received in their learning. An example of the evaluation of this skill of critical thinking is reported by Duran and Sendag (2012). The authors defined inference based on Facione (2012), which is "the ability to query evidence, conjecture alternatives, and conclude." Furthermore, the authors define that in the Test of Everyday Reasoning (TER) context, "inference skills are used to draw con-

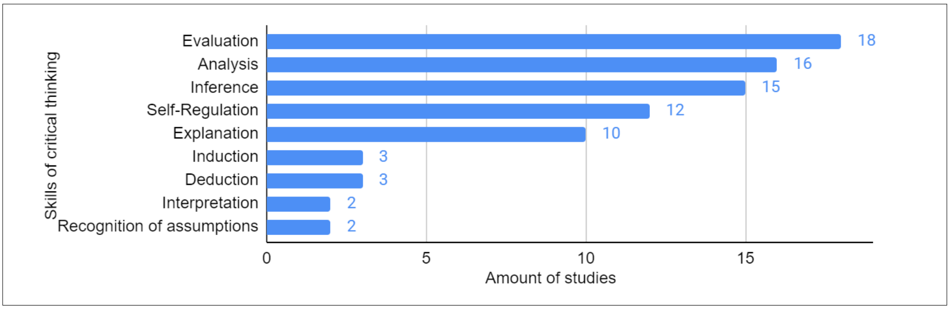


Fig. 6. Distribution of the skills of critical thinking assessed in the studies.

clusions based on reasons and evidence". The authors assessed "Inference" in students' experiences based on their IT/STEM projects.

The skill of critical thinking "Self-regulation" was also assessed in several studies (n=12). These studies primarily assessed students' willingness to learn challenging content and their ability to develop regular plans for solving complex problems (e.g., Durak, 2020; Huang and Qiao, 2024; Jiang and Li, 2019). In the studies by Jiang and Li (2019), the authors assessed this skill in students' ability to develop consistent strategies for solving highly complex issues in a programming course.

Only two studies assessed the skill of critical thinking "Interpretation": Liu *et al.* (2022) and Yang and Chang (2013). The authors analyzed students' ability to understand the information received correctly. Liu *et al.* (2022) investigated whether robotics teaching developed this skill in students compared to the pre-test. Yang and Chang (2013) examined how teaching programming associated with digital games improved students' understanding of information.

More specifically, Liu *et al.* (2022) and Yang and Chang (2013) assessed the skills of critical thinking: "Recognition of assumption", "Induction", and "Deduction". Duran and Şendağ (2012) assessed "Induction" and "Deduction". The objective of these studies was to investigate whether there was a significant increase in the scores of these critical thinking skills before and after learning programming, analyzing the student's ability to evaluate what they need to learn to obtain a more appropriate result based on known data, and, using reason, to reach a satisfactory conclusion.

AQ3. How are these critical thinking skills assessed?

The vast majority of studies (n=17) used instruments that are well-known in the literature to assess students' critical thinking skills (Fig. 7). Five studies specifically assessed critical thinking (Jin *et al.*, 2021; Li *et al.*, 2023a; Liu *et al.*, 2022; Sun and Li, 2019; Yang and Chang, 2013). The other studies assessed critical thinking as part of the assessment of computational thinking. The most widely adopted instrument was the "Computational Thinking Skill Level - Secondary School (CTLS)", developed by Korkmaz *et al.* (2015). CTLS is a scale designed to measure computational thinking levels for secondary school students. Various researchers have employed it for different purposes related to assessing computational thinking and its associated skills. This instrument was used in six studies (Durak, 2020; Durak *et al.*, 2019; Oluk and Korkmaz, 2016; Saritepeci and Durak, 2017; Saritepeci, 2020; Tonbuloğlu and Tonbuloğlu, 2019). Durak (2020) and Durak *et al.* (2019) used the CTLS to investigate computational thinking skill levels in learning programming and robotics. Oluk and Korkmaz (2016) also employed the CTLS to explore computational thinking skills, including critical thinking, as part of programming education. Saritepeci and Durak (2017) and Saritepeci (2020) utilized the CTLS to investigate the effects of computational thinking skills on programming and robotics education and the impact of programming and design-based learning activities on developing these skills. Tonbuloğlu and Tonbuloğlu (2019) used the CTLS to investigate the effect of unplugged coding activities on computational thinking skills.

Some studies (n=4) used versions derived from the CTLS, such as the Computational Thinking Scales (CTS) developed by Korkmaz *et al.* (2017). This instrument was used in the studies by Huang and Qiao (2024), Jiang and Li (2019), Qu *et al.* (2023), and Li *et al.* (2023b). Li *et al.* (2023b) and Jiang and Li (2019) used a version of the CTS translated into Chinese for students in schools in China. Jiang and Li (2019) further simplified the CTS questions to facilitate participants' understanding (between 10 and 11 years old).

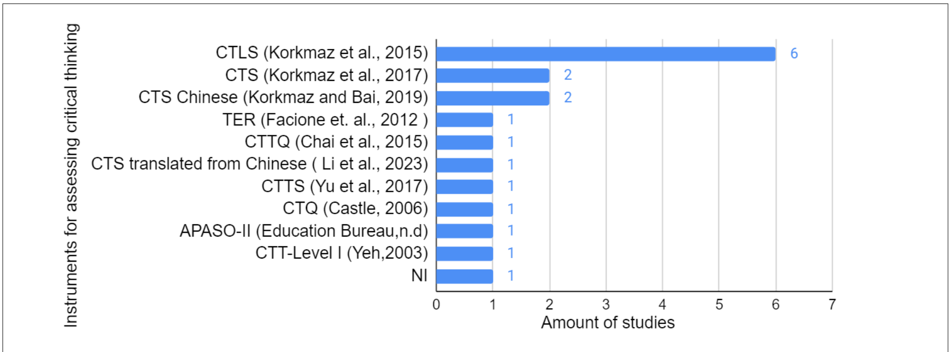
Other instruments have been used to assess critical thinking skills in specific contexts. Sun and Li (2019) used the Critical Thinking Questionnaire (CTQ) by Castle (2016), a modified version of the California Critical Thinking Skills Test (CCTST) by Facione (1990). The CTQ was initially developed to assess students' critical thinking skills in a radiography course (Castle, 2016). In the findings, Sun and Li (2019) used the CTQ to investigate game-based programming teaching to develop creativity and critical thinking.

Another instrument used was the Critical Thinking Tendency Questionnaire (CTTQ), developed and refined by Lin *et al.* (2019), Liu *et al.* (2018), and Chai *et al.* (2015) to assess students' multidimensional perceptions of 21st-century learning practices. Li *et al.* (2023a) employed the CTTQ to investigate the development of computational thinking skills through a design-based learning approach.

It was also reported that the Critical Thinking Tendency Scale (CTTS), developed by Yu *et al.* (2017) and adapted by Liu *et al.* (2022), was used to investigate the effectiveness of different teaching tools in promoting critical thinking in robot programming learning.

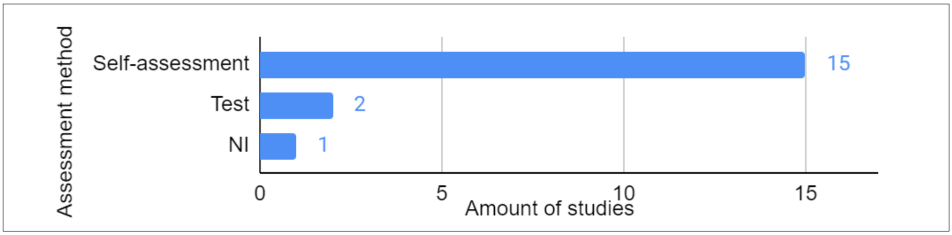
Other instruments used include the Assessment Program for Affective and Social Outcomes (APASO-II), developed by the Education Bureau, used by Wong and Cheung (2020) to investigate the impact of programming on creative thinking, critical thinking and problem-solving.

The Test of Everyday Reasoning (TER), developed by Facione (2012), employed by Duran and Şendağ (2012) to investigate the impact of an IT program in the context of STEM education on critical thinking skills.



NI - not informed or not identified

Fig.7 Instruments for assessing critical thinking skills.



NI - not informed or not identified

Fig.8 Amount of adopted assessment methods.

And, the Critical Thinking Test-Level I (CTT-Level I), developed by Yeh (2003) to assess the critical thinking skills of elementary and secondary school students. Yang and Chang (2013) used this test to investigate the impact of digital game authorship on concentration, critical thinking skills, and academic achievement.

The majority of studies (n = 17) use a student self-assessment paradigm, empowering students to evaluate the critical thinking skills they have developed (e.g. (Durak, 2020; Li *et al.*, 2023b; Tonbuloğlu and Tonbuloğlu, 2019). Additionally, most studies (n = 17) described the method used in their instruments to assess critical thinking. Of these, fifteen used the "student self-assessment" method in their instruments (Fig. 8). Only two studies reported the use of "tests" for the assessment of critical thinking (Duran and Şendağ, 2012 and Yang and Chang, 2013). Duran and Şendağ (2012) used tests in a summative assessment paradigm, administering pre-tests, mid and post-tests. This test involves the student's reasoning skills using questions that progressively invite participants to analyze or interpret information presented in texts, graphs, or images, draw accurate and secure inferences, evaluate inferences, and explain why they represent strong or weak reasoning. Yang and Chang (2013), on the other hand, used summative evaluation to measure results, and incorporated elements of formative evaluation through their collaborative game design process. Yang and Chang (2013), reported the use of 25 multiple-choice questions of "CTT-Level I" instrument, but did not detail the characteristics of the questions used in the test.

Most studies (n = 13) used 4 to 5 items in the instruments for evaluating critical thinking, while others varied from 12 to 35 items. The items for assessing critical thinking followed a 5-point Likert response scale in most studies (n = 13). Only Wong and Cheung (2020) used 4-point Likert scales.

The studies that used tests as the evaluation method employed multiple-choice questions (Duran and Şendağ, 2012; Yang and Chang, 2013).

AQ4. How has the assessment approach been evaluated?

The majority of the studies (n = 14) evaluate the quality of the critical thinking assessment instrument.

The sample size for evaluating the instrument was not reported in most studies (n = 14). The studies that reported sample sizes used a small number, from 68 students (Negoro *et al.*, 2023) to a more significant sample, applied to 580 students (Li *et al.*, 2023b).

The majority (n = 12) of the studies assessed the reliability of the instruments, while only six reported on their validity.

Reliability. Even considering that the original authors previously evaluated the assessment instrument reliability, the authors in their studies (n = 12) reassessed the instrument's reliability with their study data. These studies calculated Cronbach's alpha (Cronbach, 1951) coefficient to analyze the reliability of the data collection instrument. The results mainly indicated good reliability, with most coefficients $\alpha > .8$. Only one exception, in Negoro *et al.* (2023), obtained reliability considered as poor ($\alpha = .597$).

The reliability measured from the instrument used by Li *et al.* (2023a) obtained the highest Cronbach's alpha coefficient = .985. This result indicates excellent internal consistency of the instrument (Table 12).

Validity. Six studies reported the evaluation of the instrument's validity (Durak, 2019; Jiang and Li, 2019; Li *et al.*, 2023b; Liu *et al.*, 2022; Negoro *et al.*, 2023; Oluk and Korkmaz, 2016) (Table 13). Durak (2019) used factor-total correlation to evaluate the relationship between the observed variable and the observed latent factor. The results indicated that the individual variables had a moderate to strong correlation with the general factor, confirming the instrument's validity.

Jiang and Li (2019) analyzed the validity of the data collection instrument through exploratory and combinatorial factor analysis. Verifying KMO and Bartlett's test confirmed the possibility of performing an exploratory factor analysis. Multiple fit indices were used to evaluate the validity of the instrument. The results showed that the instrument fit the data well. Li *et al.* (2023b) analyzed the instrument's validity by per-

Table 12

Overview of reliability evaluations

Reference	Sample size	Reliability Results (Cronbach alpha)	Reliability findings
(Durak, 2020)	NI	$\alpha = .866$	Good
(Durak, 2019)	NI	α = between .78 to .94 for the subscales	Good
(Jiang and Li, 2019)	NI	$\alpha = .893$	Good
(Jin, *et al.*, 2021)	158	$\alpha = .838$	Good
(Li *et al.*, 2023a)	NI	$\alpha = .985$	Excellent
(Li *et al.*, 2023b)	580	α = between .79 to .88 for the subscales	Good
(Liu *et al.*, 2022)	485	$\alpha = .955$	Good
(Negoro, *et al.*, 2023)	68	$\alpha = .597$	Poor
(Oluk and Korkmaz, 2016)	241	$\alpha = .809$	Good
(Saritepeci and Durak,2017)	NI	$\alpha = .853$	Good
(Saritepeci,(2020)	NI	$\alpha = .867$	Good
(Sun and Li, 2019)	NI	$\alpha = .84$	Good

NI - not informed or not identified

Table 13

Overview of validity evaluations

Reference	Sample size	Validity (CFA indices)	Validity findings
(Durak, 2019)	NI	Factor-total correlation: .48-.73	-
(Jiang and Li, 2019)	NI	$\chi 2$/df = 1.989 CFI = .922 TLI = .908 RMSEA = .047	Excellent Acceptable Acceptable Excellent
(Li *et al.*, 2023b)	580	AGFI = .90 CMIN/DF = 3.232 CFI = .95 GFI = .91 IFI = .97 RMSEA = .062 SRMR = .044	Good Acceptable Excellent Good Excellent Acceptable Good
(Liu *et al.*, 2022)	485	$\chi 2$/df = 2.091 CFI = .974 GFI = .916 IFI = .975 NFI = .952 RMSEA = .047	Excellent Excellent Good Excellent Excellent Excellent
(Negoro, *et al.*, 2023)	68	AGFI = .90 CFI = .93 GFI = .91 NFI = .96 TLI/NNFI = .98 RMSEA = .01 SRMR = .044	Good Acceptable Good Excellent Excellent Excellent Good
(Oluk and Korkmaz, 2016)	241	Maximum likelihood regression: .507-.872 Item-test correlation: .655-.862	Strong to weak correlation

NI - not informed or not identified

forming exploratory and confirmatory factor analysis based on data collected from 580 participants. With a KMO greater than .5 and a significant Bartlett's test statistic, the subsequent factor analysis can reasonably proceed on the scale. The fit indices of the CFA analysis showed that the instrument fit the data sufficiently well.

Liu *et al.* (2022) reported the verification of KMO and Bartlett based on a sample of 485 participants, confirming the possibility of performing an exploratory factor analysis. The fit indices of the CFA analysis showed that the instrument was suitable for the scale. Negoro *et al.* (2023) used second-order factor analysis. The analysis showed that the critical thinking assessment instrument meets the valid criteria observed in the "Goodness of Fit." Furthermore, the values of the CFA analysis indicated the appropriate scale. Oluk and Korkmaz (2016) used combinatorial factor analysis through the maximum likelihood technique. The values found showed a moderate to solid fit of the model, demonstrating the consistency and validity of the items. The item-test correlation coefficient confirms the instrument's effectiveness in measuring critical thinking.

5. Discussion

Despite the growing importance of critical thinking as an essential 21st-century skill, the systematic mapping results revealed that few studies are dedicated to evaluating this competency's development in K-12 computing education.

AQ1. What existing studies include assessing the development of critical thinking in the context of K-12 computing education.

Of the 18 select studies, the majority evaluate the development of critical thinking in teaching algorithms and programming. This can be attributed to the fact that teaching these concepts are a more consolidated area within K-12 computing education, with more established curricular guidelines and pedagogical practices, considering, for example, computational thinking frameworks such as K12CS (K12CS.org, 2016) and the Computational Thinking for Science framework (CT-S) (Hurt *et al.*, 2023). Thus, there is a need for more studies focused on evaluating critical thinking in other areas, such as robotics, integration with STEM, and especially in teaching AI (found in only one study). Given the growing integration of artificial intelligence in everyday life and its impact on multiple domains, it becomes imperative that learners develop critical thinking skills to deal with the challenges and possibilities presented by this emerging technology (UNICEF, 2023; Lee *et al.*, 2023).

The origin of these studies is concentrated in the Asian continent. This trend may be aligned with the educational policies of Asian countries that recognize the importance of developing these skills early to prepare students for the challenges of the digital era (Jiang and Li, 2021; Wong and Cheung, 2020; Wong and Jiang, 2018).

In terms of the educational stages, most studies were conducted in elementary and middle school, which may be related to the introduction of computing curricula from the early years in Asia, especially in China, with a focus on the development of computational thinking and skills such as critical thinking (Jiang and Li, 2021). However, the smaller number of studies in high school suggests the need for more initiatives to enhance critical thinking skills at this educational stage as well, considering the importance of preparing students for higher education, the job market, and to become future leaders (Sari *et al.*, 2022; OECD, 2019).

AQ2. How is critical thinking defined in the studies, and what skills are being assessed.

There needs to be more consensus in the analyzed studies concerning the definitions of critical thinking. However, many studies define it as a subskill of Computational Thinking, aligning with the definition by Korkmaz *et al.* (2015), which defines Critical Thinking as a "high-level thinking skill" in a more general sense.

It was also observed that there needs to be a consensus on the skills that compose critical thinking. Each study assessed a specific set of skills based on their research needs and teaching objectives.

Among the skills analyzed, "Evaluation" was the most assessed, present in all the studies, to discover the students' beliefs and opinions about the results achieved. On the other hand, "Interpretation" was the least assessed. None of the studies assessed the complete set of skills defined as "core skills" by Facione (1990).

AQ3. How are these critical thinking skills assessed?

Most studies adopted "student self-assessment" as the most commonly used assessment paradigm and method for critical thinking. Only two studies used "tests," and one did not report the assessment method. In this sense, although self-assessment has positive aspects, such as helping students develop metacognitive skills and analyze their learning progress (Andrade, 2019), it can also present limitations, such as lack of objectivity and difficulty in identifying "blind spots" in learning, which may mask the degree of skills developed by students (Taylor, 2014).

AQ4. How has the assessment approach been evaluated?

Regarding the quality of the instruments used for assessing critical thinking, most studies analyzed the reliability of these instruments, reporting good internal consistency with Cronbach's alpha coefficients above .8. Only one exception reported internal consistency with a Cronbach's alpha coefficient below .6, considered "poor" by Gliem and Gliem (2003). Few studies examined the validity of the instruments; of these, the majority used exploratory and/or confirmatory factor analysis. The results obtained were favorable, validating the measurement instruments. The studies used samples ranging from 68 to 580 participants; the results can be classified as consistent and, therefore, in line with Hair *et al.* (2009), who discuss the complexity of conducting factor analysis with a sample smaller than 50 participants. However, it is observed that half of these studies should have mentioned the sample size used, making the study results difficult to analyze.

Threats to validity. A major threat of systematic mappings is the omission of relevant studies. To mitigate this threat, we precisely delimited the scope of the research, identifying the key concepts and their synonyms. In addition, we included critical thinking as a dimension of computational thinking to minimize the risk of omission.

Another limitation was that the analysis was restricted to the 200 most relevant papers from the initial search results. Relevance was determined by the search engines' algorithms, which consider factors such as citation count, publication date, and keyword matching. To mitigate this limitation, we conducted supplementary searches on Google Scholar and Google, using different ranking criteria to reduce the possibility of overlooking relevant studies (Piasecki *et al.*, 2018).

Measures to mitigate possible threats to study selection and data extraction were adopted by defining explicit inclusion/exclusion and quality criteria. The author followed the select criteria, and the findings were discussed with the co-authors until a consensus was reached.

When information was not explicitly reported in the studies analyzed, the authors inferred the data based on the context. This inference process followed the methodology of Krippendorff (2013), and was necessary to fill in gaps in the reports. However, this process could introduce potential bias. To mitigate this, all inferred data was thoroughly reviewed and discussed by the co-authors to achieve consistency, correctness, and accuracy.

6. Conclusion

This systematic literature mapping reveals that, although critical thinking is an essential skill in the 21st century, little research has been carried out on this topic in computing education.

Eighteen studies assessed the development of critical thinking in K-12 computing education, mainly in extracurricular courses in programming, logic, and algorithms. These studies indicated a need for more consensus on the definition of critical thinking and the skills that compose the studies. Thus, there is no fixed set of critical thinking skills, indicating that the skills are listed according to the needs and particularities of each study. In addition, the results showed that the instruments used to assess critical thinking are third-party, well-known in the literature. Most studies use student self-assessment as the evaluation method. Also, most studies evaluated the reliability of the instruments in contrast to their validity to assess the quality of the methods. However, the results presented demonstrate the validity and internal consistency of the instruments.

The findings from this mapping provide indications to guide educators and researchers in developing initiatives and applying practical assessments to promote critical thinking skills in K-12 computing education. In summary, the findings support the need for more comprehensive and diverse assessments of the development of critical thinking in K-12 computing education, covering different contexts, computing concepts, geographical regions, and educational stages.

Acknowledgments

A special thanks to the Federal University of Santa Catarina and the Federal Institute of Santa Catarina for supporting the research.

References

Adams, C., Cutumisu, M., & Lu, C. (2019). Measuring K-12 Computational Thinking Concepts, Practices and Perspectives: An Examination of Current CT Assessments. In: Proc. of Society for Information Technology & Teacher Education International Conference, Las Vegas, NV, U.S, 275-285.

Aktoprak, A., & Hursen, C. (2022). A bibliometric and content analysis of critical thinking in primary education. Thinking Skills and Creativity, 44, 101029.

Allen, J.P., Pianta, R.C., Gregory, A., Mikami, A.Y., & Lun, J. (2011). Observations of Effective Teacher–Student Interactions in Secondary School Classrooms: Predicting Student Achievement With the Classroom Assessment Scoring System—Secondary. Journal of Educational Psychology, 103(1), 76-98.

Anazifa, R., & Djukri, D. (2017). Project-based learning and problem-based learning: Are they effective to improve student's thinking skills? Jurnal Pendidikan IPA Indonesia, 6(2), 346-355.

Anders, P.L., Stellrecht, E.M., Davis, E.L., & McCall, W.D., Jr (2019). A Systematic Review of Critical Thinking Instruments for Use in Dental Education. Journal of Dental Education, 83(4), 381-397.

Andrade, H. (2019). A Critical Review of Research on Student Self-Assessment. Frontiers in Education, 4, 87.

Azhar, M.Q., Haynes, A., Day, M., & Wissinger, E. (2023). Implementation and evaluation of a virtual hackathon in an urban HSI community college during COVID-19. Journal of Computing Sciences in Colleges, 38(6), 72-84.

Bae, Y.-K., & Nam, J.-W. (2010). Impact of robot programming education in application of Web 2.0 on improving problem-solving ability. The Journal of the Korea Contents Association, 10(11), 468-475.

Braun, H.I., Shavelson, R.J., Zlatkin-Troitschanskaia, O., & Borowiec, K. (2020). Performance assessment of critical thinking: Conceptualization, design, and implementation. Frontiers in Education, 5, 156.

Brookhart, S.M. (2010). How to assess higher-order thinking skills in your classroom. ASCD.

Caspersen, M.E., *et al.* (2022). The Informatics Reference Framework for School. Informatics for All. `https://www.informaticsforall.org/wp-content/uploads/2022/03/Informatics-Reference-Framework-for-School-release-February-2022.pdf`

Castle, A. (2006). Assessment of the critical thinking skills of student radiographers. Radiography, 12(2), 88-95.

Chai, C.S., Deng, F., Tsai, P.S., Koh, J.H.L., & Tsai, C.C. (2015). Assessing multidimensional students' perceptions of twenty-first-century learning practices. Asia Pacific Education Review, 16(3), 389-398.

Chen, Y., Yu, X., & Wu, Y. (2021). The Game Design and Development of Critical Thinking Ability Assessment. In: Proc. of Society for Information Technology & Teacher Education International Conference, Waynesville, NC, USA, 394-401.

Clark, A.Y., Li, Y., & Jiang, Y. (2018). Using natural language processing and qualitative thematic coding to explore math learning and critical thinking. In: Proc. of the International Conference on Big Data and Education, Honolulu, HI, U.S, 38-43.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46.

Cortázar, C., Nussbaum, M., Harcha, J., Alvares, D., López, F., Goñi, J., & Cabezas, V. (2021). Promoting critical thinking in an online, project-based course. Computers in Human Behavior, 119, 106705.

Create-Learn. (2023). Benefits of Python coding for kids. Retrieved from `https://www.create-learn.us/blog/benefits-of-python-coding-for-kids/`

Creswell, J.W., & Poth, C.N. (2018). Qualitative Inquiry and Research Design: Choosing Among Five Approaches (4th ed.). Sage Publications.

Criteriacorp. (2023). Pre-Employment Critical Thinking Tests. Retrieved from `https://www.criteriacorp.com/assessments/measure-critical-thinking`

CriticalThinking (2019). Foundation for Critical Thinking. Critical thinking testing and assessment. Retrieved from `https://www.criticalthinking.org/pages/critical-thinking-testing-and-assessment/594`

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334.

DeVellis, R.F. (2017). Scale development: Theory and applications (4th ed.). SAGE Publications, Inc.

Dewey, J. (1933). How we think: A restatement of the relation of reflective thinking to the educative process. Boston, MA: Houghton Mifflin Company.

Dominguez, J., Moreira, F., Gaia, R., Da Silva, R.A., & Gomes, A. (2021). Considerations on the development of critical thinking in a virtual learning environment. In: Proc. of the 9th International Conference on Technological Ecosystems for Enhancing Multiculturality, Barcelona, Spain, 225-230.

Durak, H. (2020). The effects of using different tools in programming teaching of secondary school students on engagement, computational thinking and reflective thinking skills for problem solving. Technology, Knowledge and Learning, 25, 179–195. `https://doi.org/10.1007/s10758-018-9391-y`

Durak, H.Y., Yilmaz, F.G., & Yılmaz, R. (2019). Computational Thinking, Programming Self-Efficacy, Problem Solving and Experiences in the Programming Process Conducted with Robotic Activities. Contemporary Educational Technology, 173-197.

Duran, M. & Şendağ, S. (2012). A Preliminary Investigation into Critical Thinking Skills of Urban High School Students: Role of an IT/STEM Program. Creative Education. 03. 10.4236/ce.2012.32038, Duran, M. (2012). A Preliminary Investigation into Critical Thinking Skills of Urban High School Students: Role of an IT/STEM Program. Creative Education, 03, 241-250.

Education Bureau [EDB]. (n.d.). Assessment program for affective and social outcomes (2nd Version) (APASO-II). The Government of the Hong Kong Special Administrative Region of the People's Republic of China. Retrieved from `https://www.edb.gov.hk/attachment/en/sch-admin/sch-quality-assurance/performance-indicators/esda/APASO_User_Manual_sec_en.pdf`

Elder, L., & Paul, R. (2010). Critical thinking: Competency standards essential for the cultivation of intellectual skills, Part 1. Journal of Developmental Education, 34(2), 38-39.
Ennis, R., Millman, J., & Tomko, T. (2005). Cornell Critical Thinking Tests. Seaside, CA: The Critical Thinking Co.
Ennis, R.H. (1993). Critical thinking assessment. Theory Into Practice, 32(3), 179-186.
Ennis, R., Millman, J., & Tomko, T. (1985). Cornell Critical Thinking Tests level X and level Z – Manual. Pacific Grove, CA: Midwest Publishing.
Facione, P.A. (1990). California Critical Thinking Skills Test: CCTST. California Academic Press.
Facione, P.A., *et al.* (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations. ERIC, Institute of Education Sciences, 1-112.
Facione, P., Facione, N., & Winterhalter, K. (2012). The test of everyday reasoning – (TER): Test Manual. Milbrae, CA, USA: California Academic Press.
Fagin, B., Harper, J., Baird, L., Hadfield, S., & Sward, R. (2006). Critical thinking and computer science: Implicit and explicit connections. Journal of Computing Sciences in Colleges, 21, 171-177.
Fouché, S., & Mangle, A. (2017). Web-based programming practice with Code Hunt: Conference tutorial. Journal of Computing Sciences in Colleges, 32(3), 62.
Gentile, M., Città, G., Perna, S., Signa, A., Reale, F., Dal Grande, V., Ottaviano, S., La Guardia, D., & Allegra, M. (2019). The effect of disposition to critical thinking on playing serious games. In: Proc. of the 7th International Conference, Games and Learning Alliance. Palermo, Italy, 1-17.
Gholam, A. (2019). Inquiry-based learning: Student teachers' challenges and perceptions. Journal of Inquiry & Action in Education, 10(2), 112-133.
Gliem J.A. and Gliem R.R. (2003) Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education, Columbus, 82-88.
Glorfeld, L.W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educational and Psychological Measurement, 55(3), 377–393.
Günay, C., Doloc-Mihu, A., Gluick, T., & Moore, C.A. (2019). Project-based learning improves critical thinking for software development students. In: Proc. of the 20th Annual SIG Conference on Information Technology Education, Tacoma, WA, U.S, 105.
Haghparast, M., Abdullah, N., & Nasaruddin, F.H. (2018). Fog learning for cultivating critical thinking in the information-seeking process. Concurrency and Computation: Practice and Experience, 31(8).
Hair, J.F., *et al.* (2009). Multivariate Data Analysis. Bookman.
Halpern, D.F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. American Psychologist, 53(4), 449-455.
Hattie, J., Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81-112.
Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. Evidence-Based Nursing, 18(3), 66-67.
Hsu, F.-H., Lin, I.-H., Yeh, H.-C., & Chen, N.-S. (2022). Effect of Socratic reflection prompts via video-based learning system on elementary school students' critical thinking skills. Computers & Education, 183, 104497.
Huang, X., & Qiao, C. (2022). Enhancing computational thinking skills through artificial intelligence education at a STEAM high school. Science & Education.
Huang, X., Qiao, C. (2024) Enhancing Computational Thinking Skills Through Artificial Intelligence Education at a STEAM High School. Sci & Educ 33. `https://doi.org/10.1007/s11191-022-00392-6`
Hurt, T., *et al.* (2023). The computational thinking for science (CT-S) framework: operationalizing CT-S for K–12 science education researchers and educators. Int. Journal of STEM Education, 10.
Hussein, M.H., Ow, S.H., Cheong, L.S., & Thong, M.-K. (2019). A digital game-based learning method to improve students' critical thinking skills in elementary science. IEEE Access, 7.
İlic, U. (2021). The impact of Scratch-assisted instruction on computational thinking (CT) skills of pre-service teachers. International Journal of Research in Education and Science, 7(2), 426-444.
Insight Assessment. (2023). Critical thinking assessments for higher education. Retrieved from `https://www.insightassessment.com/article/critical-thinking-assessments-for-higher-education`
Instituto Ayrton Senna. (2022). Criatividade e pensamento crítico. Retrieved from `https://institutoayrtonsenna.org.br/o-que-defendemos/criatividade-e-pensamento-critico/`

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34(2), 183–202.

Jiang, B., Li, Z. (2021). Effect of Scratch on computational thinking skills of Chinese primary school students. J. Comput. Educ. 8, 505–525. `https://doi.org/10.1007/s40692-021-00190-z`

Jin Y., J. Sun, H. Ma and X. Wang, (2021). "The impact of different types of scaffolding in project-based learning on girls' computational thinking skills and self-efficacy," Tenth International Conference of Educational Innovation through Technology (EITT), Chongqing, China. doi: 10.1109/EITT53287.2021.00077.

K–12 Computer Science Framework. (2016). Retrieved from `http://www.k12cs.org`

Kaiser, H.F., & Rice, J. (1974). Little Jiffy, Mark IV. Educational and Psychological Measurement, 34(1), 7.

Karaahmetoğlu, K., & Korkmaz, Ö. (2019). The effect of project-based arduino educational robot applications on students' computational thinking skills and their perception of Basic STEM skill levels. Participatory Educational Research, 6(2), 1-14.

Kim, S.-W., Park, H., & Lee, Y. (2019). Development of project-based robot education program for enhancing interest toward robots and computational thinking of elementary school students. Journal of the Korea Society of Computer and Information, 24(1), 247-255.

Korkmaz, Ö., Çakır, R., & Özden, M.Y. (2015). Computational thinking levels scale (CTLS) adaptation for secondary school level. Gazi Journal of Educational Science, 1(2).

Korkmaz, Ö., Çakır, R., & Özden, M.Y. (2017). A validity and reliability study of the Computational Thinking Scales (CTS). Computers in Human Behavior, 72, 558-569.

Korkmaz, Ö., & Bai, X. M. (2019). Adapting computational thinking scale (CTS) for Chinese school students and their thinking scale skills level. Participatory Educational Research, 6, 10-26.

Krippendorff, K. (2013). Content Analysis: An Introduction to Its Methodology. 3rd ed. Thousand Oaks, CA: Sage Publications.

Kuder, G.F., & Richardson, M.W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151–160.

Lee, S.J., Francom, G.M., & Nuatomue, J. (2022). Computer science education and K-12 students' computational thinking: A systematic review. International Journal of Educational Research, 114, 102008.

Lee, S., Choi, D., Lee, M., Choi, J., & Lee, S. (2023). Fostering youth's critical thinking competency about AI through exhibition. In: Proc. of the CHI Conference on Human Factors in Computing Systems, Hamburg, Germany.

Li, W., Huang, J., Liu, C., Tseng, J.C., & Wang, S. (2023). A study on the relationship between student' learning engagements and higher-order thinking skills in programming learning. Thinking Skills and Creativity.

Li, X., Xie, K., Vongkulluksn, V., Stein, D., & Zhang, Y. (2023). Developing and Testing a Design-Based Learning Approach to Enhance Elementary Students' Self-Perceived Computational Thinking. Journal of Research on Technology in Education, 55(2), 1–22.

Li-tze Hu & Peter M. Bentler (1999): Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, Structural Equation Modeling: A Multidisciplinary Journal, 6:1, 1-55.

Lin, H.C., Hwang, G.J., & Hsu, Y.D. (2019). Effects of ASQ-based flipped learning on nurse practitioner learners' nursing skills, learning achievement and learning perceptions. Computers and Education, 139, 207-221.

Lin, P. -H., & Chen, S. -Y. (2020). Design and evaluation of a deep learning recommendation based augmented reality system for teaching programming and computational thinking. IEEE Access, 8.

Liu, J.L., McBride, R.E., Xiang, P., & Scarmardo-Rhodes, M. (2018). Physical education pre-service teachers' understanding, application, and development of critical thinking. Quest, 70(1), 12-27.

Liu, Y. (2019). Using reflections and questioning to engage and challenge online graduate learners in education. Research and Practice in Technology Enhanced Learning, 14(3).

Liu, H., Sheng, J., & Zhao, L. (2022). Innovation of Teaching Tools during Robot Programming Learning to Promote Middle School Students' Critical Thinking. Sustainability, 14, 1-14.

Mäkiö, E., and Mäkiö, J. (2023). The Task-Based Approach to Teaching Critical Thinking for Computer Science Students. Education Sciences, 13(7), 742.

Martins, R.M. *et al.* (2024a). Machine learning for all!—Introducing machine learning in middle and high school. International Journal of Artificial Intelligence in Education, (34), 185-223.

Martins, R.M. *et al.* (2024b). Teaching machine learning to middle and high school students from a low socio-economic status background. Informatics in Education. 23(1), 179-222.

McDonald, R.P. (1999). Test theory: A unified treatment. Lawrence Erlbaum Associates, Inc.
Morrison G.R., Ross S.M., Morrison J.R., Kalman H.k. (2019). Designing effective instruction, Eighth edition. Wiley
Moskal, B.M., & Leydens, J.A. (2000). Scoring rubric development: Validity and reliability. Practical Assessment, Research & Evaluation, 7(10). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=10.
Mouta, A., Torrecilla Sánchez, E., & Pinto Llorente, A. (2019). Blending machines, learning, and ethics. In: Proc. of the 7th International Conference on Technological Ecosystems for Enhancing Multiculturality, León, Spain.
Negoro, R.A., Rusilowati, A., & Aji, M.P. (2023). Scratch-assisted waves teaching materials: ICT literacy and students' critical thinking skills. Journal of Turkish Science Education, 20(1), 189-210.
New York State Education Department. (2024). Performance-Based Learning and Assessment [Fact sheet]. Retrieved from https://www.nysed.gov/sites/default/files/programs/plan-pilot/fact-sheet-performance-based-learning-assessment.pdf.
OECD. (2019). OECD Learning Compass 2030: A series of concept notes. OECD Learning Compass 2030 Concept Note Series. Retrieved from https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compass_2030_Concept_Note_Series.pdf.
Oluk, A.I., & Korkmaz, Ö. (2016). Comparing Students' Scratch Skills with Their Computational Thinking Skills in Terms of Different Variables. International Journal of Modern Education and Computer Science, 8, 1-7.
P21. (2019). P21 Framework definitions. Retrieved from https://www.battelleforkids.org/networks/p21.
Paul, J.A., Sinha, M., & Cochran, J.D. (2023). Instruments to assess students' critical thinking—A qualitative approach. Decision Sciences Journal of Innovative Education, 21(3), 123-143.
Pedrosa-de-Jesus, H., & Guerra, C. (2018). Teachers' written formative feedback on students' critical thinking: A case study. Journal of Education, 9(1), 3-22.
Petersen, K., *et al.* (2008). Systematic Mapping Studies in Software Engineering. In: Proc. of the 12th Int. Conference on Evaluation and Assessment in Software Engineering, Bari, Italy, 68-77.
Piasecki, J. *et al.* (2018). Google Search as an Additional Source in Systematic Reviews. Science and Engineering Ethics, 24(2), 809-810.
Popham, W.J. (2008). Transformative assessment. ASCD.
Popat, S., & Starkey, L. (2019). Learning to code or coding to learn? A systematic review. Computers & Education, 128, 365-376.
Qu, Z., Liu, J., Che, L., Su, Y., & Zhang, W. (2023). Research on the Application of Gamification Programming Teaching for High School Students' Computational Thinking Development. In 2023 IEEE 12th International Conference on Educational and Information Technology, 144-149.
Rehmat, A.P., & Hartley, K. (2020). Building engineering awareness: Problem-based learning approach for STEM integration. Interdisciplinary Journal of Problem-based Learning, 14(1), 2-8.
Reynders, G., Lantz, J., Ruder, S.M. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. International Journal of STEM Education, 7(9), 2-15.
Saadé *et al.* (2012). Critical thinking in E-learning environments. Computers in Human Behavior, 28(5), 1608-1617.
Sari, M.K., Sudiyanto, S., & Kurniawan, S.B. (2022). Critical thinking skills profile of fourth-grade elementary school students in science learning. In: Proc. of the 5th International Conference on Learning Innovation and Quality Education, Surakarta, Indonesia.
Saritepeci, M. & Durak, H. (2017). Analyzing the Effect of Block and Robotic Coding Activities on Computational Thinking in Programming Education. Educational Research and Practice, 49.
Saritepeci, Mustafa. (2019). Developing Computational Thinking Skills of High School Students: Design-Based Learning Activities and Programming Tasks. The Asia-Pacific Education Researcher. 29. 35-54. 10.1007/s40299-019-00480-2.
Shafiyeva, U. (2021). Assessing students' minds: Developing critical thinking or fitting into a Procrustean bed. European Journal of Education, 4(2), 79-92.
Snedecor, G.W., Cochran, W.G., & Cox, G.M. (1989). Statistical methods (8th ed.). Iowa State University Press.
Sontag, M. (2009). Critical thinking with Alice: A curriculum design model for middle school teachers. In:

Proc. of the Alice Symposium, Durham, NC, USA, 1-3.
Soland, J., Hamilton, L.S., & Stecher, B.M. (2013). Measuring 21st century competencies: Guidance for educators. Global Cities Education Network Report. RAND Corporation. Retrieved from https://www.rand.org/pubs/external_publications/EP50463.html
Spector, J.M., & Ma, S. (2019). Inquiry and critical thinking skills for the next generation: From artificial intelligence back to human intelligence. Smart Learning Environment, 6, 8.
Sun, Dan & Li, Yan. (2019). Improving Junior High School Students' Creativity, Critical Thinking and Learning Attitude in Minecraft Programming.
Tasgin, A., & Dilek, C. (2023). The mediating role of critical thinking dispositions between secondary school student's self-efficacy and problem-solving skills. Thinking Skills and Creativity, 50, 101400.
Taylor, S.N. (2014). Student Self-Assessment and Multisource Feedback Assessment: Exploring Benefits, Limitations, and Remedies. Journal of Management Education, 38(3), 359-383.
Taylor, W. (2022). Promoting critical thinking through classroom discussion. In C. L. Fuiks & L. Clark (Eds.), Teaching and Learning in Honors.
Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. International journal of medical education, 11, 245–247.
Ten Haken, B. (2017). Critical Thinking Skills, AI and the Next Generation Workplace. Retrieved from https://babettetenhaken.com/2017/07/24/critical-thinking-skills-ai-workplace/.
Tonbuloğlu, B., & Tonbuloğlu, İ. (2019). The Effect of Unplugged Coding Activities on Computational Thinking Skills of Middle School Students. Informatics Educ., 18, 403-426.
U.S. Department of Education, Office of Educational Technology. (2019). Artificial Intelligence in Education: Opportunities and Implications for Federal Support. Retrieved from https://www2.ed.gov/documents/ai-report/ai-report.pdf.
UNICEF. (2023a). National AI strategies and children. Retrieved from https://www.unicef.org/globalinsight/media/1156/file
UNICEF. (2023). Artificial intelligence chatbots. Retrieved from https://www.unicef.org/eap/blog/artificial-intelligence-chatbots
Velázquez-Iturbide, J.Á. (2013). An experimental method for the active learning of greedy algorithms. ACM Transactions on Computing Education, 13(4), article 18, 1-23.
Vincent-Lancrin, S. *et al.* (2019), Fostering Students' Creativity and Critical Thinking: What it Means in School, Educational Research and Innovation, OECD Publishing, Paris.
Voskoglou, M.G., & Buckley, S. (2012). Problem Solving and Computational Thinking in a Learning Environment. arXiv preprint arXiv:1212.0750.
Walden, J., Doyle, M., Garns, R., & Hart, Z. (2013). An informatics perspective on computational thinking. In: Proc. of the 18th ACM Conference on Innovation and Technology in Computer Science Education, Lanarca, Cyprus, 4-9.
Wong, G.K.W., & Jiang, S. (2018). Computational thinking education for children: Algorithmic thinking and debugging. Proceedings of 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE, Wollongong, Australia, 328-334.
Wong, G.K.-W., & Cheung, H.Y. (2020). Exploring children's perceptions of developing twenty-first century skills through computational thinking and programming. Interactive Learning Environments, 28(4), 438-450.
World Economic Forum. (2020). The Future of Jobs Report 2020. Retrieved from https://www.weforum.org/publications/the-future-of-jobs-report-2020/.
Yang, Y.C., & Chang, C. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. Comput. Educ., 68, 334-344.
Yeh, Y.C. (2003). Critical thinking test-level I (CTT-I). Taipei, Taiwan: Psychological Publishing.
Yu, K.-C., Lin, K.-Y., & Chang, S.-F. (2017). The development and validation of a mechanical critical thinking scale for high school students. EURASIA Journal of Mathematics, Science and Technology Education, 13(5), 1361-1376.

D.M. Arndt is a professor of Telecommunications at the Federal Institute of Santa Catarina (IFSC), São José, Brazil. She is currently a PhD student in the Postgraduate Program in Computer Science (PPGCC) at the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, and a research student in the Computing in Schools/INCoD/INE/UFSC initiative. She holds a Master's in Electrical Engineering (2012) from the Federal University of Santa Catarina (UFSC). Her research interests are computer education, machine learning, and critical thinking.

R.M. Martins is a Professor of Telecommunications at the Federal Institute of Santa Catarina (IFSC), São José, Brazil. He holds a PhD in Computer Science from the Federal University of Santa Catarina (UFSC) and an MSc in Telecommunications from the National Telecommunications Institute (INATEL). He also earned a PGC in Telecommunications Systems, Computer Networks, Systems Engineering, and an MBA in Systems Analysis and Telecommunications from ESAB, as well as a BSc in Telecommunications Engineering from the University of Southern Santa Catarina. His main research interests include Computer Education, Artificial Intelligence, and Telecommunications.

J.C.R. Hauck is a professor at the Federal University of Santa Catarina, and co-coordinator of the Software Quality Group and the initiative Computação na Escola. He holds a Ph.D. in Knowledge Engineering from the Federal University of Santa Catarina and his main research interests are in compuntin education and software engineering.

Appendix A

AQ1. What existing studies include assessing the development of critical thinking in the context of K-12 computing education?

Reference	Bibliographic reference	Brief description of the study	Computing education concepts	Educational stage (K-12)	Country
(Durak, 2020)	Durak, H. Y. (2020). The Effects of Using Different Tools in Programming Teaching of Secondary School Students on Engagement, Computational Thinking, and Reflective Thinking Skills for Problem Solving. *Technology, Knowledge and Learning*, 25.	Investigate the effects of programming practices on the development of critical thinking, reflective thinking, problem-solving skills and computational thinking.	Programming with Scratch and Alice	Elementary school (5th grade)	Türkiye
(Durak, *et al.*, 2019)	Durak, H.Y., Yilmaz, F.G., & Yılmaz, R. (2019). Computational Thinking, Programming Self-Efficacy, Problem Solving and Experiences in the Programming Process Conducted with Robotic Activities. *Contemporary Educational Technol*ogy, 10.	Investigate students' skill levels about computational thinking, critical thinking, programming and reflective thinking in problem-solving in the programming/ robotics learning process.	Robotics with Arduino	Middle school (6th and 7th grade)	Türkiye
(Duran and Şendağ, 2012)	Duran, M. and Şendağ, S. (2012). A Preliminary Investigation into Critical Thinking Skills of Urban High School Students: Role of an IT/STEM Program. *Creative Education*, 3.	Investigates the impact of an Information Technology program within the context of STEM education on critical thinking skills (analysis, evaluation, inference, induction, and deduction).	Integration of IT within STEM education	Middle school	USA
(Huang and Qiao, 2024)	Huang, X., Qiao, C. (2024). Enhancing Computational Thinking Skills Through Artificial Intelligence Education at a STEAM High School. Sci & Educ 33.	Investigate the teaching of artificial intelligence with the STEAM model with the aim of improving computational thinking skills.	Integrating STEM education and AI	High School	China
(Jiang and Li, 2019)	Jiang, B., & Li, Z. (2021). Effect of Scratch on computational thinking skills of Chinese primary school students. *Journal of Computers in Education*, 8.	Analyzes the effects of Scratch language learning on computational thinking skills such as critical thinking, creativity, algorithmic thinking, cooperativity and problem solving.	Programming with Scratch	Elementary school (5th grade)	China
(Jin, *et al.*, 2021)	Jin, Y., Sun, J., Ma, H. & Wang, X. (2021). The impact of different types of scaffolding in project-based learning on girls' computational thinking skills and self-efficacy. In: Proc. of the *10th International Conference of Educational Innovation through Technology*, Chongqing, China.	Investigates the effects of learning programming on the development of critical thinking, computational thinking skills and self-efficacy.	Programming with Scratch	Elementary school (3rd grade)	China

Continued on next page

Table – continued from previous page

Reference	Bibliographic reference	Brief description of the study	Computing education concepts	Educational stage (K-12)	Country
(Li *et al.*, 2023a)	Li, W., Huang, J.-Y., Liu, C.-Y., Tseng, J. C. R., & Wang, S.-P. (2023). A study on the relationship between student' learning engagements and higher-order thinking skills in programming learning. *Thinking Skills and Creativity*, 49.	Explore the relationship between levels of learning engagement (cognitive, emotional, and behavioral) and higher-order thinking skills such as computational thinking, problem-solving, and critical thinking engaged in programming learning.	Programming	Elementary, middle, and high school	China
(Li *et al.*, 2023b)	Li, X., Xie, K., Vongkulluksn, V., Stein, D., & Zhang, Y. (2023). Developing and testing a Design-Based Learning Approach to Enhance Elementary Students' Self-Perceived Computational Thinking. *Journal of Research on Technology in Education*, 55(2).	Investigate the development of computational thinking skills, such as critical thinking through a design-based learning approach.	Programming with Scratch	Elementary school (4th grade)	China
(Liu *et al.*, 2022)	Liu, H., Sheng, J., & Zhao, L. (2022). Innovation of Teaching Tools during Robot Programming Learning to Promote Middle School Students' Critical Thinking. *Sustainability*, 14, 6625.	Investigates the effectiveness of different teaching tools in promoting critical thinking engaged in robot programming learning.	Robotics	Middle school (8th grade)	China
(Negoro *et al.*, 2023)	Negoro, R. A., Rusilowati, A., & Aji, M. P. (2023). Scratch-Assisted Waves Teaching Materials: ICT Literacy and Students' Critical Thinking Skills. *Journal of Turkish Science Education*, 20(1)).	The study investigates the development of critical thinking skills in learning programming.	Programming with Scratch	High school	Indonesia
(Oluk and Korkmaz, 2016)	Oluk, A., & Korkmaz, Ö. (2016). Comparing Students' Scratch Skills with Their Computational Thinking Skills in Terms of Different Variables. *Int. Journal of Modern Education and Computer Science*, 11.	Investigate computational thinking skills, including critical thinking as part of teaching programming.	Programming with Scratch	Elementary school (5th grade)	Türkiye
(Qu *et al.*, 2023)	Qu, Z., Liu, J., Che, L., Su, Y., & Zhang, W. (2023). Research on the Application of Gamification Programming Teaching for High School Students' Computational Thinking Development. In 2023 IEEE 12th International Conference on Educational and Information Technology.	To investigate the effects of gamified programming on the development of computational thinking.	Programming with Minecraft, Scratch, and Python	High School	China

(Saritepeci and Durak, 2017)	Saritepeci, M., & Durak, Y. H. (2017). Analyzing the effect of block and robotic coding activities on computational thinking in programming education. In I. Koleva & G. Duman (Eds.), Educational Research and Practice (pp. 464-473). St. Kliment Ohridski University Press. (Chapter 49).	Investigate the effects of computational thinking skills, such as critical thinking, through the teaching of programming and robotics.	Programming with Scratch and robotics	Middle school (9th grade)	Türkiye
(Saritepeci, 2020)	Saritepeci, M. (2020). Developing Computational Thinking Skills of High School Students: Design-Based Learning Activities and Programming Tasks. *The Asia-Pacific Education Researcher*, 29(1).	Investigate the effects of programming and designer-based learning activities on developing computational thinking skills such as critical thinking.	Programming with Python, Scratch, and CodeCombat	Middle school (9th grade)	Türkiye
(Sun and Li, 2019)	Sun, Dan & Li, Yan. (2019). Improving Junior High School Students' Creativity, Critical Thinking and Learning Attitude in Minecraft Programming	Investigate game-based programming teaching in order to develop creativity and critical thinking.	Programming Python in the game Mine-craft	Middle school (7th grade)	China
(Tonbuloğlu and Tonbuloğlu, 2019)	Tonbuloğlu, B., & Tonbuloğlu, İ. (2019). The Effect of Unplugged Coding Activities on Computational Thinking Skills of Middle School Students. *Informatics in Education*, 18.	Investigate the effect of disconnected coding activities on computational thinking skills such as critical thinking.	Unplugged coding	Elementary school (5th grade)	Türkiye
(Wong and Cheung, 2020)	Wong, G. K.-W., & Cheung, H.-Y. (2020). Exploring children's perceptions of developing twenty-first-century skills through computational thinking and programming. *Interactive Learning Environments*, 28.	Investigate the impact of programming on creative thinking, critical thinking, and problem-solving.	Programming using Micro-soft Kodu (computer graphics environment)	Elementary school (4th, 5th and 6th grade)	China
(Yang and Chang, 2013)	Yang, C. Y.-T., & Chang, C.-H. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. *Computers & Education*, 68.	Investigate the impact of creating digital games on concentration, critical thinking skills and academic performance.	Programming based on ga-me develop-ment (inter-disciplinary with the bio-logy course)	Middle school (7th grade)	Taiwan

Appendix B
AQ2. How is critical thinking defined in the studies, and what skills are being assessed?

Reference	Definition of critical thinking in the context of the study	Assessed critical thinking skills						
		Interp-retation	Ana-lysis	Evalua-tion	Infe-rence	Expla-na-tion	Self-Regu-lation	Additional critical think-ing skills
(Durak, 2020)	Critical thinking is a subskill of Computational Thinking, which defines It as a high-level thinking skill in the most general sense (Korkmaz *et al.*, 2015).	-	X	X	X	X	X	-
(Durak, *et al.*, 2019)	Critical thinking is a subskill of Computational Thinking, which defines It as a high-level thinking skill in the most general sense (Korkmaz *et al.*, 2015).	-	X	X	X	X	X	-
(Duran and Şendağ, 2012)	A comprehensive way of thinking involves analysis, synthesis, and interpretation. It is a type of thinking that is closely associated with reasoning, decision-making, and problem-solving.	-	X	X	X	-	-	induction; deduction
(Huang and Qiao, 2024)	Critical thinking is a subskill of Computational Thinking, which defines It as a high-level thinking skill in the most general sense (Korkmaz *et al.*, 2015).	-	X	X	X		X	-
(Jiang and Li, 2019)	As the process of analyzing and assessing thinking with a view to improving it (Paul and Elder, 2007).	-	X	X	X	X	X	-
(Jin, *et al.*, 2021)	Critical thinking is a subskill of Computational Thinking, which defines It as a high-level thinking skill in the most general sense (Korkmaz *et al.*, 2015).	-	X	X	X	-	X	-
(Li *et al.*, 2023a)	Critical thinking is the ability to analyze information objectively and reflectively and make reasoned judgments to solve practical problems (Li *et al.*, 2021).	-	X	X	-	-	-	-
(Li *et al.*, 2023b)	Students' ability to clarify, analyze, and evaluate problems.	-	X	X	X	X	X	-
(Liu *et al.*, 2022)	Students' thinking about a series of interrelated critical questions, their ability to properly raise and answer critical questions, and their desire to utilize their critical thinking ability to solve problems.	X	-	X	-	-	-	recognition of assumptions; induction; deduction

(Negoro, *et al.*, 2023)	As a type of higher-order thinking that involves analytical thinking, which will then be able to form thinking frameworks such as memory to metacognition (Dwyer, 2014; Munawaroh, 2015; Permana, 2016; Prayogi, 2017; Rachmawati, 2015; Rahmawati, 2016).	-	X	X	X	-	-	-
(Oluk and Korkmaz, 2016)	Critical thinking is a subdimension of Computational Thinking skills, defining It as a high-level thinking skill in the most general sense (Korkmaz *et al.*, 2015).	-	X	X	X	X	X	-
(Qu *et al.*, 2023)	Critical thinking is a subdimension of Computational Thinking skills, defining It as a high-level thinking skill in the most general sense (Korkmaz *et al.*, 2015).	-	X	X	X	X	X	-
(Saritepeci and Durak, 2017)	As one of the processes for generating solutions to real-world problems, along with interdisciplinarity, creativity (Beer, Chiel and Drushel, 1999).	-	X	X	X	X	X	-
(Saritepeci, 2020)	As a way of dealing with an issue in a reflective and logical way which is related to beliefs and actions in a context (Ennis, 1993).	-	X	X	X	X	X	-
(Sun and Li, 2019)	The ability or skill by which the individual transcends his/her subjective self willfully to arrive rationally at conclusions (not necessarily favorable to him/her) that can be substantiated using valid information (Voskoglou and Buckley, 2012).	-	X	X	X	X	-	-
(Tonbuloğlu and Tonbuloğlu, 2019)	Critical thinking is a subskill of Computational Thinking, which defines It as a high-level thinking skill in the most general sense (Korkmaz *et al.*, 2015).	-	X	X	X	X	X	-
(Wong and Cheung, 2020)	Critical thinking involves the knowledge, skills, and attitudes necessary to think systematically, evaluate evidence, make reasoned judgments and decisions, and articulate these through clear explanations and justification of the process (Binkley *et al.*, 2012). It is integral to programming through Computational Thinking, as seen when children participate in the programming process, because solving any computational problem necessitates constant judgment and justification of the process, from algorithm design to testing and debugging (Lye and Koh, 2014).	-	X	X	X	-	X	-
(Yang and Chang, 2013)	Critical thinking involves carefully applying reason to determine whether a claim is true (Moore and Parker, 2009).	X		X				recognition of assumption; induction; deduction

Appendix C
AQ3. How are these critical thinking skills assessed?

Reference	Assessment of critical thinking			Critical thinking assessment instrument	Characteristics of assessment instrument	
	Paradigm	Method	Item format		Items on critical thinking	Response Scale
(Durak, 2020)	Self-assessment	Student self-assessment	Likert-scale	CTLS - adaptation for Secondary School Level(Korkmaz *et al.*, 2015)	Four items	5-point Likert scale
(Durak, *et al.*, 2019)	Self-assessment	Student self-assessment	Likert-scale	CTLS - adaptation for Secondary School Level (Korkmaz *et al.*, 2015)	Four items	5-point Likert scale
(Duran and Şendağ, 2012)	Formative assessment	Test	Multiple choice	TER - member of the California Critical Thinking Skills Test (CCTST) (Facione *et al.*, 2012)	35 items	NI
(Huang and Qiao, 2024)	Self-assessment	Student self-assessment	Likert-scale	CTS - adaptation for Chinese high school students (Korkmaz and Bai, 2019)	Four items	5-point Likert scale
(Jiang and Li, 2019)	Self-assessment	Student self-assessment	Likert-scale	CTS (Korkmaz *et al.*, 2017) was translated into Chinese and adapted for the target audience (Jiang and Li, 2019).	Five items	5-point Likert scale
(Jin, *et al.*, 2021)	Self-assessment	Student self-assessment	Likert-scale	CTS - adaptation for Chinese high school students (Korkmaz and Bai, 2019)	Five items	5-point Likert scale
(Li *et al.*, 2023a)	Self-assessment	Student self-assessment	Likert-scale	Critical Thinking Tendency Questionnaire CTTQ- adapted version of original (Lin *et al.*, 2019; Liu *et al.*, 2018; Chai *et al.*, 2015)	Five items	5-point Likert scale
(Li *et al.*, 2023b)	Self-assessment	Student self-assessment	Likert-scale	Computational Thinking Scales (CTS) - translated to Chinese (Li *et al.*, 2023)	Five items	5-point Likert scale
(Liu *et al.*, 2022)	Self-assessment	Student self-assessment	NI	Critical Thinking Tendency Scale CTTS-adapted version of original (Yu *et al.*, 2017)	25 items	NI
(Negoro, *et al.*, 2023)	NI	NI	NI	NI	NI	NI
(Oluk and Korkmaz, 2016)	Self-assessment	Student self-assessment	Likert-scale	CTLS - adaptation for Secondary School Level (Korkmaz *et al.*, 2015)	Four items	5-point Likert scale
(Qu *et al.*, 2023)	Self-assessment	Student self-assessment	Likert-scale	CTS (Korkmaz *et al.*, 2017).	Five items	5-point Likert scale

(Saritepeci and Durak, 2017)	Self-assessment	Student self-assessment	Likert-scale	CTLS adaptation for Secondary School Level (Korkmaz *et al.*, 2015)	Four items	5-point Likert scale
(Saritepeci, 2020)	Self-assessment	Student self-assessment	Likert-scale	CTLS adaptation for Secondary School Level (Korkmaz *et al.*, 2015)	Four items	5-point Likert scale
(Sun and Li, 2019)	Self-assessment	Student self-assessment	Likert-scale	Critical Thinking Questionnaire -CTQ modified version of California Critical Thinking Skills Test (CCTST) (Castle, 2006)	12 items	5-point Likert scale
(Tonbuloğlu and Tonbuloğlu, 2019)	Self-assessment	Student self-assessment	Likert-scale	CTLS - adaptation for Secondary School Level (Korkmaz *et al.*, 2015)	Four items	5-point Likert scale
(Wong and Cheung, 2020)	Self-assessment	Student self-assessment	Likert-scale	Assessment program for affective and social outcomes (2nd Ver-sion) (APASO-II) for secondary school (Education Bureau, n.d)	Five items	4-point Likert scale
(Yang and Chang, 2013)	Summative and Formative assessments	Test	Multiple-choice	Critical Thinking Test-Level I -CTT-Level I (Yeh,2003)	25 items	NI

NI - not informed or not identified

Appendix D
AQ4. How has the assessment approach been evaluated?

Reference	Sample size	Reliability		Validity	
		Analysis	Findings	Analysis	Findings
(Durak, 2020)	NI	Cronbach alpha coefficient = .866	Good internal consistency.	NI	NI
(Durak, 2019)	NI	Cronbach's alpha coefficient varied between .78 to .94 for the subscales	Acceptable internal consistency.	Factor-total correlation of all factors in the scale range between .48-.73	The validity level of the factors in the scale is moderate.
(Duran and Şendağ, 2012)[§]	NI	NI	NI	NI	NI
(Huang and Qiao, 2024)	NI	Cronbach's alpha coefficient =.85	Good internal consistency.	NI	NI
Jiang and Li, 2019)	NI	Cronbach alpha coefficient = .893	Good internal consistency.	Exploratory factor analysis: Kaiser-Meyer-Olkin (KMO)= .892, Bartlett values ($\chi 2$ =3673.36; SD= 406; p< .000), Confirmatory factor analysis: ($\chi 2$/df = 467.404/235 = 1.989; p= .000; CFI=.922>.9; TLI=.908; RMSEA= .047< .05)	The KMO and Bartlett's test results confirm the suitability of the factor analysis. The CFA shows an excellent fit with χ^2/df and RMSEA values and acceptable CFI and TLI indices, validating the model's factors against the data.
(Jin, *et al.*, 2021)	150 elementary school students	Cronbach's alpha coefficient =.838	Good internal consistency.	NI	NI
(Li *et al.*, 2023a)	NI	Cronbach alpha coefficient = .985	Excellent internal consistency.	NI	NI

(Li *et al.*, 2023b)	580 students	Cronbach's alpha coefficient varied between .79 to .88 for the subscales	Good internal consistency.	Exploratory factor analysis: Kaiser-Meyer-Olkin = .880, Significant Bartlett values ($\chi 2$ = 7727.897; SD = 406; $p < .001$) Confirmatory factor analysis: ($\chi 2$ = 1169.932 (SD = 362, N = 580), $\chi 2$ p-value < .001, CMIN/DF = 3.232, RMSEA = .062, SRMR = .044, GFI = .91, AGFI = .90, CFI = .95 and IFI = .97)	The KMO value above .50 and the significant Bartlett's test support the feasibility of factor analysis. The CFA demonstrates acceptable fit indices, indicating that the measurement model adequately fits the data.
(Liu *et al.*, 2022)	485 students from middle school, 8th grade	Cronbach alpha coefficient =.955	Excellent internal consistency.	Exploratory factor analysis: Kaiser-Meyer-Olkin = .955, Significant Bartlett values ($\chi 2$ =11384.868; df = 300; $p < .001$) Confirmatory factor analysis (GFI = .916; f2/df = 2.091; CFI = .974; NFI = .952; IFI = .975; RMSEA = .047).	These values indicated that the scale was suitable.
(Negoro, *et al.*, 2023)	68 students	Cronbach alpha coefficient =.597	Unacceptable internal consistency	Confirmatory factor analysis: ($\chi 2$ = 1285, NFI = .96 ,RMSEA = .01, SRMR = .044, GFI = .91, AGFI = .9, CFI = .93 and TLI/NNFI = .98)	These values indicated that the scale was suitable.
(Oluk and Korkmaz, 2016)	NI	Cronbach alpha coefficient =.809	Good internal consistency.	Confirmatory factor analysis: Maximum likelihood regression value = between .507 and .872 Item-test correlation coefficient = between .655 and .862	The values reflect a moderate to robust model fit, with item consistency and validity. The value range confirms the instrument's efficacy in measuring the targeted construct, indicating item alignment with the construct.
(Qu *et al.*, 2023)	NI	Cronbach alpha coefficient = .837	Good internal consistency.	NI	NI
(Saritepeci and Durak, 2017)	NI	Cronbach alpha coefficient =.853	Good internal consistency.	NI	NI
(Saritepeci, 2020)	NI	Cronbach alpha coefficient =.714	Good internal consistency.	NI	NI
(Sun and Li, 2019)	NI	Cronbach alpha coefficient =.84	Good internal consistency.	NI	NI

Continued on next page

Table – continued from previous page

Reference	Sample size	Reliability		Validity	
		Analysis	Findings	Analysis	Findings
(Tonbuloğlu and Tonbuloğlu, 2019)[§]	NI	NI	NI	NI	NI
(Wong and Cheung, 2020)[§]	NI	NI	NI	NI	NI
(Yang and Chang, 2013)[§]	NI	NI	NI	NI	NI

NI - not informed or not identified
[§]A assessment instrument was used without modifications that had been previously evaluated statistically.