Artículos de investigación

Instruments for Evaluating Computational Thinking: A Systematic Review

Instrumentos de evaluación del pensamiento computacional: una revisión sistemática

Milena Corrales-Álvarez
Universidad del Quindío, Colombia
mcorrales@uniquindio.edu.co
https://orcid.org/0000-0002-7433-6602
Lina Marcela Ocampo
Universidad del Quindío, Colombia
lmocampo@uniquindio.edu.co
https://orcid.org/0000-0001-6510-5880
Sergio Augusto Cardona Torres
Universidad del Quindío, Colombia
sergio_cardona@uniquindio.edu.co
https://orcid.org/0000-0002-1053-6322

Recepción: 30 Noviembre 2023
Aprobación: 08 Marzo 2024
Publicación: 24 Abril 2024

Acceso abierto diamante

Abstract

Computational Thinking (CT) is considered a key literacy skill in the digital age. It encompasses problem-solving, mathematical thinking, critical thinking, creativity, and communication. Since research on CT evaluation is in a consolidation phase, there is still a lack of systematic grouping of assessment instruments across different educational levels. This review aimed to identify the instruments used to measure CT, the evaluated skills, and the psychometric properties of these instruments. For such purpose, a systematic review of 52 articles published between 2012 and 2022 was conducted. The results revealed a significant growth in publications on the design and validation of CT measurement instruments in recent years. Over 80 % of the instruments demonstrated validity and reliability, particularly in terms of content validity, construct validity, and internal consistency. Furthermore, some instruments also evaluated affective and social skills, as well as attitudes, which enhanced the assessment of cognitive skills. However, the absence of contributions from Central and South American countries in the analyzed literature, along with the scarcity of instruments aimed at early childhood and teachers, highlights the need for further research into CT assessment in specific populations.

Keywords: Computational thinking, assessment instruments, psychometric properties, thinking skills, statistical methods.

Resumen

El pensamiento computacional (PC) es una nueva forma de alfabetización y se considera como una competencia clave para los ciudadanos de la era actual. Es un constructo compuesto que tiene relación con la resolución de problemas, el pensamiento matemático, el pensamiento crítico, la creatividad y la comunicación. La investigación sobre la evaluación del PC se encuentra en consolidación, sin embargo, se evidencia ausencia de agrupación sistemática de instrumentos de medición del PC en diferentes

Notas de autor
lmocampo@uniquindio.edu.co

niveles educativos. El objetivo de esta revisión consistió en identificar los instrumentos usados como herramientas para medir el PC, las habilidades evaluadas y las propiedades psicométricas de los instrumentos. Esta revisión sistemática presentó el análisis de 52 artículos encontrados del 2012 al 2022. Los resultados de la revisión demostraron un crecimiento significativo en las publicaciones relacionadas con el diseño y la validación de instrumentos de medición del PC en los últimos años. Se encontró que más del 80 % de los instrumentos presentaron evidencia de validez y confiabilidad, destacando la validez de contenido, la validez de constructo y la consistencia interna. Así mismo, en algunos instrumentos se consideraron la evaluación de habilidades afectivas, sociales y actitudes, lo cual enriquecía la valoración de las habilidades cognitivas. Sin embargo, se evidenció la ausencia de los países de Centro y Sur América en los artículos analizados sobre esta temática, al igual que la escasez de instrumentos dirigidos a la primera infancia y a los docentes. Estos hallazgos resaltan la necesidad de continuar investigando el PC desde la perspectiva de la evaluación en poblaciones específicas.

Palabras clave: Pensamiento computacional, instrumentos de evaluación, propiedades psicométricas, habilidades de pensamiento, métodos estadísticos.

Highlights

- Las escalas y las pruebas son los instrumentos más usados para evaluar el pensamiento computacional.
- La abstracción es la habilidad más evaluada en instrumentos de medición de pensamiento computacional.
- En Latinoamérica se evidencia ausencia de instrumentos de medición del pensamiento computacional.
- Turquía, USA y China lideran la publicación de artículos de evaluación del pensamiento computacional.
- La validez es la propiedad psicométrica más reportada en los instrumentos de medición del pensamiento computacional

Highllights

- Scales and tests are the most widely used instruments to evaluate computational thinking.
- Abstraction is the most evaluated skill in computational thinking measurement instruments.
- In Latin America, there is an absence of instruments for measuring computational thinking.
- Türkiye, USA, and China lead the publication of articles that evaluate computational thinking.
- Validity is the most reported psychometric property in computational thinking measurement instruments

1. INTRODUCTION

In recent decades, the scientific and educational communities have stressed the importance of incorporating Computational Thinking (CT) into curricula across all levels of education. Nonetheless, given its status as an emerging field, there is still no consensus on its definition and practical application. Consequently, a variety of approaches have been adopted to integrate it in students' learning process. This lack of a standardized definition for CT makes it challenging to design methods and tools for its evaluation [1]. Moreover, the rapid advancement of information and communication technologies underscores the need for 21st-century individuals to develop digital skills [2], [3], which brings about changes in how people think, act, communicate, and solve problems.

From a conceptual standpoint, various definitions have been put forth for CT. For instance, [4] serves as a starting point, defining it as the process of applying basic computer science principles to solve problems, design systems, and understand human behavior. [5], for his part, analyzed the different definitions of CT that have been proposed from the generic, operational, psychological-cognitive, and educational-curricular perspectives. In their literature review, [6] suggested classifying the definitions based on two approaches. The first approach is concerned with the relationship between computational concepts and programming, where authors [7]–[9] stand out. The second approach pertains to the set of competencies that students should develop, encompassing domain-specific knowledge and problem-solving skills. In this latter approach, the authors highlight the proposals of the International Society for Technology in Education (ISTE) and the Computer Science Teachers Association (CSTA) [10], along with [11] and [12].

Several initiatives have been developed to integrate CT into curricula, as well as tools for its accurate and reliable assessment. These tools include questionnaires [13]–[15], task-based tests [16], coding activities [17], and observation. Nevertheless, for widespread application across various educational levels, it is imperative to enhance the measurement of this construct using instruments with psychometric properties.

This systematic review was motivated by the need to measure CT and identify the instruments that have been designed for its assessment. Specifically, the goal is to analyze a set of bibliometric indicators and variables of interest, such as the type of instrument, number of items, target population, sample size, evidence of pilot testing, identification of skills/competencies, theoretical foundations, and psychometric properties.

1.1 Literature review

An examination of previous studies on the use of CT assessment tools provided valuable insights into the current landscape of this field. The search yielded 15 reviews, classified into different categories: scoping reviews [18]–[20], systematic or bibliometric mappings [21], [22], systematic reviews [6], [23]–[29], and meta-analyses [30], [31]. The analysis of these reviews underscored the need to identify the instruments employed for measuring CT, their psychometric properties, and the various variables associated with CT.

The retrieved review articles contributed to shedding light on research related to CT assessment. For instance, a scoping review of CT assessments in higher education [19] unveiled empirical studies focusing on CT assessments in post-secondary education. The majority of the analyzed instruments sought to measure CT skills by combining various dimensions, including concepts, practices, and perspectives. Among the skills frequently evaluated in these studies were algorithmic thinking, problem-solving, data handling, logic, and abstraction. However, it is worth noting that only four of these instruments provided sufficient evidence of their reliability and validity.

In another scoping review of empirical research on recent CT assessments [18], the authors classified features related to graphical or block-based programming, web-based simulation, robotics-based games, tests, and scales. Most studies in this review adopted a quasi-experimental approach, with only a few providing evidence of their validity. This review highlighted the need to carry out assessments aimed at different levels of higher-order thinking skills.

In their systematic review [6], evaluated 96 articles, considering variables such as educational level, subject matter domain, educational setting, and assessment tool. The findings emphasized the need for more assessments targeting high school students, college students, and teachers, in addition to evidence of the validity and reliability of the instruments. [25], for their part, analyzed 64 studies on CT measurement, identifying the psychometric properties of instruments primarily aimed at determining levels and measuring skills.

In the identified scoping reviews, mapping reviews, and meta-analyses, Scopus, ScienceDirect, ERIC, and Web of Science (WOS) were the most frequently consulted sources. As for the target population,Figure 1 shows that the most commonly selected population was students in K-12 educational settings. No reviews targeting teachers were found in the analysis.

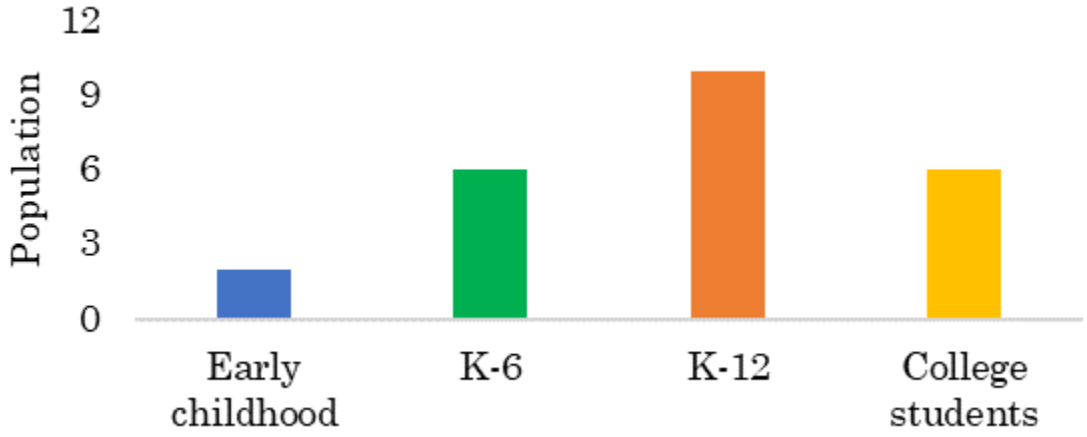

Figure 1.
Target population in the retrieved reviews.
Source: Own work.

Out of the 15 reviews, the one conducted by [22] included the largest number of sample articles (321 in total), while [23] had the smallest sample (15 articles). The sample sizes of the other reviews ranged from 17 to 101 articles. Figure 2 illustrates the number of articles included in the 15 reviews.

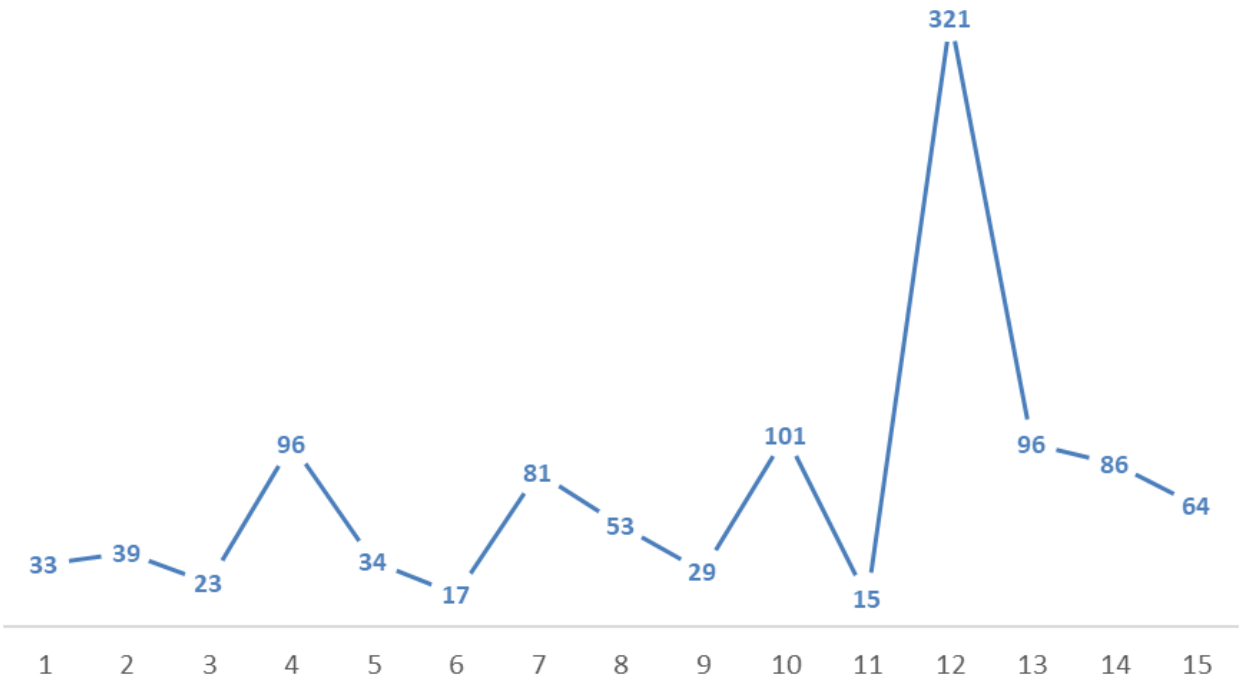

Figure 2.
Target population in the retrieved reviews.
Source: Own work.

Additionally, Table 1 lists the titles of the review articles, along with the country where the research was conducted.

Table 1
Identified review articles Source Own work

#	Title	Country
1	A scoping review of computational thinking assessments in higher education [19]	Canada
2	A Scoping Review of Empirical Research on Recent Computational Thinking Assessments [18]	Canada
3	Approaches to Assess Computational Thinking Competences Based on Code Analysis in K-12 Education: A Systematic Mapping Study [21]	Brazil
4	Assessing computational thinking: A systematic review of empirical studies [6]	USA
5	Computational thinking and academic achievement: A meta-analysis among students [32]	China
6	Computational thinking learning experiences, outcomes, and research in preschool settings: a scoping review of literature [20]	USA
7	Computational Thinking Through an Empirical Lens: A Systematic Review of Literature [27]	China
8	Computational thinking in primary education: a systematic literature review [26]	Italy
9	How to Develop Computational Thinking: A Systematic Review of Empirical Studies [28]	Türkiye
10	Mapping Computational Thinking through Programming in K-12 Education: A Conceptual Model based on a Systematic Literature Review [29]	Greece
11	Preschool children, robots, and computational thinking: A systematic review [23]	USA Uruguay
12	Trends and development in research on computational thinking [22]	Türkiye
13	Unleashing the Potential of Abstraction From Cloud of Computational Thinking: A Systematic Review of Literature [24]	China
14	Which way of design programming activities is more effective to promote K-12 students' computational thinking skills? A meta-analysis [31]	China
15	An investigation of the data collection instruments developed to measure computational thinking [25]	Türkiye

Source: Own wor

2. METHODOLOGY

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [33]. Such a protocol involves the following steps: (a) defining the research questions, objectives, and study variables (bibliometric indicators and variables of interest); (b) conducting a literature search (definition of search strings, period of analysis, inclusion and exclusion criteria, and sources of information, and study selection); and (c) identifying relevant articles.

2.1 Research questions, objectives, and variables

The following research questions were proposed for this systematic review:

Which studies have used instruments to assess CT?

What tools have been proposed for measuring CT?
What tools have been proposed for measuring CT?
What population is targeted for instrument application?
What constructs or skills are evaluated or measured?
What are the psychometric properties of the employed instruments?
hat statistical methods were employed for analyzing psychometric properties?
What factors are considered when measuring CT?

All these questions serve to identify the tools that have been used for CT assessment, their psychometric properties, and the evaluated skills. The proposed variables were divided into two categories: (i) bibliometric indicators, encompassing title, source of information, publication year, country, language, authors, journal, quartile, and the Scientific Journal Rank (SJR) and Journal Citation Reports (JCR) indices; and (ii) variables of interest, including type of instrument, number of items, age of target population, evaluated skills/ competencies, theoretical foundations, authors, sample size, pilot testing, and method for determining instrument validity and reliability.

2.2 Literature search

To conduct the search, the following eight search strings were formulated, incorporating key concepts such as computational thinking, measurement, and instruments, while adhering to the syntax required by the employed databases:

"Pensamiento Computacional" AND medición
"Computational Thinking" AND measuring
"Computational thinking" + "measuring instruments"
"Computational thinking" + "measurement"
"Computational thinking" + "measure instruments"
"Computational thinking" + "measurement tool"
"Computational thinking" AND ("measur* instruments" OR "measur* tool*")
"Computational thinking" AND (“assess” OR “validity” OR “reliability” OR “test” OR “scale”)

The search spanned from 2012 to 2022 because, as indicated by [34], this is when CT started to consolidate as a construct. For the search, five sources of information were consulted: ScienceDirect, EBSCO Discovery, Scopus, WOS, and Springer.

Regarding exclusion and inclusion criteria, only research articles and reviews were considered for analysis, while publications in book formats, posters, conference proceedings, or articles that did not employ a specific instrument for measuring CT were excluded.

2.3 Identified articles

Initially, the search yielded 439 articles. After removing duplicates, 204 articles remained. Following further screening for relevance, 115 articles were retained. Finally, by applying exclusion criteria, a total of 52 articles were selected for the systematic review. Figure 3 provides a summary of the articles identified at each stage of the search process, which was conducted following the PRISMA statement.

Identification	EBSCO Discovery	Scopus	ScienceDirect	Springer	WOS	Total
	131	146	47	45	70	439
Screening	Duplicate records	Records after duplicates were removed	Records excluded after screening	Records (after exclusions)		
	235	204	89	115		
Eligibility	Documents excluded based on exclusion criteria	Records (after exclusions)				
	48	67				
Included	Documents excluded for relevance	Records (after exclusions)				
	15	52				

Figure 3.
Systematic review flowchart.
Source: Adapted from the PRISMA statement [35].

3. RESULTS AND DISCUSSION

3.1 Analysis of bibliometric indicators

3.1.1 Consulted databases

The majority of the articles (approximately 63.4 %) were found in Scopus. Figure 4 shows the number of articles retrieved from each consulted database, with several appearing in multiple sources.

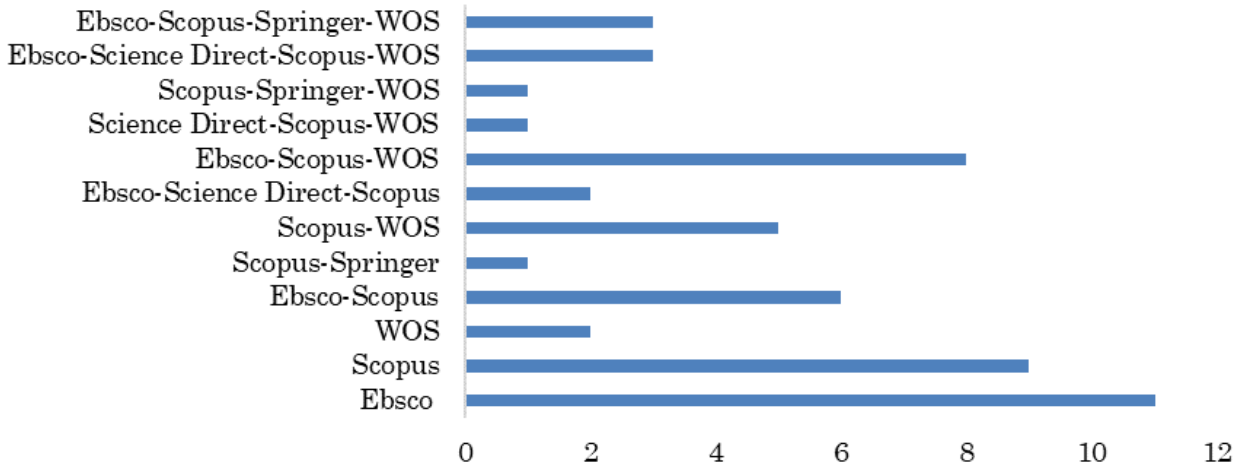

Figure 4.
Consulted databases.
Source: Own work.

3.1.2 Title keywords

According to the analysis, computational thinking was the most prevalent term in the titles of the examined articles, often accompanied by valid, scale, evaluate, and test, all of which allude to important features of the measurement instruments.

3.1.3 Publication year

As mentioned earlier, the search spanned from 2012 to 2022. Remarkably, none of the five sources yielded publications related to instrument construction before 2017. Figure 5 depicts the increase in the number of publications dedicated to CT measurement instruments throughout the analyzed period.

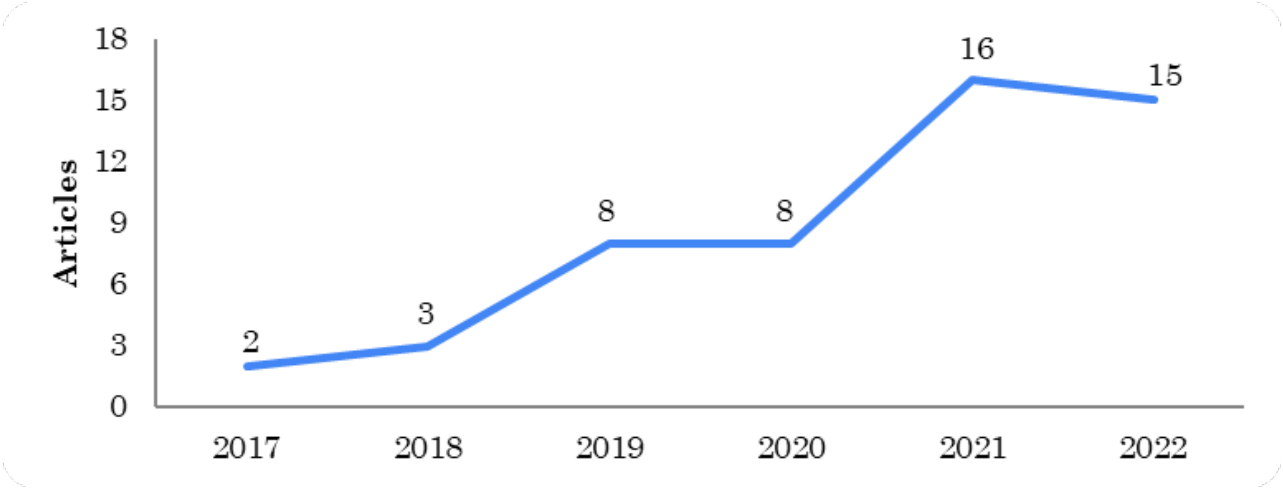

Figure 5.
Publication year of articles into CT measurement instruments.
Source: Own work.

3.1.4 Country where the research was conducted

Türkiye was the country with the highest number of articles—ten in total—followed by China and the United States, with eight and seven articles, respectively. Only one study was carried out in Latin America, specifically in Venezuela. Figura 6 displays the distribution of articles by country.

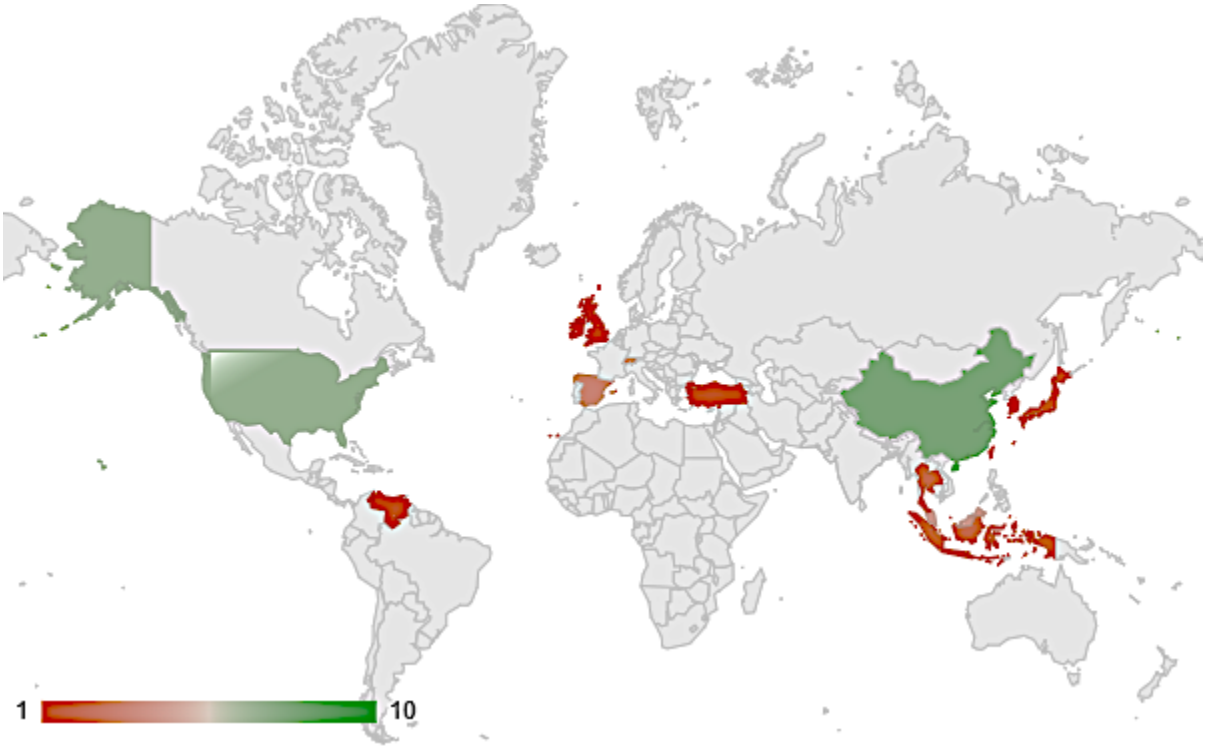

Figure 6.
Distribution of articles by country.
Source: Own work.

3.1.5 Language

English was the most prevalent language, with 90 % of the articles being written in this language. Spanish accounted for 4 %, Turkish 4 %, and Japanese 2 %.

3.1.6 Authors

Table 2 presents the most prominent authors based on the number of published articles.

Table 2.
Most prominent authors.

Author	Published articles
Yan Li [16], [36]	2
Juan Carlos Pérez González [37], [38]	2
Jungwon Cho [39], [40]	2
Saralah Sovey and Mohd Effendi [41], [42]	2
Siu Cheung Kong [43], [44]	2
Barbara Bruno, Laila El-Hamamsy, and Estefanía Martín-Barroso [15], [45]	2
Kamisah Osman [41], [42], [46]	3
Özgen Korkmaz [46]–[48]	3
Jessica Dehler Zufferey [15], [45], [49]	3
Marcos Román González [15], [37], [38], [50]	4

Source: Own work

Also, one important aspect considered in the analysis was the most cited authors in the analyzed articles (see Figure 7). Prominent authors include Brennan and Resnick, Selby and Woollard, the International Society for Technology in Education (ISTE), the Computer Science Teacher Association (CSTA), Román et al., and Korkmaz et al. The latter authors are notable references because the instruments they proposed—the Computational Thinking Test (CTt) and the Computational Thinking Scale (CTS)—are frequently employed for CT measurement.

Figure 7.
Most cited authors.
Source: Own work

3.1.7 Journal, quartile, and JCR indices

In the analysis, two impact indicators evaluating the excellence of published content were employed. JCR, on the one hand, primarily focuses on citation counts, providing the impact factor and quartile of a journal. SJR, on the other hand, considers the quality, relative importance of citations, and quartile of a journal. According to the findings, the Journal of Educational Computing Research and Education and Information Technologies stood out as the most productive journals, with six and four publications, respectively. Regarding the two indices and quartiles, 21 % of the journals had no classification in any of the indices. Information on each journal is provided in Table 3.The most prominent journals, ranked by the number of CT-related publications, are listed in [22] .

Table 3.
Information about the journals.

Journal	Number of articles	Impact factor	JCR quartile	SJR indicator	SJR quartile
Journal of Educational Computing Research	6	0.14	Q1	1.28	Q1
Education and Information Technologies	4	0.23	Q1	1.06	Q1
Computer Science Education	2	0.29	Q2	1	Q1
Computers in Human Behavior	2	0.033	Q1	2.17	Q1
European Journal of Educational Research	2	No	No	0.31	Q3
Frontiers in Psychiatry	2	No	No	1.28	Q1
AERA Open	1	0.26	Q2	0.86	Q1
British Journal of Educational Technology	1	0.019	Q1	1.87	Q1
Computers & Education	1	0.054	Q1	3.68	Q1
Computers in Education	1	0.094	Q1	1.04	Q1
Computers in the Schools	1	0.38	Q2	0.92	Q1
Current Psychology	1	0.41	Q2	0.51	Q2
Revista Digital del Doctorado en Educación de la Universidad Central de Venezuela	1	No	No	No	No
Espacios	1	No	No	0	No
Hipotenusa: Journal of Mathematical Society	1	No	No	No	No
Open Conference on Computers in Education	1	No	No	No	No
Informatics in Education	1	0.22	Q1	0.96	Q1
Interactive Learning Environments	1	0.096	Q1	1.17	Q1
International Journal of Advanced Computer Science and Applications (IJACSA)	1	No	No	0.28	Q3
International Journal of Child-Computer Interaction	1	No	No	1.03	Q1
International Journal of Educational Methodology	1	No	No	No	No
International Journal of Learning, Teaching and Educational Research	1	No	No	No	No
International Journal of Recent Technology and Engineering (IJRTE)	1	No	No	No	No
International Journal on Informatics Visualization	1	No	No	0.18	Q4
Journal of Computer and Mathematics Education	1	No	No	No	No

Journal of Research on Technology in Education	1	0.28	Q1	1.08	Q1
Journal of Science Education and Technology	1	0.16	Q1	1.15	Q1
Mathematics Teaching Research Journal	1	0.26	Q2	0.15	Q4
Pacific Rim Psychology	1	0.62	Q3	0.5	Q2
Participatory Educational Research	1	No	No	0.25	Q3
Information and Technology in Education and Learning (ITEL)	1	No	No	No	No
Revista Iberoamericana de Evaluación Educativa	1	No	No	No	No
Sosyal Bilimler Enstitüsü Dergisi	1	No	No	No	No
Sustainability	1	0.48	Q2	0.66	Q1
Technology, Knowledge and Learning	1	No	No	1.14	Q1
The All Ireland Journal of Teaching and Learning in Higher	1	No	No	No	No
Journal of the Human and Social Sciences Researches	1	No	No	No	No
Thinking Skills and Creativity	1	0.23	Q1	1.16	Q1
Transactions on Computing Education	1	0.55	Q3	0.99	Q1

Source: Own work.

Figure 8 summarizes the quartiles assigned to the journals in which the articles were published. A total of 40 different journals were identified, of which 47.5 % have been classified in a JCR quartile and 65 % have been classified with the SJR indicator.

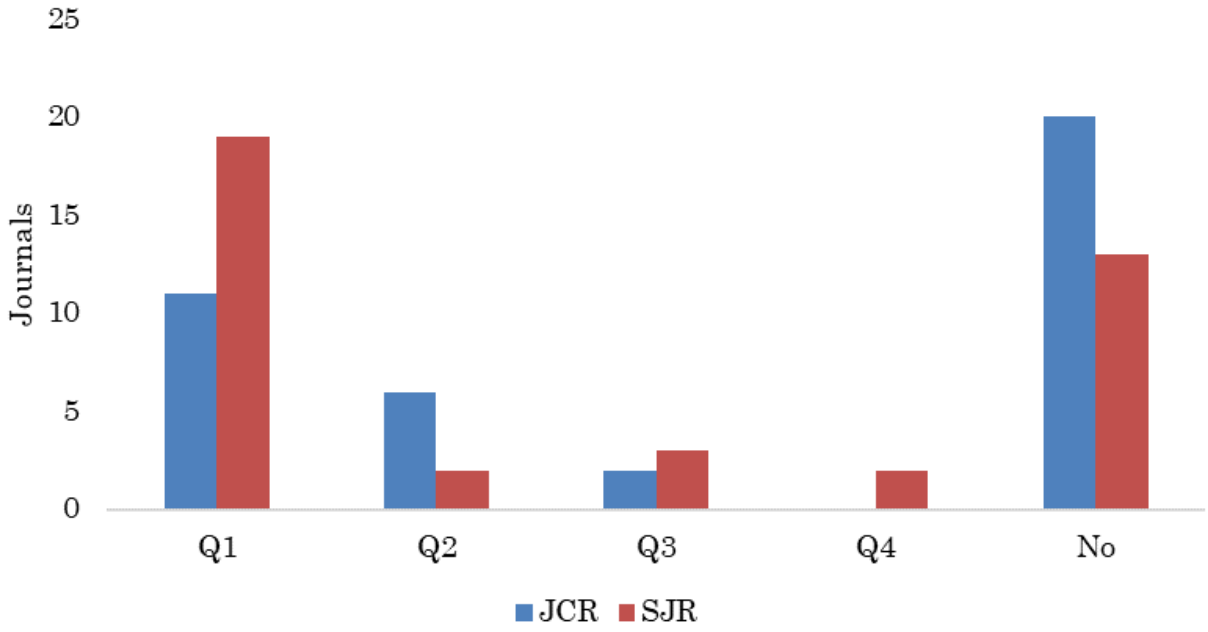

Figure 8.
Journals classified into quartiles.
Source: Own work.

3.2 Analysis of the variables of interest

This systematic literature review included 52 articles, of which only 40 introduced new instruments. The remaining 12 articles examined adaptations of the latter. Table 4 lists the instruments, the reference to the original instrument, and the reference to the adapted version.

The evidence reported above regarding the CTS suggests that all the adapted versions of this instrument underwent thorough validation of their psychometric properties. In the case of the CTt, there have been some linguistic adaptations and changes to a number of items.

3.2.1 Type of tool

The 40 CT assessment tools analyzed in this paper can be classified as shown in Figure 9.This classification is based on the name given by each author in their article. Scales were the most common format (28 %), followed by assessments (22.5 %) and tests (22.5 %).

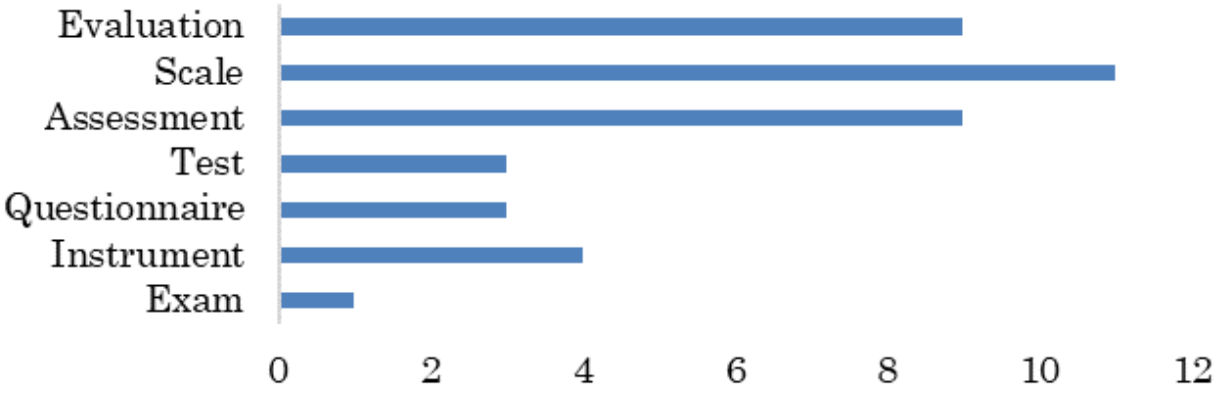

Figure 9.
Type of tools.
Source: Own work.

3.2.2 Number of items

In these articles, 5.8 % of the instruments have up to ten items; 73 %, between 10 and 30 items; 9.7 %, between 31 and 40; and 11.5 %, more than 40.

Table 4.
Instruments and references to original and adapted versions.

Instrument	Original reference	Adapted reference	Adaptation
Holistic Assessment of Computational Thinking (Hi-ACT)	[51]	[52]	This article confirmed the psychometric properties of the instrument (validity and reliability). In total, 41 items were removed from the instrument. Ten constructs were evaluated: abstraction, algorithmic thinking, decomposition, debugging, generalization, evaluation, problem-solving, teamwork, communication, and spiritual intelligence.
Computational Thinking Scale (CTS)	[46]	[50]	This paper confirmed the construct validity of the CTS and its five dimensions. Two factors were identified: (1) creative thinking ability, cooperativity, and critical thinking skills and (2) algorithmic thinking.
		[53]	This paper confirmed the construct validity of the CTS and its five dimensions. The wording of six questions in the scale was adapted because they were written from a negative perspective.
		[47]	This study confirmed the construct validity of the CTS and its five dimensions. In the process of translating the scale, the authors determined the consistency between the structures in the original language and those in Chinese.
		[54]	This study confirmed the psychometric properties of the instrument (validity and reliability). Back-translation was used to verify the consistency between the structures in the original language and those in Chinese. The wording of items about problem-solving was changed.
		[55]	This paper confirmed the construct validity of the CTS and its five dimensions. Two items were removed from the *creativity*dimension, one from *critical thinking*, and three from *problem-solving*.
		[56]	This study confirmed the construct validity of the CTS and its five dimensions. Two factors were identified: (1) creative thinking ability, cooperativity, and critical thinking skills and (2) algorithmic thinking.
Computational Thinking Test (CTt)	[37], [38]	[50]	Rasch scalability was applied as a technique to validate the psychometric properties of the skills in the CTt. Likewise, the Item Response Theory (IRT) was employed to verify the objectivity of the test. The CTt was not modified.

		[53]	This study confirmed the psychometric properties of the instrument (validity and reliability). Back-translation was used to verify the consistency between the structures in English and those in Turkish.
		[37]	This article examined the predictive validity of the CTt with respect to academic performance and learning on a virtual platform (code.org).
		[57]	This study confirmed the reliability of the instrument. Expert judgement was applied for the validation, and the final version had 28 items.
		[58]	The Item Response Theory (IRT) was used to verify the objectivity of the test and the difficulty of the items. The final version had 24 items because some questions about conditionals and loops were left out.
		[59]	Rasch scalability was applied as a technique to validate the psychometric properties of the skills in the CTt. The final version had 28 items because some questions about conditionals and loops were left out.
Computational Thinking Disposition Instrument (CTDI)	[42]	[41]	This study confirmed the psychometric properties of the instrument (validity and reliability). Nine items were removed from the *cognitive*and *affective*dimensions.

Table 4. Instruments and references to original and adapted versions.

3.2.3 Study population

In terms of study population, 25 % of the papers focused on college students, 3.8 % on teachers in training, 35 % on high school students, 21 % on primary students, and 6% on early childhood education. Among these publications, 5.7 % are about teachers. Figure 10 presents the frequency of each type of study population.

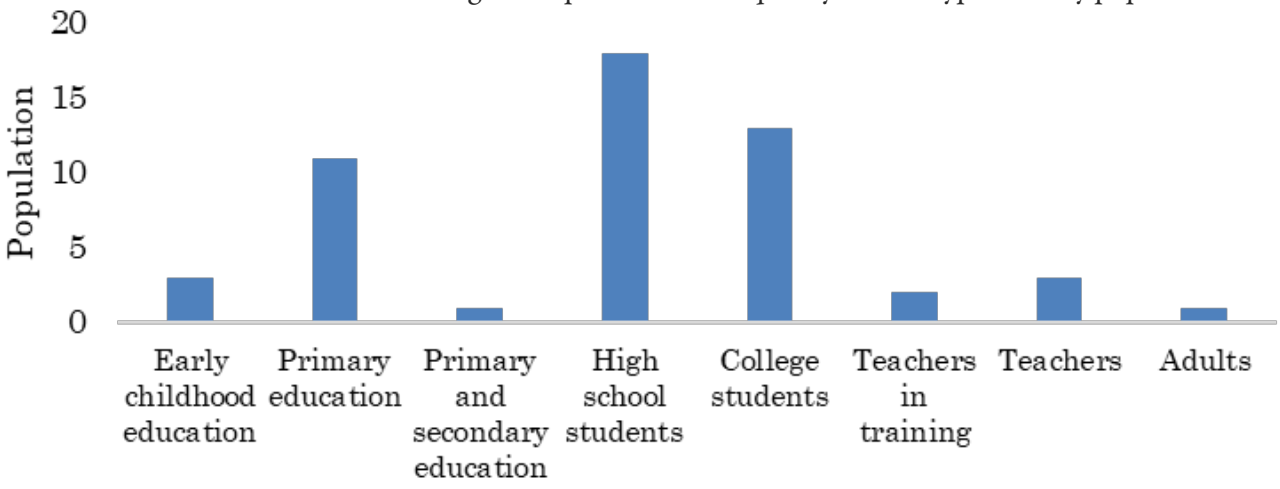

Figure 10.
Study populations.
Source: Own work.

3.2.4 Skills/competencies assessed in CT

The diversity of definitions of CT indicates that the articles have addressed this construct from the perspectives of different skills or competencies. The most frequent skills/competencies they have discussed are abstraction, logarithmic thinking, problem-solving, decomposition, debugging, algorithms, and modularizing. Based on these 40 instruments, abstraction, logarithmic thinking, problem-solving, debugging, modularizing, and affective competencies have been evaluated since 2017. Decomposition was included in 2018. Some of the instruments assess cognitive skills along with affective and social skills, as well as attitudes. Table 5 presents the constructs assessed in each of the 40 instruments.

Table 5.
CT instruments and assessed constructs.

#	Instruments / Constructs assessed	Abstraction	Algorithmic thinking	Problem-solving	Decomposition	Debugging	Algorithms	Modularizing
1	Holistic Assessment of Computational Thinking (Hi-ACT)	X	X	X	X	X		
2	Programming-oriented Computational Thinking Scale (P-CTS)	X			X			
3	Computational Thinking Skills (CTS) scale		X	X				
4	Computational Thinking Scale (CTS)		X	X				
5	CT Skill Level Scale	X	X	X	X			
6	Tufts Assessment of Computational Thinking in Children-KIBO robot version (TACTIC-KIBO)	X				X	X	X
7	Computer-based assessment			X				
8	Computational Thinking Skills Scale		X	X				
9	CT Test (CTt)	X		X		X		X
10	Computational Thinking Self-Efficacy Scale	X			X			
11	Computational Thinking Disposition Questionnaire							
12	Computational Thinking Assessment of Chinese Elementary School Students (CTA-CES)	X	X		X			
13	Evaluación del PC basado en la resolución de problemas complejos [*CT evaluation based on complex problem-solving*]			X				

14	Computational Thinking Disposition Instrument (CTDI)											
15	Generic test to assess CT practices	X		X								X
16	Assessment using card-based games					X				X		X
17	Triangle examination using Bebras Challenge									X		X
18	Assessment of Computational Thinking in Early Childhood (TechCheck)							X		X		X
19	Competent CT Test (cCTt)									X		
20	Computational Thinking Scale (CTS) for computer literacy education	X		X		X						
21	Computational Thinking Competency Assessment (CTCA)			X						X		
22	Computational Thinking Test Tool from Existing Models									X		
23	Computational Thinking Concepts Test for Primary Education Adopting an ECD Approach									X		
24	Computational Thinking Concepts Assessment	X			X			X				X
25	Mathematical Computational Thinking Skill Test	X		X		X						
26	Algorithmic Thinking Test for Adults (ATTA)	X		X		X		X				
27	CT test, questionnaire, and interview											
28	Questionnaire to assess CT components in teachers				X							
29	College Students' Computational Thinking Multidimensional Test	X				X				X		
30	Computer Programming Self-Efficacy Scale (CPSES)							X		X		

31	Instrument Test for Computational Thinking Skills Based on the Realistic Mathematics Education (RME) Approach				X		X	
32	Computational Thinking Scale (CTS)		X	X				
33	Questionnaire of Computational Thinking (QCT)	X			X		X	
34	Scale of Self-Efficacy Perception Towards Teaching Computational Thinking	X	X	X				
35	Teacher Beliefs about Coding and Computational Thinking (TBaCCT)	X		X	X		X	
36	Teacher Efficacy and Attitudes Towards STEM for Teaching Computational Thinking (T-STEM-CT)							
37	Assessment Tool for Measuring Computational Thinking Skills	X	X		X	X		
38	Early assessment				X	X		
39	Computational Thinking Test for Elementary School Students (CTT-ES)	X			X		X	
40	Beginners' CT test (BCTt)							
		17	13	12	16	9	14	7

Source: Own work.

3.2.5 Validity and reliability

It was found that 87 % of the instruments showed evidence of validity; and 69 %, evidence of reliability. Figures 11 and 12 display the types of validity and reliability reported in the articles.

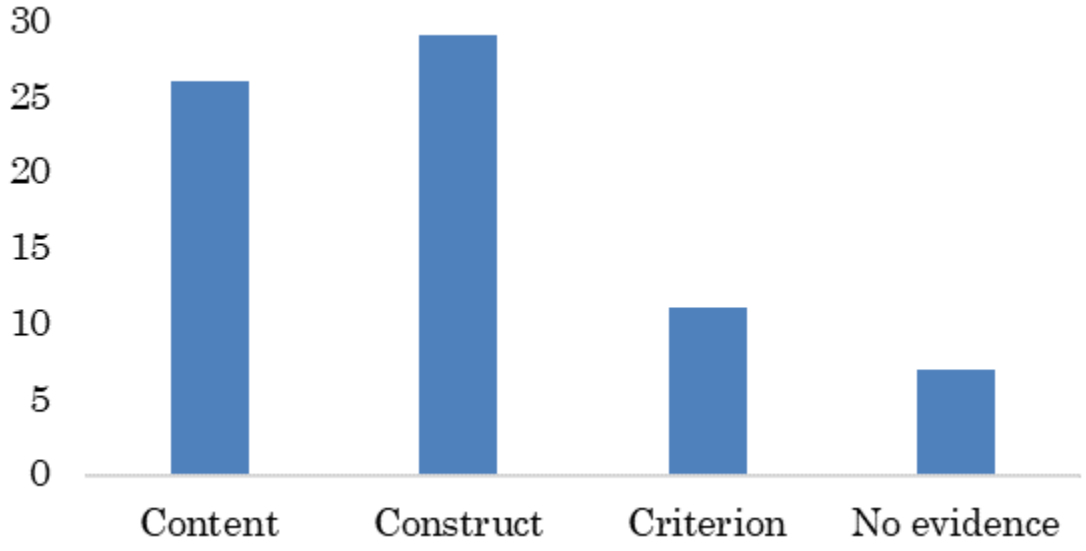

Figure 11.
Validity criterion.
Source: Own work.

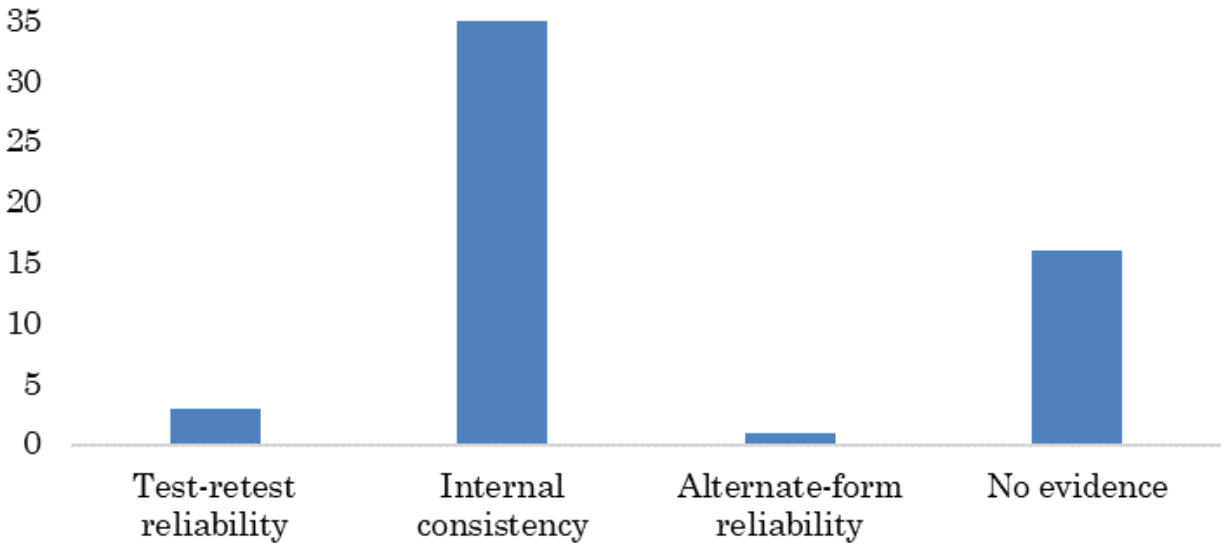

Figure 12.
Reliability criterion
Source: Own work.

Table 5.
CT instruments and assessed constructs.

#	Instruments / Constructs assessed	Abstraction	Algorithmic thinking	Problem-solving	Decomposition	Debugging	Algorithms	Modularizing
1	Holistic Assessment of Computational Thinking (Hi-ACT)	X	X	X	X	X		
2	Programming-oriented Computational Thinking Scale (P-CTS)	X			X			
3	Computational Thinking Skills (CTS) scale		X	X				
4	Computational Thinking Scale (CTS)		X	X				
5	CT Skill Level Scale	X	X	X	X			
6	Tufts Assessment of Computational Thinking in Children-KIBO robot version (TACTIC-KIBO)	X				X	X	X
7	Computer-based assessment			X				
8	Computational Thinking Skills Scale		X	X				
9	CT Test (CTt)	X		X		X		X
10	Computational Thinking Self-Efficacy Scale	X			X			
11	Computational Thinking Disposition Questionnaire							
12	Computational Thinking Assessment of Chinese Elementary School Students (CTA-CES)	X	X		X			
13	Evaluación del PC basado en la resolución de problemas complejos [*CT evaluation based on complex problem-solving*]			X				

14	Computational Thinking Disposition Instrument (CTDI)											
15	Generic test to assess CT practices	X		X								X
16	Assessment using card-based games						X			X		X
17	Triangle examination using Bebras Challenge									X		X
18	Assessment of Computational Thinking in Early Childhood (TechCheck)							X		X		X
19	Competent CT Test (cCTt)									X		
20	Computational Thinking Scale (CTS) for computer literacy education	X		X			X					
21	Computational Thinking Competency Assessment (CTCA)			X						X		
22	Computational Thinking Test Tool from Existing Models									X		
23	Computational Thinking Concepts Test for Primary Education Adopting an ECD Approach									X		
24	Computational Thinking Concepts Assessment	X			X			X				X
25	Mathematical Computational Thinking Skill Test	X		X			X					
26	Algorithmic Thinking Test for Adults (ATTA)	X		X			X	X				
27	CT test, questionnaire, and interview											
28	Questionnaire to assess CT components in teachers				X							
29	College Students' Computational Thinking Multidimensional Test	X					X			X		
30	Computer Programming Self-Efficacy Scale (CPSES)							X		X		

31	Instrument Test for Computational Thinking Skills Based on the Realistic Mathematics Education (RME) Approach				X		X	
32	Computational Thinking Scale (CTS)		X	X				
33	Questionnaire of Computational Thinking (QCT)	X			X		X	
34	Scale of Self-Efficacy Perception Towards Teaching Computational Thinking	X	X	X				
35	Teacher Beliefs about Coding and Computational Thinking (TBaCCT)	X		X	X		X	
36	Teacher Efficacy and Attitudes Towards STEM for Teaching Computational Thinking (T-STEM-CT)							
37	Assessment Tool for Measuring Computational Thinking Skills	X	X		X	X		
38	Early assessment				X	X		
39	Computational Thinking Test for Elementary School Students (CTT-ES)	X			X		X	
40	Beginners' CT test (BCTt)							
		17	13	12	16	9	14	7

Source: Own work.

Table 6.
Instruments, articles, samples, and evidence of validity and reliability.

				Validity				Reliability		
Year	Instrument	Article title	Sample	Content	Construct	Criterion	No evidence	Test-retest	Internal consistency	Alter form
2017	Computational Thinking Scale (CTS)	A validity and reliability study of the computational thinking scales (CTS)	580		X				X	
2017	CT Test (CTt)	Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test	1521			X			X	
2018	Computational Thinking Skills (CTS) scale	A valid and reliable tool for examining computational thinking skills		X	X			X	X	
2018	Computational Thinking Test (CTt)	Can computational talent be detected? Predictive validity of the Computational Thinking Test	314			X				
2018	Scale of Self-Efficacy Perception Towards Teaching Computational Thinking	The scale of self-efficacy perception towards teaching computational thinking: a validity and reliability study	378		X				X	
2019	Holistic Assessment of Computational Thinking (Hi-ACT)	A proposal for holistic assessment of computational thinking for undergraduate: Content validity	0	X						

2019	Computational Thinking Scale (CTS)	Adapting computational thinking scale (CTS) for Chinese high school students and their thinking scale skills level	1015		X			X	X
2019	Computational Thinking Self-Efficacy Scale	Computational thinking self-efficacy scale: Development, validity, and reliability	319	X	X				X
2019	Questionnaire to assess CT components in teachers	Computational thinking for preservice teachers in Thailand: A confirmatory factor analysis.	747	X					
2019	Computer Programming Self-Efficacy Scale (CPSES)	Developing the Computer Programming Self-Efficacy Scale for Computer Literacy Education	106		X				X
2019	Computational Thinking Scale (CTS)	Development of Computational Thinking Scale: Validity and Reliability Study	426	X	X				X
2019	Triangle examination using Bebras Challenge	Multivocal Challenge Toward Measuring Computational Thinking: Bebras Challenge Versus Computer Programming	150				X		
2019	Computational Thinking Test Tool from Existing Models	Toward developing a real-world computational thinking test tool from existing models	204	X					X
2020	Holistic Assessment of Computational Thinking (Hi-ACT)	A Pilot Study of an Instrument to Assess Undergraduates' Computational Thinking Proficiency	548		X				X

2020	Programming-oriented CTS (P-CTS)	A Valid and Reliable Scale for Developing Programming-Oriented Computational Thinking	360		X			X
2020	Adaption of the Computational Thinking Test	Adaption of the computational thinking test into Turkish	502				X	X
2020	Computational Thinking Skills Scale	The Development of Computational Thinking Skills Scale: Validity and Reliability Study	254	X	X			X
2020	Computational Thinking Disposition Questionnaire	Development and Predictive Validity of the Computational Thinking Disposition Questionnaire	907		X			
2020	Computational Thinking Assessment of Chinese Elementary School Students (CTA-CES)	Development and Validation of Computational Thinking Assessment of Chinese Elementary School Students	280	X	X	X		X
2020	Assessment of Computational Thinking in Early Childhood (TechCheck)	TechCheck: Development and Validation of an Unplugged Assessment of Computational Thinking in Early Childhood Education	768			X		X
2020	Computational Thinking Test	Analysis of a Novel Computational Thinking Test in First Year Undergraduate Computer Science Course	292				X	
2021	Adapted Computational Thinking Test (CTt)	A comprehensive assessment of secondary school students computational thinking skills.	328				X	X

2021	Computational Thinking Concepts Assessment	A principled approach to designing computational thinking concepts and practices assessments for upper elementary grades	5698	X	X			X
2021	Assessment Tool for Measuring Computational Thinking Skills	An alternative approach for measuring computational thinking: Performance-based platform	156	X	X	X		X
2021	Adapted Computational Thinking Test (CTt)	Assessing computational thinking abilities among Singapore secondary students: a Rasch model measurement analysis	153	X				X
2021	Computer-based assessment	Beyond Programming: A Computer-Based Assessment of Computational Thinking Competency	119		X			X
2021	CT test, questionnaire, and interview	Computational thinking evaluation tool development for early childhood software education	0	X				
2021	Early assessment	Design and validation of learning trajectory-based assessments for computational thinking in upper elementary grades	144	X				

2021	Instrument Test for Computational Thinking Skills Based on the Realistic Mathematics Education (RME) Approach	Development of Instrument Test Computational Thinking Skills IJHS/JHS Based RME Approach	102	X				X	
2021	Adapted Computational Thinking Test (CTt)	Computational thinking in elementary and middle school students	176	X				X	
2021	Evaluación del PC basado en la resolución de problemas complejos [*CT assessment based on complex problem-solving*]	Evaluar el PC mediante Resolución de Problemas: Validación de un Instrumento de Evaluación. (Spanish)	38		X	X			
2021	Questionnaire of Computational Thinking (QCT)	Examination of Turkish Middle School STEM Teachers' Knowledge about Computational Thinking and Views Regarding Information and Communications Technology	121				X		
2021	Generic test to assess CT practices	Item response analysis of computational thinking practices: Test characteristics and students' learning abilities in visual programming contexts	13956	X					X

2021	Assessment using card-based games	Measuring coding ability in young children: relations to computational thinking, creative thinking, and working memory	15	X	X	X	X	X
2021	Teacher Beliefs about Coding and Computational Thinking (TBaCCT)	Measuring teacher beliefs about coding and computational thinking	245		X			
2021	Teacher Efficacy and Attitudes Towards STEM for Teaching Computational Thinking (T-STEM-CT)	Measuring in-service teacher self-efficacy for teaching computational thinking: development and validation of the T-STEM CT	330		X			X
2021	Computational Thinking Scale (CTS) for computer literacy education	The Computational Thinking Scale for Computer Literacy Education	388		X			X
2022	Mathematical Computational Thinking Skill Test	Analysis of Content Validity on Mathematical Computational Thinking Skill Test for Junior High School Student Using Aiken Method	7	X				
2022	Algorithmic Thinking Test for Adults (ATTA)	Assessing Computational Thinking: Development and Validation of the Algorithmic Thinking Test for Adults	289	X				X

2022	Tufts Assessment of Computational Thinking in Children-KIBO robot version (TACTIC-KIBO)	Assessing young Korean children's computational thinking: A validation study of two measurements	450	X	X	X		X
2022	Beginners' CT test (BCTt)	Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt)	575				X	
2022	Computational Thinking Test (CTt)	Computational Thinking Assessment – Towards More Vivid Interpretations	202				X	
2022	College Students' Computational Thinking Multidimensional Test	Developing College students' computational thinking multidimensional test based on Life Story situations	450		X			
2022	Computational Thinking Test for Elementary School Students (CTT-ES)	Development and Validation of the Computational Thinking Test for Elementary School Students (CTT-ES): Correlate CT Competency With CT Disposition.	631	X	X	X		X
2022	Adapted self-report scale	Development of the Japanese Version of the Computational Thinking Scales for First-Year University Students in Humanities	511		X	X		X

2022	Scale of CT Skill Levels	Evaluation and developmental suggestions on undergraduates' computational thinking: a theoretical framework guided by Marzano's new taxonomy	737		X			X
2022	Computational Thinking Disposition Instrument (CTDI)	Exploratory and Confirmatory Factor Analysis for Disposition Levels of Computational Thinking Instrument Among Secondary School Students	500	X	X			X
2022	Computational Thinking Disposition Instrument (CTDI)	Gender differential item functioning analysis in measuring computational thinking disposition among secondary school students	500	X	X			X
2022	Thai Self-Rating Version of the Computational Thinking Scale	Reliability and Construct Validity of Computational Thinking Scale for Junior High School Students: Thai Adaptation	3241		X	X		X
2022	Competent CT Test (cCTt)	The competent Computational Thinking Test: Development and Validation of an Unplugged Computational Thinking Test for Upper Primary	1519	X	X			X
2022	Computational Thinking Competency Assessment (CTCA)	The Use of Cognitive Diagnostic Modeling in the Assessment of Computational Thinking	564	X				

2022	Computational Thinking Concepts Test for Primary Education Adopting an ECD Approach	Validating a computational thinking concepts test for primary education using item response theory: An analysis of students' responses	13670	X			X

Source: Own work.

4. Discussion

This systematic literature review aimed to identify studies that have introduced instruments for measuring CT, as well as bibliometric variables and other variables of interest to delve deeper into this object of study.

A total of 52 research papers and 15 meta-analyses, mapping reviews, and systematic literature reviews were selected. Four studies are noteworthy ([6], [18], [19], [25]) because they make evident what supports, adds value, and justifies this literature review: the need for an analysis of the psychometric properties of those instruments.

For this purpose, it was necessary to establish what methods have been used to determine their validity and reliability. This process can be addressed from the perspective of the Classical Test Theory (CTT) and the Item Response Theory (IRT). CTT is based on methods that evaluate the quality of tests by measuring the internal consistency and validity of the content, criterion, and construct. On the other hand, the IRT offers a more advanced approach as it considers the individual characteristics of the items and participants, enabling a more accurate estimation of the skills under evaluation and a more sensitive assessment of performance. Integrating both theories allows for a more comprehensive and reliable evaluation of tests, facilitating decision-making in various educational and professional contexts. This review includes several articles that refer to adaptations of two instruments: The Computational Thinking Test (CTt) and the self-report Computational Thinking Scale (CTS). All the adaptations of the CTS [46] and CTt [37] , [38] have shown evidence of psychometric properties.

Considering the authors of the 52 articles and those most cited within them, Marcos Román-Gonzaléz and Özgen Korkmaz were found to be at the top of both lists, demonstrating their extensive research experience in CT assessment.

The protocol for this review included six questions that can be used to delve deeper into this discussion:

What is the target population of the instrument?

The 52 papers analyzed in this study address populations at various educational levels, ranging from early childhood education to college, in addition to teachers in training. Only one instrument, the Algorithmic Thinking Test for Adults (ATTA), was exclusively designed for adults. In general, the target populations were high school (35 %) and college (25 %) students. Two articles [60], [61] focused on teachers in training; three [62]–[64] , on teachers; and one [49] , on adults. Given this distribution, there is an open space for research on the design and validation of instruments that assess CT in adults, teachers, or early childhood.

What instruments have been proposed to measure CT?

The selected studies propose 40 tools to measure CT in different formats (exam, instrument, questionnaire, test, assessment, scale, and evaluation)—with scales being the most common. Three articles included three ways to assess CT using complements to the instrument: (1) a web interactive application [48] , (2) tasks from Bebras cards combined with KIBO kits [65] , and (3) a game-based strategy [16] . The latter two were used to conduct research in early childhood education. This inventory of types, formats, constructs evaluated, number of items, and other information about the instruments can aid in making decisions for future research on CT measurement. It is worth noting that 2017 marked a milestone with the first publication of an instrument designed to assess CT. This differs from [66] perspective on the matter. It should be clarified that this information is based on a systematic literature review that did not identify any other instruments prior to that year.

What constructs or skills do the instruments assess?

The instruments assess various skills associated with CT, including concepts, attitudes, and procedures. Some authors have also included feelings. Several skills were assessed in the articles reviewed here, which

means that the construct can be evaluated in different contexts. This also indicates that future research should include skills, concepts, attitudes, procedures, and feelings for a comprehensive CT assessment [67] , [68] . The results of this review are in line with previous reviews [18], [19], [24], [29], [32] , which established that algorithmic thinking, problem-solving, and abstraction are the CT skills most commonly assessed. Likewise, it was found that computational concepts (sequences, conditionals, and loops) are widely assessed in various educational environments, which is consistent with [20], [23], [27] .

What are the psychometric properties of the instruments?

The psychometric properties of the instruments were studied from the perspectives of validity and reliability. Of the instruments assessed, 86 % demonstrated validity and 69 %, reliability. Content and construct validity were predominant. Regarding reliability, internal consistency was the most commonly used criterion in the selected studies. All articles that adapted CTS presented evidence of its psychometric properties. Among those in which the CTt was adapted, only one did not provide evidence of validity. These results differ from those reported in [6] , where the authors noted that an important number of CT assessments lacked evidence of reliability and validity.

What statistical methods were used to analyze the psychometric properties?

To determine the validity of the instruments, the most common statistical methods were correlation, Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA). In most cases, reliability was tested using Cronbach's alpha. In the adaptations of the CTS [47], [50], [53]–[56] , the most popular validation method was CFA. CFA also appeared in [25] as the most common method to validate scales.

What elements are considered in CT assessment according to the literature reviews?

CT assessment focuses on the basic concepts of educational technology, highlighting its foundations and contributions to the field of teaching [69]. Table 7 below outlines the elements of CT assessment that were considered in the 15 systematic literature reviews, mapping reviews, and meta-analyses.

Table 7.
Elements of CT assessment considered in the review articles.

#	Title	Analysis
1	[19]	Empirical studies on CT assessment in college students are summarized. Elements of CT assessment reviewed in this article: block-based assessments, knowledge/skill tests, self-report Likert scales, text-based programming projects, academic achievements of CS courses, as well as interviews and observations.
2	[18]	Key characteristics of CT assessments for K-12 students are identified and classified. Elements of CT assessment reviewed in this article: tangible tasks, programming projects, self-report Likert scales, and single- and multiple-choice questions.
3	[21]	Approaches for assessing block-based programming activities for K-12 students are analyzed. Elements of CT assessment reviewed in this article: block-based assessments and programming projects that emphasize computational concepts.
4	[6]	CT implementation contexts and CT assessment tools across all educational levels are reviewed. Elements of CT assessment reviewed in this article: portfolio, interviews, knowledge tests, and a combination of tools.
5	[32]	This article analyzes the relationship between CT and academic performance in primary school students. Elements of CT assessment reviewed in this article: academic performance and knowledge tests.
6	[20]	Existing CT studies with pre-school age participants are examined. Elements of CT assessment reviewed in this article: block-based assessments and computational concepts and perspectives.
7	[27]	This paper describes the different ways in which CT has been operationalized and implemented in practice. Elements of CT assessment reviewed in this article: computational concepts, practices, and perspectives.
8	[26]	Educational contexts where CT has been implemented are presented, highlighting the ways in which CT can be assessed/measured. Elements of CT assessment reviewed in this article: computational concepts, practices, and perspectives.
9	[28]	This study investigates the relationship between CT skills development in learning settings, conceptual understanding, and CT-related dimensions. Elements of CT assessment reviewed in this article: programming-related and non-programming activities.
10	[29]	A conceptual model is designed for six CT areas: knowledge, learning strategies, assessment, tools, factors, and capacity building. Elements of CT assessment reviewed in this article: self-report scales, tests, artifact analysis, and observations.
11	[23]	Robots and processes used in CT assessment are reviewed. Elements of CT assessment reviewed in this article: portfolio, tests, and surveys.
12	[22]	Research trends in the field of CT are analyzed. Elements of CT assessment reviewed in this article: computational concepts, practices, and perspectives.
13	[24]	This review proposes the operationalization of abstraction in the context of CT. Elements of CT assessment reviewed in this article: abstraction and generalization.
14	[31]	This study establishes the impact of programming teaching on K-12 students' CT skills. Elements of CT assessment reviewed in this article: programming tools.
15	[25]	This study determines the properties of the instruments developed to measure CT. Elements of CT assessment reviewed in this article: psychometric properties and thinking skills.

Source: Own work.

Skills, attitudes, and perceptions are dimensions widely used to measure CT. Some studies [18]–[22], [27] focus on measuring CT through computational practices and concepts. All these reviews cite Brennan and Resnick's [7] curriculum guide as a foundational resource for CT. Other studies [6], [18], [19], [23], [29] use self-report scales to analyze students' perceptions and preferences. Only one systematic literature review [24] defines abstraction from a multi-dimensional perspective.

5. CONCLUSIONS

This study reviewed 52 research articles about CT assessment and measurement published between 2012 and 2022 in academic journals. Additionally, it analyzed scoping reviews, systematic mapping reviews, and literature reviews, which revealed (a) the interest in consolidating the evidence on CT assessment and (b) research gaps for this review. Consequently, this literature review was conducted to learn about the psychometric properties of CT assessment and measurement instruments, as well as CT-related variables.

This systematic review implemented a process that ensures the repeatability of the review protocol. The research questions helped to define the limits of the bibliometric variables and variables of interest. The bibliometric variables indicate that the number of articles on CT measurement instruments has increased since 2019. Most documents on CT measurement have been published in Türkiye, the US, and China. Based on the JCR index, 27.5 % of the articles were published in Q1 journals; and based on the SJR index, 47.5 % were featured in Q1 outlets. There is no evidence of CT measurement instruments in Colombia. Regarding key authors, Brennan and Resnick, as well as Selby and Woollard, are commonly cited due to their widely recognized CT curriculum designs. Marcos Román González and Özgen Korkmaz stand out for their numerous publications and the international adaptation of their instruments. The Computational Thinking Scale (CTS) has been adapted and psychometrically validated in Europe and Asia, while the Computational Thinking Test (CTt) has been adapted and its validity has been established in the same regions. However, no adaptations of these instruments to Latin American countries were identified. As the instruments were mostly applied to high school and college students, future research should address other populations, such as young children or adults.

The results of this review highlight the diverse range of CT skills that can be evaluated. Among these skills (that have a multidimensional origin), algorithmic thinking, cognitive skills, and problem-solving capabilities are the most common. Computational capabilities are also widely assessed, especially concepts that are directly related to computer programming, such as sequences, conditionals, loops, and events. It should be noted that abstraction has been commonly evaluated across all populations, but there is little scientific evidence of a rigorous evaluation of this construct. Only Ezeamuzie et al. have formally operationalized this skill.

This review revealed a variety of instruments to measure CT—with scales being the most frequently used format. This suggests that CT should be assessed in a comprehensive manner by addressing a wide range of associated skills, concepts, attitudes, and procedures. Most reviewed instruments demonstrated both validity and reliability, with content and construct validity, as well as internal consistency, being the predominant criteria. The statistical methods most commonly employed to analyze these properties are correlation, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Cronbach's alpha.

This literature review makes a contribution to future studies by demonstrating the progress made in CT assessment through the use of measurement instruments with strong psychometric properties. In conclusion, this review accomplished its objective, i.e., it identified the tools that have been used to measure CT, along with their psychometric properties and the skills they assess.

References

[1] H. Arranz de la Fuente, and A. Pérez García, "*Evaluación del pensamiento computacional en educación primaria,*" Rev. Interuniv. Investig. en Tecnol. Educ. Educ., no. 3, pp. 25–39, 2017. http://dx.doi.org/10.6018/riite/2017/267411

[2] C. Fadel, M. Bialik, and B. Trilling, *Educación en cuatro dimensiones: las competencias que los estudiantes necesitan para su realización*, Santiago, Chile: Grafhika Impresores, 2016. [Online]. Available: https://www.educarchile.cl/recursos-para-el-aula/educacion-en-cuatro-dimensiones

[3] C. L. Scott, "*El futuro del aprendizaje 2 ¿Qué tipo de aprendizaje se necesita para el siglo XXI?," Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura*, París, Rep. ERF No. 14, 2015. [Online]. Available: https://repositorio.minedu.gob.pe/bitstream/handle/20.500.12799/4661/El%20futuro%20del%20aprendizaje%202%20Qu%c3%a9%20tipo%20de%20aprendizaje%20se%20necesita%20para%20el%20siglo%20XXI.pdf?sequence=1&isAllowed=y

[4] J. M. Wing, "Computational thinking and thinking about computing," *Philos. Trans. R. Soc. A Math. Phys.Eng. Sci.*, vol. 366, no. 1881, pp. 3717–3725, Oct. 2008. https://doi.org/10.1098/rsta.2008.0118

[5] M. Román-González, "*Codigoalfabetización y pensamiento computacional en educación primaria y secundaria: validación de un instrumento y evaluación de programas*," Ph.D. dissertation, Escuela Internacional de Doctorado, Universidad Nacional de Educación a Distancia, España, 2016. [Online]. Available: http://e-spacio.uned.es/fez/view/tesisuned:Educacion-Mroman

[6] X. Tang, Y. Yin, Q. Lin, R. Hadad, and X. Zhai, "Assessing computational thinking: A systematic review of empirical studies," *Comput. Educ.*, vol. 148, p. 103798, Apr. 2020. https://doi.org/10.1016/j.compedu.2019.103798

[7] K. Brennan, and M. Resnick, "New frameworks for studying and assessing the development of computational thinking," in *Proceedings of the 2012 annual meeting of the American educational research association, Vancouver, Canada, Apr. 2012*, p. 25. [Online]. Available: https://scratched.gse.harvard.edu/ct/files/AERA2012.pdf

[8] D. Weintrop et al., "*Defining Computational Thinking for Mathematics and Science Classrooms*," J. Sci. Educ. Technol., vol. 25, pp. 127–147, Feb. 2016. https://doi.org/10.1007/s10956-015-9581-5

9] L. Werner, J. Denner, and S. Campe, "*Children programming games: A strategy for measuring computational learning*," ACM Trans. Comput. Educ., vol. 14, no. 4, pp. 1-22, Dec. 2014. https://doi.org/10.1145/2677091

[10] ISTE, and CSTA, "*Operational definition of computational thinking for K-12 education.*," National Science Foundation, 2011. [Online]. Available: https://cdn.iste.org/www-root/Computational_Thinking_Operational_Definition_ISTE.pdf

[11] C. Selby, and J. Woollard, "*Computational thinking: the developing definition*," University of Southampton, Inglaterra, Rep. Soton356481, Oct. 2013. [Online]. Available: http://eprints.soton.ac.uk/id/eprint/356481

[12] A. Yadav, C. Mayfield, N. Zhou, S. Hambrusch, and J. T. Korb, "*Computational thinking in elementary and secondary teacher education*," ACM Trans. Comput. Educ., vol. 14, no. 1, pp. 1–16, Mar. 2014. https://doi.org/10.1145/2576872

[13] E. Relkin, L. de Ruiter, and M. U. Bers, "TechCheck: Development and Validation of an Unplugged Assessment of Computational Thinking in Early Childhood Education," *J. Sci. Educ. Technol.*, vol. 29, no. 4, pp. 482–498, Aug. 2020. https://doi.org/10.1007/s10956-020-09831-x

[14] E. Relkin, and M. Bers, "TechCheck-K: A measure of computational thinking for kindergarten children," in *2021 IEEE Global Engineering Education Conference, Vienna, Austria, 2021*, pp. 1696-1702. https://doi.org/10.1109/EDUCON46332.2021.9453926

[15] L. El-Hamamsy et al., "Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt)," *Front. Psychol.*, vol. 13, Dec. 2022. https://doi.org/10.3389/fpsyg.2022.1082659

[16] L. Wang, F. Geng, X. Hao, D. Shi, T. Wang, and Y. Li, "Measuring coding ability in young children: relations to computational thinking, creative thinking, and working memory," *Curr. Psychol.*, vol. 42, no. 10, pp. 8039-8050, Apr. 2021. https://doi.org/10.1007/s12144-021-02085-9

[17] K. Kanaki, and M. Kalogiannakis, "Assessing Algorithmic Thinking Skills in Relation to Gender in Early Childhood," *Educ. Process Int. J.*, vol. 11, no. 2, pp. 44–59, Jun. 2022. https://doi.org/10.22521/edupij.2022.112.3

[18] M. Cutumisu, C. Adams, and C. Lu, "A Scoping Review of Empirical Research on Recent Computational Thinking Assessments," *J. Sci. Educ. Technol.*, vol. 28, no. 6, pp. 651–676, Dec. 2019. https://doi.org/10.1007/s10956-019-09799-3

[19] C. Lu, R. Macdonald, B. Odell, and V. Kokhan, "A scoping review of computational thinking assessments in higher education," *J. Comput. High. Educ.*, vol. 34, no. 2, pp. 416–461, Aug. 2022. https://doi.org/10.1007/s12528-021-09305-y

[20] K. I. McCormick, and J. A. Hall, "Computational thinking learning experiences, outcomes, and research in preschool settings: a scoping review of literature," *Educ. Inf. Technol.*, vol. 27, no. 3, pp. 3777–3812, Apr. 2022. https://doi.org/10.1007/s10639-021-10765-z

[21] N. Da Cruz Alves, C. Gresse Von Wangenheim, and J. C. R. Hauck, "Approaches to assess computational thinking competences based on code analysis in K-12 education: A systematic mapping study," *Informatics Educ.*, vol. 18, no. 1, pp. 17–39, Apr. 2019. https://doi.org/10.15388/infedu.2019.02

[22] M. Tekdal, "Trends and development in research on computational thinking," *Educ. Inf. Technol.*, vol. 26, no. 5, pp. 6499–6529, Sep. 2021. https://doi.org/10.1007/s10639-021-10617-w

[23] E. Bakala, A. Gerosa, J. P. Hourcade, and G. Tejera, "Preschool children, robots, and computational thinking: A systematic review," *Int. J. Child-Computer Interact.*, vol. 29, p. 100337, Sep. 2021. https://doi.org/10.1016/j.ijcci.2021.100337

[24] N. O. Ezeamuzie, J. S. C. Leung, and F. S. T. Ting, "Unleashing the Potential of Abstraction From Cloud of Computational Thinking: A Systematic Review of Literature," *J. Educ. Comput. Res.*, vol. 60, no. 4, pp. 877–905, Dec. 2022. https://doi.org/10.1177/07356331211055379

[25] H. İ. Haseki, and U. Ilic, "An Investigation of the Data Collection Instruments Developed to Measure Computational Thinking," *Informatics Educ.*, vol. 18, no. 2, pp. 297–319, Oct. 2019. https://doi.org/10.15388/infedu.2019.14

[26] P. Kakavas, and F. C. Ugolini, "Computational thinking in primary education : a systematic literature review," vol. 11, no. 2. Dec. 2019. https://doi.org/10.2478/rem-2019-0023

[27] N. O. Ezeamuzie, and J. S. C. Leung, "Computational Thinking Through an Empirical Lens: A Systematic Review of Literature," *J. Educ. Comput. Res.*, vol. 60, no. 2, pp. 481–511, Apr. 2022. https://doi.org/10.1177/07356331211033158

[28] E. Taslibeyaz, E. Kursun, and S. Karaman, "How to Develop Computational Thinking: A Systematic Review of Empirical Studies," *Informatics Educ.*, vol. 19, no. 4, pp. 701–719, Dec. 2020. https://doi.org/10.15388/infedu.2020.30

[29] C. Tikva, and E. Tambouris, "Mapping Computational Thinking through Programming in K-12 Education: A Conceptual Model based on a Systematic Literature Review," *Comput. Educ.*, vol. 162, p. 104083, Mar. 2020. https://doi.org/10.1016/j.compedu.2020.104083

[30] H. Lei, M. M. Chiu, F. Li, X. Wang, and G. Ya-Jing, "Computational thinking and academic achievement: A meta-analysis among students," *Child. Youth Serv. Rev.*, vol. 118, p. 105439, Nov. 2020. https://doi.org/10.1016/j.childyouth.2020.105439

[31] L. Sun, L. Hu, and D. Zhou, "Which way of design programming activities is more effective to promote K-12 students' computational thinking skills? A meta-analysis," *J. Comput. Assist. Learn.*, vol. 37, no. 4, pp. 1048–1062, Aug. 2021. https://doi.org/10.1111/jcal.12545

[32] L. Sun, L. Hu, and D. Zhou, "The bidirectional predictions between primary school students' STEM and language academic achievements and computational thinking: The moderating role of gender," *Think. Ski. Creat.*, vol. 44, p. 101043, Jun. 2022. https://doi.org/10.1016/j.tsc.2022.101043

[33] A. C. Tricco et al., "PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation," *Ann. Intern. Med.*, vol. 169, no. 7, pp. 467–473, Sep. 2018. https://doi.org/10.7326/M18-0850

[34] J. M. Wing, "Computational thinking," Commun. *ACM*, vol. 49, no. 3, pp. 33–35, Mar. 2006. https://dl.acm.org/doi/pdf/10.1145/1118178.1118215

[35] G. Urrútia, and X. Bonfill, "PRISMA declaration: A proposal to improve the publication of systematic reviews and meta-analyses," *Medicina Clínica*, vol. 135, no. 11, pp. 507–511, Oct. 2010. https://doi.org/10.1016/j.medcli.2010.01.015

[36] Y. Li, S. Xu, and J. Liu, "Development and Validation of Computational Thinking Assessment of Chinese Elementary School Students," *J. Pacific Rim Psychol.*, vol. 15, Aug. 2021. https://doi.org/10.1177/18344909211010240

[37] M. Román-González, J. C. Pérez-González, J. Moreno-León, and G. Robles, "Can computational talent be detected? Predictive validity of the Computational Thinking Test," Int. J. Child-Computer Interact., vol. 18, pp. 47–58, Nov. 2018. https://doi.org/10.1016/j.ijcci.2018.06.004

[38] M. Román-González, J. C. Pérez-González, and C. Jiménez-Fernández, "Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test," *Comput. Human Behav.*, vol. 72, pp. 678–691, Jul. 2017. https://doi.org/10.1016/j.chb.2016.08.047

[39] K. Lee, and J. Cho, "Computational Thinking Evaluation Tool Development for Early Childhood Software Education," *JOIV Int. J. Informatics Vis.,* vol. 5, no. 3, p. 313, Sep. 2021. https://doi.org/10.30630/joiv.5.3.672

[40] Y. Lee, and J. Cho, "Toward Developing a Real-World Computational Thinking Test Tool from Existing Models," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2S6, pp. 389–393, Jul. 2019. https://doi.org/10.35940/ijrte.B1073.0782S619

[41] S. Sovey, K. Osman, and M. E. E. M. Matore, "Gender differential item functioning analysis in measuring computational thinking disposition among secondary school students," Front. Psychiatry, vol. 13, Nov. 2022. https://doi.org/10.3389/fpsyt.2022.1022304

[42] S. Sovey, K. Osman, and M. E. E. Mohd-Matore, "Exploratory and Confirmatory Factor Analysis for Disposition Levels of Computational Thinking Instrument Among Secondary School Students," *Eur. J. Educ. Res.*, vol. 11, no. 2, pp. 639–652, Apr. 2022. https://doi.org/10.12973/eu-jer.11.2.639

[43] S. C. Kong, and Y. Q. Wang, "Item response analysis of computational thinking practices: Test characteristics and students' learning abilities in visual programming contexts," *Comput. Human Behav.*, vol. 122, p. 106836, Sep. 2021. https://doi.org/10.1016/j.chb.2021.106836

[44] K. Siu-Cheung, and M. Lai, "Effects of a teacher development program on teachers' knowledge and collaborative engagement, and students' achievement in computational thinking concepts," *Br. J. Educ. Technol.*, vol. 54, no. 2, pp. 489–512, Mar. 2023. https://doi.org/10.1111/bjet.13256

[45] L. El-Hamamsy, M. Zapata-Cáceres, E. M. Barroso, F. Mondada, J. D. Zufferey, and B. Bruno, "The Competent Computational Thinking Test: Development and Validation of an Unplugged Computational Thinking Test for Upper Primary School," *J. Educ. Comput. Res.*, vol. 60, no. 7, pp. 1818–1866, May. 2022. https://doi.org/10.1177/07356331221081753

[46] Ö. Korkmaz, R. Çakir, and M. Y. Özden, "A validity and reliability study of the computational thinking scales (CTS)," *Comput. Human Behav.*, vol. 72, pp. 558–569, Jul. 2017. https://doi.org/10.1016/j.chb.2017.01.005

[47] Ö. Korkmaz, and X. Bai, "Adapting Computational Thinking Scale (CTS) for Chinese High School Students and Their Thinking Scale Skills Level," *Particip. Educ. Res.*, vol. 6, no. 1, pp. 10–26, Jun. 2019. https://doi.org/10.17275/per.19.2.6.1

[48] E. Çoban, and Ö. Korkmaz, "An alternative approach for measuring computational thinking: Performance-based platform," *Think. Ski. Creat.*, vol. 42, p. 100929, Dec. 2021. https://doi.org/10.1016/j.tsc.2021.100929

[49] M. Lafuente Martínez, O. Lévêque, I. Benítez, C. Hardebolle, and J. D. Zufferey, "Assessing Computational Thinking: Development and Validation of the Algorithmic Thinking Test for Adults," *J. Educ. Comput. Res.*, vol. 60, no. 6, pp. 1436–1463, Feb. 2022. https://doi.org/10.1177/07356331211057819

[50] J. Guggemos, S. Seufert, and M. Román-González, "Computational Thinking Assessment – Towards More Vivid Interpretations," *Technol. Knowl. Learn.*, vol. 28, no. 2, pp. 539–568, Jun. 2023. https://doi.org/10.1007/s10758-021-09587-2

[51] D. E. Sondakh, K. Osman, and S. Zainudin, "A Proposal for Holistic Assessment of Computational Thinking for Undergraduate: Content Validity," *Eur. J. Educ. Res.*, vol. 9, no. 1, pp. 33–50, Jan. 2020. https://doi.org/10.12973/eu-jer.9.1.33

[52] D. E. Sondakh, K. Osman, and S. Zainudin, "A Pilot Study of an Instrument to Assess Undergraduates' Computational thinking Proficiency," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, 2020. https://doi.org/10.14569/IJACSA.2020.0111134

[53] E. Polat, S. Hopcan, S. Kucuk, and B. Sisman, "A comprehensive assessment of secondary school students' computational thinking skills," *Br. J. Educ. Technol.*, vol. 52, no. 5, pp. 1965–1980, Sep. 2021. https://doi.org/10.1111/bjet.13092

[54] L. Sun, L. Hu, D. Zhou, and W. Yang, "Evaluation and developmental suggestions on undergraduates' computational thinking: a theoretical framework guided by Marzano's new taxonomy," *Interact. Learn. Environ.*, vol. 31, no. 10, pp. 6588–6610, Mar. 2023. https://doi.org/10.1080/10494820.2022.2042311

[55] M. Junpho, A. Songsriwittaya, and P. Tep, "Reliability and Construct Validity of Computational Thinking Scale for Junior High School Students: Thai Adaptation," *Int. J. Learn. Teach. Educ. Res.*, vol. 21, no. 9, pp. 154–173, Sep. 2022. https://doi.org/10.26803/ijlter.21.9.9

[56] H. Nakahara et al., "Development of the Japanese Version of the Computational Thinking Scales for First-Year University Students in Humanities," *Japan J. Educ. Technol.*, vol. 46, no. 1, 2022. https://doi.org/10.15077/jjet.45089

[57] L. Mujica De Statzewitch, "Evaluación del Desarrollo del Pensamiento Computacional en Estudiantes de Educación Primaria y Media General.," *Rev. Digit. del Dr. en Educ. la Univ. Cent. Venez.*, vol. 7, no. 13,

pp. 35–56, Jan-Jun. 2021. http://saber.ucv.ve/ojs/index.php/rev_arete/article/view/21327/144814487622

[58] İ. Çetin, T. Otu, and A. Oktaç, "Adaption of the Computational Thinking Test into Turkish," *Turkish J. Comput. Math. Educ.*, vol. 11, no. 2, May. 2020. https://doi.org/10.16949/turkbilmat.643709

[59] C. Shiau-Wei, L. Chee-Kit, and B. Sumintono, "Assessing computational thinking abilities among Singapore secondary students: a Rasch model measurement analysis," *J. Comput. Educ.*, vol. 8, no. 2, pp. 213–236, Jun. 2021. https://doi.org/10.1007/s40692-020-00177-2

[60] A. Dolmaci, and N. E. Akhan, "The Development of Computational Thinking Skills Scale: Validity and Reliability Study," *İnsan ve Toplum Bilim. Araştırmaları Derg.*, vol. 9, no. 3, pp. 1970–1991, Sep. 2020. https://doi.org/10.15869/itobiad.698736

[61] B. Ertugrul-Akyol, "Development of Computational Thinking Scale: Validity and Reliability Study," *Int. J. Educ. Methodol.*, vol. 5, no. 3, pp. 421–432, Aug. 2019. https://doi.org/10.12973/ijem.5.3.421

[62] K. Bati, and M. İkbal Yetişir, "Examination of Turkish Middle School STEM Teachers' Knowledge about Computational Thinking and Views Regarding Information and Communications Technology," *Comput. Sch.*, vol. 38, no. 1, pp. 57–73, Mar. 2021. https://doi.org/10.1080/07380569.2021.1882206

[63] D. C. Boulden, A. Rachmatullah, K. M. Oliver, and E. Wiebe, "Measuring in-service teacher self-efficacy for teaching computational thinking: development and validation of the T-STEM CT," *Educ. Inf. Technol.*, vol. 26, no. 4, pp. 4663–4689, Jul. 2021. https://doi.org/10.1007/s10639-021-10487-2

[64] P. J. Rich, R. A. Larsen, and S. L. Mason, "Measuring teacher beliefs about coding and computational thinking," *J. Res. Technol. Educ.*, vol. 53, no. 3, pp. 296–316, Jul. 2021. https://doi.org/10.1080/15391523.2020.1771232

[65] J. Sung, "Assessing young Korean children's computational thinking: A validation study of two measurements," *Educ. Inf. Technol.*, vol. 27, no. 9, pp. 12969–12997, Nov. 2022. https://doi.org/10.1007/s10639-022-11137-x

[66] M. Zapata-Cáceres, "Enseñanza, evaluación y análisis de habilidades de pensamiento computacional en etapas tempranas," Ph.D. dissertation, Universidad Rey Juan Carlos, Madrid, España, 2022. [Online]. Available: https://burjcdigital.urjc.es/handle/10115/19965?show=full

[67] S. Grover, and R. D. Pea, "'"*Systems of Assessments" for Deeper Learning of Computational Thinking in K-12*," in Annual Meeting of the American Educational Research Association, Chicago, IL, USA, Apr. 15-20, 2015. [Online]. Available: https://www.researchgate.net/publication/275771253

[68] K. Siu-Cheung, and H. Abelson, *Eds., Computational Thinking Education*, 1st. ed. Gateway East, Singapore: Springer Singapore, 2019. https://doi.org/10.1007/978-981-13-6528-7

[69] M. del M. Sánchez Vera, "Computational Thinking in Educational Environments: An Approach from Educational Technology," *Research in Education and Learning Innovation Archives*, no. 23, p. 24-39, Jul-Dec. 2019. https://doi.org/10.7203/realia.23.15635

Notes

5. ACKNOWLEDGEMENT AND FUNDING :

The authors thank the Ph.D. program in Education Sciences at Universidad del Quindío, which provided funding for Milena Corrales Álvarez and Lina Marcela Ocampo to complete their doctoral studies.

They would also like to thank Universidad del Quindío for funding the research project (Code 1187 of the Call for Proposals No. 14 of 2022).

CONFLICTS OF INTEREST :

The authors declare that there is no conflict of interest.

AUTHORS CONTRIBUTIONS :
- Milena Corrales-Álvarez: Methodology, Conceptualization, Investigation, and Writing - Review & Editing.
- Lina Marcela Ocampo: Methodology, Conceptualization, Investigation, and Writing - Review & Editing.
- Sergio Augusto Cardona-Torres: Conceptualization, Investigation, and Writing - Review & Editing

Información adicional

How to cite / Cómo citar: M. Corrales-Álvarez, L. M. Ocampo, and S. A. Cardona Torres, "Instruments for Evaluating Computational Thinking: a Systematic Review," *TecnoLógicas*, vol. 27, no. 59, e2950, Apr. 2024. https://doi.org/10.22430/22565337.2950

Enlace alternativo

https://revistas.itm.edu.co/index.php/tecnologicas/article/view/2950 (html)

Milena Corrales-Álvarez, Lina Marcela Ocampo,
Sergio Augusto Cardona Torres

Instruments for Evaluating Computational Thinking: A Systematic Review

Instrumentos de evaluación del pensamiento computacional: una revisión sistemática

TecnoLógicas
vol. 27, núm. 59, p. 1 - 26, 2024
Instituto Tecnológico Metropolitano, Colombia
tecnologicas@itm.edu.co

ISSN: 0123-7799 **/ ISSN-E:** 2256-5337

DOI: https://doi.org/10.22430/22565337.2950