Jupiter: Jurnal Pendidikan Terapan
Homepage : https://journal.diginus.id/jupiter/index
Email: jurnalpendidikanterapan@gmail.com

p-ISSN : 2985-3214; e-ISSN : 2964-3171
Jupiter, Vol. 4, No. 5, November 2026
Page 322-341

Computational Thinking Across Educational Levels in Mathematics Education: A Systematic Literature Review of the Evidence on Cross-Level Development

Sarwan[1], Kartika Yulianti[2*], Jarnawi Afghani Dahlan[3], Tia Purniati[4]
[1,2,3,4]Universitas Pendidikan Indonesia, Bandung, Indonesia

ARTICLE INFO

Article history:
Submitted: May 19, 2026
Final Revised: November 10, 2026
Accepted: November 20, 2026
Published: November 20, 2026

Keywords:
Computational Thinking;
Mathematics Education;
Systematic Literature Review;
Cross-Level Development;
Pedagogical Factors.

ABSTRACT

Purpose: This study synthesises how computational thinking (CT) is evidenced, emphasised, and shaped by pedagogical factors in mathematics education across primary, secondary, and higher education, and examines how far the available literature supports a cross-level developmental account of CT. **Methods:** A systematic literature review (SLR) was conducted following PRISMA 2020. Scopus was searched on 19 May 2026, and 1,155 records were screened against explicit eligibility criteria (open-access, English-language journal articles published in 2022–2025), leaving 20 included articles. Data were coded through directed content analysis and synthesised narratively in an evidence extraction matrix. Screening and coding were carried out mainly by one reviewer without an inter-rater statistic, and empirical studies, review-type articles, student-based evidence, and teacher-based evidence were kept separate in the synthesis. No formal risk-of-bias tool was applied; instead, each synthesised claim was graded as strong, moderate, or tentative according to the number, design, and participants of the supporting studies. **Findings:** Primary-level studies most often reported debugging, decomposition, abstraction, and algorithm design in concrete, hands-on tasks. Secondary-level studies, which were few, reported algorithm design, pattern recognition, and debugging. Higher-education studies, drawn largely from prospective-teacher samples, linked CT to creativity, critical thinking, and integrated knowledge. These patterns come from different samples and instruments rather than from longitudinal data, so they indicate a plausible concrete-to-abstract progression, not a demonstrated developmental trajectory. A possible discontinuity at the primary-to-secondary transition is a tentative interpretation that no included study tested directly. **Research Implications:** Technology type, instructional design, teacher preparedness, and affective factors were the factors most often associated with CT development, although much of this evidence is small-scale or correlational. The findings point to teacher professional learning and cross-level curriculum coherence as priorities, and to longitudinal research and standardised, level-sensitive CT assessment as research needs. **Originality:** Earlier reviews addressed K-12 instructional effects, effect sizes, bibliometric mapping, or national research trends. This review compares the evidence on CT indicators across primary, secondary, and higher education within one corpus, and offers a conceptual, hypothesis-generating spiral-expansive model that still requires empirical validation.

Doi: https://doi.org/10.61255/jupiter.v4i5.1160

INTRODUCTION

The pace of digital transition is unlike anything we have seen before, requiring individuals to exhibit cognitive abilities that go far beyond linear or procedural thinking. In this context, computational thinking (CT) has emerged as one of the most talked-about cognitive conceptions in global educational discourse in the last two decades. Wing (2006) introduced CT as a fundamental analytical skill that draws on concepts from computer science to solve problems, design systems, and understand human behaviour. CT is considered to be more than just a technical computing talent but a basic way of thinking that can be useful in solving real world problems including those in the domain of

mathematics (Shute et al., 2017). Mathematics and CT have profound historical, conceptual links. Both the sciences rely on abstract reasoning, pattern recognition, logical deduction and generalisation (Ye et al., 2023). CT principles were first popularised by Papert (1980) through the introduction of the Logo programming language and he believed that computational experience may provide pupils with powerful new ways of comprehending and expressing mathematical ideas. This conviction has since been supported by increasing empirical evidence, with several studies reporting gains in students' mathematical understanding when CT is integrated into mathematics instruction (Kaup et al., 2023; Suparman et al., 2025).

Several reviews have already synthesised CT in mathematics education, and they define what the present review can add. Ye et al. (2023) reviewed CT-based mathematics instruction and student learning in K-12 settings and found that cyclical, interactive reasoning processes can co-develop CT and mathematics across several mathematical domains; their scope, however, stopped at K-12, and it centred on instructional approaches and learning outcomes rather than on how the emphasis of individual CT indicators changes across levels. Suparman et al. (2025) meta-analysed CT in mathematics instruction integrated with STEAM education and reported moderate to large effects on mathematical understanding; an effect-size synthesis, however, says little about which CT indicators are enacted at which level. Ersozlu et al. (2023) produced a scientific mapping of research on developing CT through mathematics, describing publication patterns and thematic foci such as the growing link between CT and higher-order thinking, but not the profile of CT within individual studies. Park and Kwon (2022) analysed research trends on CT in science and mathematics education in a single national context, the Republic of Korea. Taken together, existing reviews describe whether CT-integrated instruction works, where research is concentrated, and which themes dominate. To the authors' knowledge, none has compared the evidence on CT indicators across primary, secondary, and higher education within one corpus, or stated how strongly that evidence supports cross-level developmental claims. Because Suparman et al. (2025), Ersozlu et al. (2023), and Park and Kwon (2022) also meet the eligibility criteria of the present review, they are treated here as review-type studies and distinguished from primary empirical studies in the synthesis.

Against this background, three gaps remain. First, most primary studies focus on a single educational level, and the literature offers little cross-level comparison of how CT is evidenced as learners move from primary school to higher education. Second, it remains unclear which CT indicators are emphasised at which level, which makes it difficult for instructors and curriculum designers to set developmentally appropriate learning goals. Third, the pedagogical and systemic factors that shape CT development in different educational settings have rarely been documented together (Park & Kwon, 2022), and the strength of the supporting evidence is seldom stated. These gaps can lead to undifferentiated, one-size-fits-all teaching strategies, regardless of where pupils are in their development. Addressing them also calls for caution: cross-level comparisons built from different studies are not longitudinal evidence, so developmental claims should be graded by the strength of what supports them. This study therefore employs a Systematic Literature Review (SLR) to synthesise recent evidence on CT in mathematics education and to state explicitly how strongly that evidence supports each cross-level claim. It is guided by four research questions: (RQ1) How is CT evidenced by primary, secondary, and higher education students in the context of mathematics? (RQ2) What are the differences in emphasis of CT indicators by educational level? (RQ3) What does the evidence suggest about the developmental path of CT from primary school to higher education? (RQ4) What systemic and pedagogical factors are consistently associated with CT development across these levels? The review analyses 20 articles indexed by Scopus from 2022 to 2025.

Discussions of CT in education can be found in the scholarly literature prior to Wing's (2006) seminal article in Communications of the ACM. In his influential book Mindstorms: Children, Computers, and Powerful Ideas, Papert (1980) laid the philosophical foundation for viewing computers not as tools, but as media for thinking, allowing children to construct mathematical knowledge through active, playful interaction with machines. His constructionist learning theory that the most effective learning happens when learners produce shareable artefacts remains a key theoretical anchor for CT research today (Yunianto, Jarvis, et al., 2025).

Wing (2006) operationalised CT as problem solving, system design and understanding human behaviour through concepts fundamental to computer science as a universal cognitive skill rather than the exclusive province of computer scientists. Later, International Society for Technology in Education and Computer Science Teachers Association (ISTE & CSTA, 2011) transformed CT into operational skills for K-12 education, which included formulating the problem, logically organising data, representing data based on abstraction, automating the solution of the problem, and evaluating the solution. One of the frameworks that has achieved the broadest currency in mathematics and science education research is the six-facet model of (Shute et al., 2017). The facets are: (1) decomposition breaking complex problems into manageable sub-problems; (2) abstraction identifying essential information while discarding irrelevant details; (3) algorithm design creating systematic, ordered instructions for solving problems; (4) debugging identifying and correcting errors in processes or programs; (5) iteration repeating processes to converge on optimal solutions; and

(6) generalisation identifying patterns and transferring solutions across contexts. This framework is consistently used in several studies analysed in this review (Yunianto, Jarvis, et al., 2025; Fang et al., 2023; Mumcu et al., 2023).

The connection between CT and mathematics is not only conceptual but is supported by a growing body of empirical evidence. Ye et al. (2023) performed a systematic literature review and found that cyclical, interactive reasoning processes can co-develop CT and mathematics, with various mathematical domains (arithmetic, algebra, geometry, calculus) acting as productive contexts for CT learning. In practice, this integration can be of two types. Plugged approaches use digital software such as Scratch, GeoGebra or App Inventor to enable students to code to investigate and represent mathematical ideas (Fang et al., 2025; Yunianto, Jarvis, et al., 2025). Unplugged approaches use card games, physical activities, and logic puzzles that represent CT principles such as decomposition and pattern recognition (Mumcu et al., 2023) to achieve the same goals without digital technology. Existing studies suggest that these approaches may be complementary and can be combined. The infusion of CT into mathematics is also increasingly framed within STEAM (Science, Technology, Engineering, Arts, Mathematics) perspectives, where CT is not an add-on but rather a core part of scientific and mathematical thinking. In a meta-analysis, Suparman et al. (2025) reported that the integration of CT in mathematics instruction in STEAM contexts significantly improves students' mathematical understanding across levels with moderate to large effect sizes.

CT development is not the result of a single factor but rather of a number of interacting factors. Technologically, the type and mode of use of digital platforms affect which dimensions of CT are developed. Block-based platforms such as Scratch have been used to support mathematics learning and CT at primary level and in teacher education (Fang et al., 2023; Molina-Ayuso et al., 2022, 2024), and text-based GeoGebra commands have been used to engage primary students in algorithm design and debugging within mathematical construction tasks (Yunianto, Jarvis, et al., 2025). Design-based and collaborative approaches, including the 5E instructional model, have been used to integrate CT into mathematics lessons (Fang et al., 2023, 2025). Teacher readiness, in terms of both CT content knowledge and pedagogical capacity to embed CT into mathematics curricula, is often identified as the most critical barrier to effective classroom implementation (Djidu et al., 2025; Moon et al., 2023). Affectively, Kaya et al. (2025) found that 21st-century skills of entrepreneurship, innovation, leadership, and information literacy accounted for 57% of the variance in prospective teachers' CT levels, suggesting that CT development may be closely tied to dispositional development more broadly.

METHOD

Research Design

This study employs a Systematic Literature Review (SLR) guided by the PRISMA 2020 statement (Page et al., 2021), which supports a transparent and reproducible account of how studies were identified, screened, and included, and which gives stronger grounds for conclusions than traditional narrative reviews. No review protocol was prepared or registered before the search, which is acknowledged as a limitation. The review was designed to answer the four research questions stated in the Introduction. Because the search was confined to one database and to open-access articles (see below), the review is best read as a synthesis of a bounded corpus, not of the whole literature on CT in mathematics education.

Search Strategy and Selection of Studies

The search was conducted in Scopus, selected for its broad coverage of peer-reviewed education research and its structured subject classification. Scopus was the only database searched; the consequences of this decision are discussed under Limitations. The search was run on 19 May 2026 with the following string: TITLE-ABS-KEY ("computational thinking") AND TITLE-ABS-KEY ("mathematics education" OR "mathematics learning" OR "mathematics teaching" OR "mathematical pedagogy" OR "mathematics instruction" OR "mathematics"), which yielded 1,155 records. These were filtered by document type (article only), language (English), subject area (Social Sciences and Mathematics), and access type (open access), and publication year (2022–2025), reducing the pool to 213 records. Open access was used as a pragmatic criterion so that every full text could be assessed without access barriers; it may exclude relevant paywalled studies. Titles and abstracts were screened against the eligibility criteria in Table 1, which excluded a further 130 records and left 83 articles for full-text assessment. After full-text review, 63 articles were excluded because they did not address at least one of the four research questions or lacked sufficient information about the educational level of participants, resulting in a final corpus of 20 articles. An article was therefore eligible if it addressed at least one research question, not all four. The selection process is illustrated in the PRISMA 2020 flow diagram (Figure 1), and the inclusion and exclusion criteria are presented in Table 1. Screening of records and coding were carried out by the first author. Two co-authors validated the decisions on the included articles; only a sample of the coded data was checked, the proportion checked was not recorded, and no inter-rater agreement statistic

was calculated. Excluded records were not independently re-screened. Disagreements were resolved through discussion until consensus. Because the screening step was not independently replicated and the access and language filters were applied at the database level, the effect of these restrictions on the corpus cannot be quantified from this review.

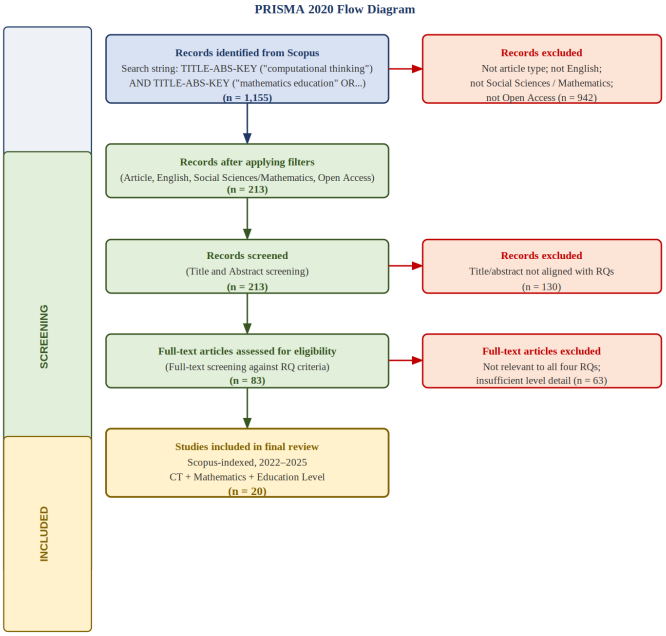

Figure 1. PRISMA 2020 Flow Diagram of the Article Selection Process

Table 1. Inclusion and Exclusion Criteria for Article Selection

Category	Inclusion Criteria	Exclusion Criteria
Document Type	• Peer-reviewed journal articles only	• Book chapters, conference papers, theses, technical reports, editorials
Language	• Published in English	• Articles in languages other than English
Subject Area	• Social Sciences or Mathematics (Scopus subject classification)	• Articles outside Social Sciences and Mathematics subject categories
Access Type	• Open Access publications	• Non-Open Access / paywalled articles without accessible full text
Database	• Indexed in Scopus	• Articles not indexed in Scopus
Topical Relevance	• Explicitly addresses computational thinking (CT) in the context of mathematics education or mathematics learning	• Discusses CT without connection to mathematics; mathematics studies without mention of CT
Educational Level	• Identifiable educational level: primary school, secondary school, or higher education	• Studies without clear information about participants educational level
Research Questions Alignment	• Addresses at least one of the four RQs: CT characteristics, CT indicator emphasis, CT developmental trajectory, or factors influencing CT. Empirical studies and review-type studies (systematic reviews, meta-analyses, mappings) were both eligible and are labelled separately in the synthesis	• Studies that do not address any of the four RQs; purely theoretical or policy papers without empirical data
Publication Year	• ● Published between 2022 and 2025	• Published before 2022 or after 2025

Characteristics of the Included Articles

The 20 included articles were published between 2022 and 2025: 2022 (n = 2), 2023 (n = 7), 2024 (n = 3), and 2025 (n = 8). As coded in Table 2, six articles concern primary education only, two span primary and higher/teacher education, two concern secondary education only, two span secondary and higher education, five concern higher education only, and three are cross-level (K-12 or general) studies. Seventeen articles are empirical studies and three (S12, S14, and S17) are review-type studies: a meta-analysis, a research-trend review, and a scientific mapping. These were retained as secondary evidence. The participants also differ: eight articles centre on students (S1, S5, S7, S9, S10, S15, S18, S19), eight on teachers or prospective teachers (S2, S3, S4, S8, S11, S13, S16, S20), and one on higher-education participants as reported in the original article (S6). Cross-level comparisons therefore combine evidence about students' CT with evidence about teachers' knowledge and conceptions of CT. Where sample sizes were extracted, they ranged from two teachers (S13) to 300 prospective teachers (S20), and all articles were cross-sectional or short-term. Geographically, the articles originate from Indonesia, Turkey, Spain, Norway, the United States, Denmark, Hong Kong, Hungary, Brazil, Lithuania, and the Republic of Korea. Table 2 presents the complete list of included articles, identified as S1–S20, and Table 3 summarises the design, sample, and participant type of each article by assigned level.

Table 2. List of Included Articles (S1–S20) and Participant Characteristics

ID	Author (Year)	Article Title	Journal	Level (participants)	DOI
S1	Yunianto, Jarvis, et al. (2025)	Primary school students' problem-solving strategies in creating artworks with GeoGebra	LUMAT	Primary (students)	10.31129/LUMAT.13.2.2547
S2	Dahshan & Galanti (2024)	Teachers in the Loop: Integrating CT and Mathematics to Build Early Place Value Understanding	Education Sciences	Primary & HE (in-service early-grade teachers in a graduate course)	10.3390/educsci14020201
S3	Humble & Mozelius (2023)	Grades 7–12 teachers' perception of computational thinking for mathematics and technology	Frontiers in Education	Secondary (Grades 7–12 teachers)	10.3389/feduc.2023.956618
S4	Djidu et al. (2025)	Mathematics Teachers' Perspectives on CT: Knowledge, Strategies, and Challenges	Qualitative Report	Secondary & HE (mathematics teachers)	10.46743/2160-3715/2025.6494
S5	Kaup et al. (2023)	Integrating computational thinking to enhance students' mathematical understanding	J. of Pedagogical Research	Primary (students)	10.33902/JPR.202318531
S6	Dos Santos et al. (2025)	Automatic Feedback in Mathematics Education: A Pathway to Robotics and CT	Acta Scientiae	Higher Ed (participants as reported in the article)	10.17648/acta.scientiae.8240
S7	Fang et al. (2025)	Design mobile CT-integrated mathematics lessons based on the 5E model for primary students	Educational Technology R&D	Primary (students)	10.1007/s11423-025-10513-x

ID	Author (Year)	Article Title	Journal	Level (participants)	DOI
S8	Molina-Ayuso et al. (2022)	Introduction to CT with Scratch for Teacher Training for Spanish Primary Teachers in Mathematics	Education Sciences	Primary & HE (pre-service primary teachers, n = 149)	10.3390/educsci12120899
S9	Sarmasági et al. (2025)	Bridging Algebraic and Computational Thinking: Impacts on Student Development in K–12	Informatics in Education	K-12 (students)	10.15388/infedu.2025.13
S10	Yunianto, Prodromou, et al. (2025)	Examining students' CT skills in enhanced-technology mathematics lessons: gender and grade effects	Discover Education	Secondary (students, Grades 7–10)	10.1007/s44217-025-00734-w
S11	Moon et al. (2023)	Developing preservice teachers' intuitions about CT in a math and science methods course	J. of Pedagogical Research	Higher Ed (pre-service teachers)	10.33902/JPR.202318599
S12	Suparman et al. (2025)	CT in Mathematics Instruction Integrated to STEAM Education: A Systematic Review & Meta-Analysis	TEM Journal	K-12 (general); meta-analysis	10.18421/TEM141-84
S13	Nordby et al. (2024)	Computational thinking in primary mathematics classroom activities	Frontiers in Education	Primary (two teachers' classroom activities)	10.3389/feduc.2024.1414081
S14	Park & Kwon (2022)	Research trends and issues including CT in science and mathematics education in Korea	J. Baltic Science Education	K-12; research-trend review (Korea)	10.33225/jbse/22.21.875
S15	Molina-Ayuso et al. (2024)	Computational Thinking with Scratch: A Tool to Work on Geometry in the Fifth Grade	Sustainability	Primary (Grade 5 students, n = 66)	10.3390/su16010110
S16	Sala-Sebastià et al. (2023)	Didactic–Mathematical–Computational Knowledge of Future Teachers When Solving Robotics Problems	Axioms	Higher Ed (prospective teachers)	10.3390/axioms12020119
S17	Ersozlu et al. (2023)	Developing Computational Thinking through Mathematics: An Evaluative Scientific Mapping	Education Sciences	Higher Ed; scientific mapping	10.3390/educsci13040422

ID	Author (Year)	Article Title	Journal	Level (participants)	DOI
S18	Fang et al. (2023)	Integrating CT into primary mathematics: A case study of fraction lessons with Scratch	Asian J. for Mathematics Education	Primary (Grade 4 students, Hong Kong)	10.1177/27527263231181963
S19	Mumcu et al. (2023)	Integrating CT into mathematics education through an unplugged computer science activity	J. of Pedagogical Research	Secondary & HE (middle school students, n = 80; pre-service teachers, n = 12)	10.33902/JPR.202318528
S20	Kaya et al. (2025)	The relationship between 21st-century skills and CT skills of prospective math and science teachers	J. of Pedagogical Research	Higher Ed (prospective teachers)	10.33902/JPR.202531498

Note. HE = higher education. S12, S14, and S17 are review-type studies; all other articles are empirical. Study identifiers (S1–S20) are used in the text and in Tables 3–6.

Table 3. Design, Sample, and Participant Type of the Included Articles, by Assigned Level

Assigned level and participants	Articles (n)	Design and sample as reported in this review	Contribution to indicator counts (Table 4)
Primary: students	5 (S1, S5, S7, S15, S18)	S1: design study, 15 Grade 5 students (3 analysed in detail); S5: quasi-experimental, non-equivalent groups; S7: design of mobile 5E CT lessons, with engagement and out-of-class practice reported; S15: quasi-experimental, 66 Grade 5 students in three schools, geometry achievement; S18: four-session design-based case study, 23 Grade 4 students	S1 and S18 report discrete indicators; S5, S7, and S15 report learning or engagement outcomes
Primary: teachers	2 (S2, S13)	S2: qualitative exploratory study, 13 in-service teachers in an online graduate course working on K-2 place-value activities; S13: classroom activities of two primary teachers	S2 reports discrete indicators; S13 reports none
Secondary: students	2 (S10, S19)	S10: Educational Design Research, 88 students in Grades 7–10 in four schools, non-parametric tests and Rasch analysis; S19: unplugged activity with 80 middle school students in four classes, implemented by 12 pre-service teachers	Both report discrete indicators
Secondary: teachers	2 (S3, S4)	S3: conceptions of CT among Grades 7–12 mathematics teachers; S4: mathematics teachers' knowledge, strategies, and challenges; sample sizes not reported in this synthesis	Neither reports discrete indicators
Higher education: prospective teachers or unspecified	5 (S6, S8, S11, S16, S20)	S6: automatic feedback in robotics-integrated mathematics, participants as reported in the article; S8: quasi-experimental, 149 pre-service primary teachers, CT test with a control group; S11: methods course for pre-service teachers; S16: prospective teachers solving robotics problems; S20: cross-sectional questionnaire study, 300 prospective mathematics and science teachers	S6, S8, and S20 report discrete indicators; S11 and S16 report integrated knowledge

Assigned level and participants	Articles (n)	Design and sample as reported in this review	Contribution to indicator counts (Table 4)
Cross-level: students	1 (S9)	Unplugged tasks linking algebraic and computational thinking, drawn from the curricula of six European countries and tested with students aged 9–14	Reports discrete indicators
Review-type	3 (S12, S14, S17)	S12: systematic review and meta-analysis of CT in STEAM-integrated mathematics instruction (K-12); S14: research-trend review of CT in science and mathematics education in Korea; S17: evaluative scientific mapping	Not counted; used as context only

Note. Levels follow the assignment rules described in the Method section; Table 2 shows the levels as labelled by each article. All articles were cross-sectional or short-term. Sample sizes are given only where they are reported in this review.

Data Extraction, Coding, and Synthesis

Data were analysed using directed content analysis (Krippendorff, 2004) with deductive and inductive coding. The four RQs served as initial deductive categories, and emergent themes enriched the analysis inductively. For each article, the following information was extracted into a coding sheet: (1) bibliographic data and country context; (2) educational level and type of participants (students, in-service teachers, or prospective teachers); (3) study type and design (empirical or review-type; qualitative, quantitative, or mixed, where reported); (4) sample size, where reported; (5) CT indicators examined, coded against the six facets of Shute et al. (2017), namely decomposition, abstraction, algorithm design, debugging, iteration, and generalisation, plus pattern recognition, which many included studies treat as a separate indicator; (6) instructional methods and tools; (7) main findings for each RQ; and (8) practical implications suggested by the authors.

Extracted findings were arranged in an evidence extraction matrix that links each synthesised claim to its supporting studies, to qualifying or contrasting studies, and to the type of evidence (Table 5). To avoid impressionistic judgements about which indicators "dominate" at each level, emphasis was operationalised as the number of articles, assigned to a level, that explicitly reported an indicator as a targeted focus or an observed outcome (Table 4). Articles spanning two levels were assigned to a single level according to the rules described under Study Classification and Rules for Comparison below. An indicator reported as a difficulty for participants was recorded separately. An indicator that no article reported at a level was recorded as "not reported", which does not mean that it is absent from practice. Counts are descriptive and were not weighted by sample size or study quality.

Study Classification and Rules for Comparison

Three rules governed how findings were categorised and compared. First, each article was classified as empirical or review-type (Table 3). Review-type articles (S12, S14, and S17) were retained as contextual evidence, were not included in the indicator counts (Table 4), and did not count towards the number of supporting studies in the grading criteria, so that no finding is counted twice through a review and the primary studies it may have synthesised. Whether these reviews synthesised primary studies that are also in the corpus was not checked.

Second, empirical articles were classified by participant type: students, in-service teachers, or prospective teachers. Findings about students' CT were compared only with other findings about students; findings about teachers' knowledge, conceptions, or CT performance were treated as a separate stream; and teacher-based evidence was never used alone to support a claim about students. Third, Table 2 shows the educational levels as labelled by each article, including articles that span two levels. For the indicator counts, a two-level article was assigned to the level whose participants or tasks produced the reported indicators: S2 to primary education, because in-service teachers worked on activities designed for pupils in the early grades; S8 to higher education, because its indicators came from a CT test taken by prospective teachers; and S19 to secondary education, because its indicators were reported for middle school students. S9, a K-12 study, was kept in a separate cross-level column. Comparisons between levels are descriptive only. No statistical comparison, pooling, or weighting was attempted because designs, samples, and instruments differ across studies (Table 3). This approach is consistent with the SWiM reporting guideline for syntheses without meta-analysis (Campbell et al., 2020), which asks reviewers to state how studies were grouped, which metric was used, and how certainty was assessed.

Appraisal of Evidence Strength

No formal risk-of-bias or quality-appraisal instrument was applied in this review, and this is acknowledged as a limitation. To keep claims proportionate to the evidence, every synthesised claim was instead graded using the number and consistency of supporting studies, the type of participants (students or teachers), and whether the claim was observed directly or inferred by comparing different studies. A claim was graded strong when at least three consistent student-based empirical studies from at least two countries supported it, no contradicting study existed, and at least one study used a comparison group or a large sample; moderate when at least two consistent empirical studies supported it, or when a claim that otherwise met the strong criteria was qualified by a contrasting or limiting study; and tentative when it rested on a single study, on teacher-only evidence for a student-level claim, on review-type studies alone, or on an inference from comparing different studies. These grades describe the strength of evidence within this corpus, not the truth of a claim.

The grades are a device for keeping claims proportionate and are not a substitute for a validated critical-appraisal instrument such as the Mixed Methods Appraisal Tool (MMAT) Hong et al., (2018), which was not applied. The review therefore cannot rule out that individual studies have methodological weaknesses, such as very small samples, the absence of comparison groups, or instruments of unreported validity, that bear on the claims they support. To let readers judge this for themselves, Table 3 reports the design, sample, and participant type of each article, and Table 5 lists qualifying and contrasting studies next to each claim.

RESULTS

RQ1: Characteristics of CT in Different Educational Levels

CT Profile at the Elementary School Level

Eight articles addressed primary-level contexts: six concerned primary education only and two also involved teacher education (Table 2). Their findings point to concrete, activity-based CT in which debugging, decomposition, abstraction, and algorithm design were each reported by two or more articles, while pattern recognition and iteration were each reported by one (Table 4). Yunianto, Jarvis, et al. (2025) studied 15 Grade 5 students in Indonesia in a design study in which students used GeoGebra commands to create mandalas, and analysed the work of three students in detail. They found three CT facets: pattern recognition (as a nascent form of abstraction), simple algorithm design, and debugging. For example, one student recognised that segment commands could be entered in any order as long as each pair of points was connected, and another reused an unpresented command form for later circles. Fang et al. (2023) conducted a four-session design-based study with 23 Grade 4 students in Hong Kong, in which students built a Scratch "fraction magic calculator" in groups and used computational abstractions in the form of variables, functions, and iterations to formulate mathematical models. Students were also engaged in decomposing fraction problems and debugging, with scaffolding provided by the 5E instructional model.

Molina-Ayuso et al. (2024) conducted a quasi-experimental study with Grade 5 students (66 students from three schools) and reported a more positive learning process in geometry among students who worked with Scratch; the study measured geometry achievement and did not report individual CT indicators. Nordby et al. (2024), studying two Norwegian primary teachers who used a range of CT tools in mathematics lessons, raised a pedagogical concern: the balance between CT and mathematics tended to tip towards CT, to the detriment of mathematical substance, which points to a genuine tension in primary CT integration. Kaup et al. (2023) examined, in a quasi-experimental study with a non-equivalent group design, whether a CT intervention in number and arithmetic, algebra, and geometry affected students' learning performance in primary schools. Dahshan and Galanti (2024) provided complementary evidence from the teacher side. In a qualitative exploratory study, 13 in-service elementary teachers in an online graduate CT course explored Scratch place-value activities designed for the early grades (K-2). Most groups applied decomposition effectively and understood the program's algorithm, and most understood abstraction, although one group struggled to translate its understanding of abstraction; three groups identified the logic error in the program, but their fixes were incomplete. Their work suggests that designing CT-integrated primary mathematics activities also requires building teachers' own CT understanding. Overall, primary-level evidence is the most abundant in the corpus, but it comes from small, single-context studies, and three of the eight articles (S2, S8, and S13) centre on teachers rather than pupils.

CT Profile at the Secondary School Level

Evidence at the secondary level is limited to four articles (Table 2): two involving students (Yunianto, Prodromou, et al., 2025; Mumcu et al., 2023) and two involving teachers (Humble & Mozelius, 2023; Djidu et al., 2025). The patterns below are therefore indicative rather than conclusive. Yunianto, Prodromou, et al. (2025) used Educational Design Research with 88 students in Grades 7–10 (ages 12–16) in four Indonesian schools, analysing GeoGebra-based CT

tasks with non-parametric tests and Rasch analysis. They reported no significant overall difference between male and female students, although a few items functioned differently by gender, and a significant, large difference by grade level (the lowest mean score in Grade 7, the highest in Grade 8, and a slightly lower score in Grade 10 than in Grade 8). Task-level analysis addressed algorithm design, pattern recognition (as a form of abstraction), and debugging, and grade differences appeared in all three. Mumcu et al. (2023) developed an unplugged computer science activity that 12 pre-service mathematics teachers implemented with 80 middle school students in four classes. Students performed data collection, data analysis, decomposition, pattern recognition, algorithm design, and testing and debugging, and said they enjoyed the activity, although they needed help connecting it with computer science concepts.

Humble and Mozelius (2023) examined the conceptions of CT held by mathematics teachers in Grades 7–12 and found considerable conceptual variation, with some participants viewing CT as computer programming and others as a more general reasoning ability. The authors argue that this inconsistency among teachers risks fragmented and suboptimal integration of CT in classrooms. Sarmasági et al. (2025), a K-12 study, developed unplugged tasks linking algebraic thinking (AT) and CT, drawing on curricula from six European countries and testing the tasks with students aged 9–14. Their framework explicitly maps AT generalisation onto CT generalisation, and the study reports that the integrated learning pathways enhanced proficiency in both AT and CT, although they required extended practice time. This suggests the importance of curriculum coherence across the grades.

CT Profile at Higher Education Level

Nine articles involve higher-education or teacher-education contexts (Table 2), but much of this evidence concerns prospective or in-service teachers (Moon et al., 2023; Sala-Sebastià et al., 2023; Kaya et al., 2025) rather than university students in general. The profile described here is therefore mainly a profile of prospective teachers. Within this evidence, CT is described as increasingly integrated with mathematical and pedagogical knowledge and with higher-order thinking. In a quantitative study of 300 prospective mathematics and science teachers in Turkey, Kaya et al. (2025) found that CT levels differed significantly by gender, semester, and GPA. Male prospective teachers scored higher on creativity and critical thinking, while no significant gender difference was found on algorithmic thinking or problem solving. With a cross-sectional design, this pattern cannot be attributed to educational experience or to socio-cultural factors. CT levels were higher in later semesters, which may reflect the cumulative effect of formal education, although comparisons between different cohorts cannot confirm this, and the pattern was not linear across all dimensions.

Sala-Sebastià et al. (2023) observed that prospective teachers who integrated didactic, mathematical, and computational knowledge when solving robotics problems showed a more comprehensive understanding of how mathematical concepts can be externalised through computational artefacts. Molina-Ayuso et al. (2022), in a quasi-experimental study of 149 pre-service primary teachers, found a larger gain in a CT test (covering sequences, loops, conditionals, and functions) among those who completed Scratch practices than among a control group. A scientific mapping of CT research in higher education by Ersozlu et al. (2023) revealed a growing focus on the links between CT and higher-order thinking skills, including critical thinking and mathematical problem solving, which may indicate a shift from viewing CT as a technical skill towards viewing it as an advanced cognitive capability. Dos Santos et al. (2025) reported that iteration and generalisation, both hallmark CT capacities, were developed through adaptive automated feedback in robotics-integrated mathematics education.

RQ2: Level-Based Prioritisation of CT Indicators

RQ2 asked how the emphasis of CT indicators differs by educational level. Table 4 summarises, for each indicator and level, the number of articles that explicitly reported it, using the operationalisation described in the Method section. The counts are small and the articles differ in design and instruments, so no indicator can be called dominant at any level, and the patterns below are interpreted cautiously. At the primary level, debugging was reported by three articles, decomposition, abstraction, and algorithm design by two each, and pattern recognition and iteration by one each. At the secondary level, algorithm design, pattern recognition, and debugging were each reported by both student-based articles (S10, S19), and decomposition by one (S19). Iteration was reported by one primary article (S18) and by two higher-education articles (S6, S8), and generalisation only by one higher-education article (S6) and one cross-level article (S9). The higher-education articles otherwise emphasised integrated constructs such as creativity, critical thinking, and the integration of didactic, mathematical, and computational knowledge (S16, S17, S20) rather than discrete indicators, and no higher-education or cross-level article reported debugging.

Table 4. Number of Articles Reporting Each CT Indicator, by Educational Level

CT indicator	Primary	Secondary	Higher education	Cross-level (K-12)
Decomposition	2 (S2(T), S18)	1 (S19)	–	1 (S9)
Pattern recognition	1 (S1)	2 (S10, S19)	–	1 (S9)
Abstraction	2 (S2(T), S18)	–	–	1 (S9)
Algorithm design	2 (S1, S2(T))	2 (S10, S19)	2 (S8(T), S20(T))	1 (S9)
Debugging	3 (S1, S2(T), S18)	2 (S10, S19)	–	–
Iteration	1 (S18)	–	2 (S6, S8(T))	–
Generalisation	–	–	1 (S6)	1 (S9)
Empirical articles assigned / reporting at least one indicator	7 / 3 (5 student, 2 teacher)	4 / 2 (2 student, 2 teacher)	5 / 3 (prospective teachers; S6 unspecified)	1 / 1 (student)

Note. Cells give the number of articles that explicitly reported the indicator as a targeted focus or an observed outcome, followed by study identifiers from Table 2. "–" = not reported in the extracted data (this does not mean the indicator is absent in practice). (T) = teacher or prospective-teacher participants. One article can report several indicators. For S8, test items on sequences were coded as algorithm design and items on loops as iteration; for S1 and S10, pattern recognition reported as a form of abstraction was coded as pattern recognition. The final row gives the number of empirical articles assigned to each level and, after the slash, how many of them reported at least one discrete indicator; S8 is counted under higher education because its indicators come from prospective teachers' CT test. Review-type articles (S12, S14, S17) are not included in any count. Counts are descriptive and not weighted by sample size or study quality.

The final row of Table 4 gives the denominators for each column. Only three of seven empirical articles at primary level, two of four at secondary level, and three of five at higher-education level reported any discrete indicator, so most articles contribute nothing to the counts. Restricting attention to student-based evidence, discrete indicators come from two primary articles (S1, S18), two secondary articles (S10, S19), and one cross-level article (S9). No higher-education article reported discrete indicators for students; that column reflects prospective teachers' performance on a CT test (S8), a questionnaire scale (S20), or participants not specified in the article (S6). Comparisons between primary and secondary levels therefore rest on two student-based articles each, and any comparison with higher education contrasts students with prospective teachers measured by different instruments. Differences between columns are accordingly differences between sets of studies, not between age groups.

Three patterns follow from Table 4. First, discrete indicators were reported mostly in primary- and secondary-level student studies, and in similar sets: algorithm design and debugging appear at both levels. In the available data, the profile of students' CT at these two levels is therefore more alike than different. Second, iteration and generalisation were reported mainly by higher-education or cross-level articles (Dos Santos et al., 2025; Molina-Ayuso et al., 2022; Sarmasági et al., 2025), which is compatible with the view that more abstract indicators receive more emphasis at later levels. This reading is tentative because each pattern rests on one or two articles and one primary article (Fang et al., 2023) also reported iteration. Third, higher-education articles emphasise integrated and higher-order constructs: Kaya et al. (2025) reported associations between CT and the creativity and critical thinking dimensions, and Sala-Sebastià et al. (2023) described the integration of didactic, mathematical, and computational knowledge. Piagetian stage explanations (concrete-operational thinking at primary level and formal-operational thinking at secondary level) are an interpretive frame that the included studies did not test.

RQ3: What the Evidence Suggests About the Developmental Path of CT

RQ3 asked what the evidence suggests about the developmental path of CT from primary school to higher education. No included study followed the same learners across levels, and the studies differ in participants, instruments, and designs. The trajectory described below is therefore an inference from cross-study comparison, graded tentative in Table 5, and not an empirically observed developmental path.

In the primary phase (ages 6–12), the reviewed studies describe CT as experiential. Learners work with pattern recognition and decomposition through the manipulation of concrete or virtual objects; debugging is a natural response to immediate feedback from digital or physical environments; and algorithms are simple, linear, and context-specific. This is compatible with Piagetian concrete operational development, in which children need tangible representations to build conceptual understanding (Yunianto, Jarvis, et al., 2025; Fang et al., 2023, 2025), although these studies did not test Piagetian stages.

In the secondary phase (ages 12–18), the evidence is thin (four articles coded as secondary). The two student-based studies report algorithm design, pattern recognition, and debugging in GeoGebra-based and unplugged tasks (Yunianto, Prodromou, et al., 2025; Mumcu et al., 2023), and the cross-level study of Sarmasági et al. (2025) links algebraic and computational thinking through generalisation. These findings are compatible with a gradual move from concrete towards symbolic-formal representation, but they do not show it, and none followed the same learners over

time. Park and Kwon (2022) analysed trends in Korean CT research and found that secondary education has historically received the most research attention, which suggests that it is recognised as a crucial stage for more formal CT development in that context; this contrasts with the thin secondary-level evidence in the present corpus.

In higher education (age 18+), the evidence, drawn mostly from prospective teachers, points to CT that is linked to creativity, critical thinking, and the integration of mathematical, didactic, and computational knowledge (Kaya et al., 2025; Sala-Sebastià et al., 2023; Ersozlu et al., 2023). Kaya et al. (2025) reported higher CT levels in later semesters, which is compatible with a cumulative educational effect, but the pattern differed across dimensions, so the data do not show that growth occurs automatically. Whether CT becomes an internalised, flexible mode of thinking at this level cannot be established from the included studies, which did not assess it directly.

A possible discontinuity at the primary-to-secondary transition emerges from this comparison, but it is best treated as a hypothesis. It rests on the contrast between the relatively rich indicator-level evidence at primary level and the thinner, more varied evidence at secondary level, on the conceptual variation among Grades 7–12 teachers (Humble & Mozelius, 2023), and on the finding that integrated algebraic–computational learning pathways enhanced proficiency in both domains, which points to the value of a coherent progression (Sarmasági et al., 2025). No included study measured the same learners across this transition, and the smaller number of secondary-level studies may itself explain part of the apparent gap. The term "CT gap" is therefore used in this article as a working hypothesis for future longitudinal research, not as an established finding.

RQ4: Systemic and pedagogical influences on the development of CT

Technology as a Mediator of CT

Many articles in the corpus explicitly or implicitly address the role of technology in enabling CT development. The reviewed studies suggest that the kind of technology and the way it is used, rather than technology use per se, shape which dimensions of CT are developed. Block-based Scratch activities were associated with gains in mathematics learning or CT-test performance, or with the use of abstractions such as variables, functions, and iterations, in primary classrooms and in teacher education (Fang et al., 2023; Molina-Ayuso et al., 2022, 2024). Text-based GeoGebra commands were used with primary students to elicit algorithm design, pattern recognition, and debugging in mathematical construction tasks (Yunianto, Jarvis, et al., 2025). None of these studies compared platforms directly. Unplugged approaches also show pedagogical promise: in Mumcu et al. (2023), middle school students performed data collection, data analysis, decomposition, pattern recognition, algorithm design, and testing and debugging during an unplugged activity, and said they enjoyed it. Fang et al. (2025) reported that mobile-based CT learning with the 5E model increased student engagement and the frequency of out-of-class CT practice among primary learners. Robotics systems and automated feedback were examined mainly at the higher-education level, where Dos Santos et al. (2025) reported that adaptive automated feedback supported independent iteration and generalisation, which are hallmark CT capacities.

Educational Design & Instructional Techniques

Instructional design is frequently identified as a major pedagogical determinant of CT quality. Across the corpus, there are several effective design principles. Firstly, the principle of the "half-baked artifact" (Kynigos, 2007, used by Yunianto, Jarvis, et al., 2025) was reported to be effective in stimulating CT: presenting deliberately incomplete artefacts challenged students to complete, explore and modify, and stimulated algorithmic creativity and debugging without overly prescribing solutions. Second, the 5E instructional model (Engage, Explore, Explain, Elaborate, Evaluate) used by Fang et al. (2023, 2025) created learning sequences intended to support the holistic, progressive development of CT, in which the Explore phase targets decomposition and pattern recognition, Elaborate targets generalisation and abstraction, and Evaluate targets debugging and reflection. Third, several included studies used collaborative, group-based tasks (Dahshan & Galanti, 2024; Fang et al., 2023), but few compared instructional approaches directly, so the relative effectiveness of collaborative and problem-based approaches remains untested within this corpus. Fourth, studies suggest that putting CT into culturally relevant or personally meaningful contexts increases student engagement and motivation. For example, Yunianto, Jarvis, et al. (2025) incorporated mandala art, a culturally important art form across Asian contexts, into mathematics CT activities and reported that students' CT skills were supported through engagement with the activities. This suggests that CT need not be acultural; CT based on familiar cultural contexts can be a useful entry point for learners from diverse backgrounds.

Teacher Preparation and Professional Development

Teacher capacity is the most commonly identified barrier and enabler of CT implementation. Djidu et al. (2025) identified three important dimensions of teacher readiness for CT: knowledge of CT content (knowing what CT is), knowledge of instructional strategies (knowing how to teach CT) and awareness of challenges (knowing why CT is

hard to implement). In their study, most teachers had fragmented understandings, often reducing CT to programming, a misconception that, according to the authors, affects the quality of implementation. Humble and Mozelius (2023) confirmed that teachers of Grades 7-12 perceived CT differently and that their disciplinary background was a factor in this perception: mathematics teachers who understood formal algorithmic thinking leaned towards a more thorough understanding of CT than science teachers. Moon et al. (2023) demonstrated that direct CT experience in teacher education methods courses changed prospective teachers' intuitions about CT, and increased their readiness to integrate CT. Molina-Ayuso et al. (2022), in a quasi-experimental study of 149 pre-service primary teachers, found a larger CT-test gain after Scratch practices than in a control group, and participants described the practices as positive, useful, and practical.

Demographics and Affective Factors

Affective and demographic variables are important and complex in the development of CT. Kaya et al. (2025) found that the components of 21st century skills including entrepreneurship and innovation, social responsibility and leadership, and information and technology literacy accounted for 57% of the variance in prospective teachers' CT levels. This finding suggests that CT may be best developed as part of the broader ecosystem of 21st-century skills, although a correlational result of this kind does not establish causation. Gender was a constant but nuanced variable. At the secondary level, gender differences were limited to a few items (Yunianto, Prodromou, et al., 2025), whereas at the higher-education level significant gender differences were reported on some dimensions (Kaya et al., 2025). However, Yunianto, Prodromou, et al. (2025) found no significant overall gender difference, whereas Kaya et al. (2025) found differences on some dimensions (creativity and critical thinking) but not on algorithmic thinking or problem solving. Gender effects therefore appear to be dimension- and context-specific, and neither study tested interventions aimed at reducing such gaps. Sarmasági et al. (2025) emphasised the importance of designing learning pathways: students who followed integrated learning pathways linking algebraic and computational thinking enhanced their proficiency in both, although the pathways required extended practice time; this points to the potential value of a coherent, integrated curriculum design.

Summary of Evidence Strength

Table 5 links each synthesised claim to its supporting and qualifying studies and grades its strength using the criteria described in the Method section. No claim met all the criteria for strong evidence in the extracted data. Claims about primary-level characteristics and indicators, secondary-level indicators, Scratch-based activities, and teachers' conceptions and preparation reached the moderate level. Claims about higher education, about cross-level progression and discontinuity, and about instructional design, gender, and integrated algebra–CT pathways remain tentative, because they rest on single studies, on teacher-based samples, or on comparisons across different studies. Grades should be read together with Table 3, which shows the design, sample, and participant type behind each claim, and claims graded tentative are best read as hypotheses.

Table 5. Evidence Extraction Matrix: Synthesised Claims, Supporting Studies, and Evidence Strength

Claim (research question)	Supporting studies	Qualifying or contrasting studies	Evidence type	Grade
(RQ1) Primary-level CT is evidenced mainly through concrete, activity-based tasks.	S1, S5, S7, S18	S13: CT–mathematics balance tipped towards CT	Student-based empirical; small, single-context studies	Moderate
(RQ2) Debugging, decomposition, abstraction, and algorithm design are each reported by two or more primary-level articles; pattern recognition and iteration by one.	S1, S2(T), S18	S2(T): one group struggled with abstraction; debugging fixes incomplete	Student- and teacher-based empirical, small samples	Moderate
(RQ1–2) At secondary level, algorithm design, pattern recognition, and debugging are reported by both student-based studies, with performance differing by grade level (S10).	S10, S19	S3(T), S4(T): teachers' conceptions of CT vary	Two student-based studies; cross-sectional	Moderate

Claim (research question)	Supporting studies	Qualifying or contrasting studies	Evidence type	Grade
(RQ1–2) Higher-education evidence (mostly prospective teachers) links CT to creativity, critical thinking, and integrated knowledge, and reports gains in sequences and loops (S8) and in iteration and generalisation (S6).	S6, S8(T), S16(T), S17, S20(T)	S11(T)	Mainly prospective-teacher samples; one survey, one mapping, one quasi-experiment	Tentative
(RQ2) Emphasis moves from discrete indicators (primary, secondary) to integrated, higher-order constructs (higher education).	S1, S18, S10, S19 compared with S6, S8(T), S16(T), S17, S20(T)	Different samples and instruments; no common measure	Cross-study inference	Tentative
(RQ3) Findings across levels are compatible with a concrete-procedural to abstract-conceptual progression; it has not been observed in the same learners.	S1, S6, S10, S19, S20(T)	Not tested directly; S14 shows research attention concentrated at secondary level in Korea	Cross-study inference; no longitudinal data	Tentative
(RQ3) A discontinuity may exist at the primary-to-secondary transition (working hypothesis).	S3(T), S9	Thin secondary evidence may explain the apparent gap	Inference; working hypothesis	Tentative
(RQ4) Scratch-based activities were associated with gains in mathematics learning or CT-test performance.	S8(T), S15, S18	S13: CT–mathematics balance tipped towards CT; no direct platform comparison	Quasi-experimental and design studies	Moderate
(RQ4) Instructional design (5E model, half-baked artefacts, group-based tasks, culturally relevant contexts) supports CT.	S1, S7, S18	Few direct comparisons of instructional approaches	Mostly single-group classroom studies	Tentative
(RQ4) Teachers' conceptions of CT are fragmented (S3, S4), and hands-on CT experience improves prospective teachers' CT understanding or performance (S8, S11).	S3(T), S4(T), S8(T), S11(T)	–	Teacher-based empirical	Moderate
(RQ4) Gender differences in CT are small and dimension-specific; grade level matters.	S10, S20(T)	S10 found no overall gender difference; S20 found differences on some dimensions	Two studies; one correlational	Tentative
(RQ4) Integrated learning pathways linking algebraic thinking and CT enhanced proficiency in both.	S9	–	Single K-12 study	Tentative

Note. (T) = teacher or prospective-teacher participants. Grades follow the criteria in the Method section and describe the strength of evidence within this corpus. Review-type articles (S12, S14, S17), where listed, provide context only and did not count towards the grade criteria.

DISCUSSION

CT as a Developmental Concept: Critical Analysis

Taken together, the 20 articles are compatible with a view of CT as a developmental concept that evolves alongside learners' broader cognitive and mathematical development, but they do not demonstrate it. This view challenges reductionist notions that CT is a separate technical skill learned in isolation, and it is aligned with Grover and Pea's (2013) perspective of CT as a cluster of concepts and practices that develops over time rather than a set of discrete competencies acquired in a linear fashion. The comparison across levels must, however, be read in light of differences in research designs, samples, instruments, and evidence quality. Primary-level evidence comes mostly from small-scale, single-context studies with students; secondary-level evidence is scarce and includes two teacher-perception studies; and higher-education evidence is dominated by prospective teachers, a cross-sectional survey of 300 participants (Kaya et al., 2025), and a scientific mapping (Ersozlu et al., 2023). Measures range from classroom observation and artefact analysis to questionnaire scales and bibliometric mapping. Differences between levels may therefore reflect what was measured and how, as well as learners' development, and interpretations of progression across levels are necessarily weaker than they would be with longitudinal observation of the same learners.

A cautious implication is that introducing formal abstraction too early may overload younger learners, whereas deferring richer CT components to later levels may leave students less prepared for them. Neither possibility was tested directly in the included studies; the abstraction difficulty observed in one teacher group by Dahshan and Galanti (2024), for example, concerns teachers rather than pupils.

The connection between CT and mathematical thinking also warrants a nuanced discussion. Ye et al. (2023) argued that CT and mathematical thinking share much but also retain their own domains. To design an effective curriculum, these boundaries and intersections need to be understood. Over-emphasising CT components that are already part of mathematical thinking wastes opportunities to create genuinely new CT capacities, while emphasising CT at the expense of mathematical substance, as observed in primary classrooms by Nordby et al. (2024), can weaken the mathematical coherence of lessons. Curriculum designers and teachers therefore need deep knowledge of both CT and mathematical epistemology, a level of expertise that is often not sufficiently developed in current teacher education programmes.

Differences and Tensions Among the Findings

The included studies do not simply converge, and several divergences limit how far their findings can be combined. Gender is one example. Yunianto, Prodromou, et al. (2025) found no significant overall gender difference among Grade 7–10 students, apart from a few items that functioned differently, whereas Kaya et al. (2025) found differences in creativity and critical thinking, but not in algorithmic thinking or problem solving, among prospective teachers. The two studies differ in age group, instrument (Rasch-analysed tasks versus a questionnaire scale), and design, so the discrepancy cannot be attributed to educational level or to gender itself. Evidence on Scratch is a second example. Molina-Ayuso et al. (2024) measured geometry achievement without reporting individual CT indicators, Molina-Ayuso et al. (2022) measured CT-test performance among prospective teachers, and Fang et al. (2023) described students' use of abstractions in a design-based case study. These studies are consistent in direction, but they concern different outcomes and participants, so the moderate grade in Table 5 reflects a consistent association across heterogeneous outcomes, not a common effect size. A third example is the balance between CT and mathematics: Nordby et al. (2024) observed CT overshadowing mathematical substance, a tension that no other included study was coded as reporting, so it is unclear whether it is general or specific to the two classrooms studied.

Divergence also appears between the corpus and the wider literature. Park and Kwon (2022) found that secondary education received the most research attention in Korea, whereas only four articles in the present corpus concern secondary education; this difference may reflect the national scope of the Korean review and the language, access, and topic filters of the present review as much as any difference in research volume. Kaya et al. (2025) reported higher CT levels in later semesters, but not uniformly across dimensions, and because the design compares cohorts rather than following students, the pattern cannot be read as growth. The clearest convergence concerns teachers: Humble and Mozelius (2023) and Djidu et al. (2025) both describe fragmented conceptions of CT among mathematics teachers, and Moon et al. (2023) and Molina-Ayuso et al. (2022) report improvements after direct CT experience among prospective teachers. This agreement across four studies with different designs and participants is the main reason the corresponding claims reach the moderate grade, and it contrasts with the tentative status of most claims about learners' development.

Generalisability of the Findings

The findings generalise to the reviewed corpus and, with caution, as hypotheses to similar settings; they do not describe learners in general. The student-based evidence is small: the primary-level student articles include samples of 15 students (three analysed in detail), 23 students, and 66 students, and the two secondary-level student articles involve 88 and 80 students. The only large sample is a cross-sectional survey of 300 prospective teachers in Turkey (Kaya et al., 2025), and the one quasi-experimental study at higher-education level involved Spanish pre-service primary teachers (Molina-Ayuso et al., 2022). Three articles report Indonesian samples (Yunianto, Jarvis, et al., 2025; Yunianto, Prodromou, et al., 2025; Djidu et al., 2025), and the corpus as a whole is weighted towards Western and East Asian contexts. Further constraints come from the search: English-language, open-access, Scopus-indexed articles from 2022–2025, with CT operationalised through the six facets of Shute et al. (2017) plus pattern recognition. Practitioners in other settings, including the Global South, should therefore treat the indicator profiles as examples of what has been reported, and the cross-level propositions as questions to test locally, not as norms.

A Conceptual Spiral-Expansive Model of CT Development

Drawing together the four RQs, this review proposes a conceptual spiral-expansive model of CT development in mathematics education. The model is an interpretation of the cross-study pattern, intended to generate testable hypotheses. It has not been empirically validated, and each proposition is set against the evidence in the corpus in

Table 6. It is offered as a conceptual framework for organising hypotheses and designing future studies, not as an established developmental model, and the corpus does not allow a claim that CT develops in this way. In contrast to linear models, the model conceptualises CT as a spiral that becomes more complex with each educational level. Its propositions are: (P1) recurrence, because basic CT components such as pattern recognition, decomposition, and algorithm design are introduced early and recur at later levels with more complex mathematical content and representations, so that no component is "completed" at one level and discarded at the next; (P2) expansion, because additional components such as iteration, generalisation, and integrated higher-order thinking become more prominent at later levels; (P3) variability, because individual, cultural, and contextual factors may accelerate or shift the pace of development of each component; and (P4) continuity, because coherent tools, representations, and pedagogy across level transitions may sustain development, whereas abrupt changes may disrupt it. Table 6 specifies, for each proposition, what would count as evidence, how well the present corpus supports it, and how it could be tested.

Table 6. Propositions of the Spiral-Expansive Model, Operational Indicators, and Proposed Validation

Proposition	Operational indicator (what would count as evidence)	Support in the present corpus	Proposed validation
P1 Recurrence	The same CT indicator is observed at two or more levels, with higher task complexity or more symbolic representations at later levels	Algorithm design reported at primary (S1, S2(T)), secondary (S10, S19), and higher-education (S8(T), S20(T)) levels. Tentative: indicator levels are not comparable across studies	Longitudinal or cross-sequential design using a common, level-adapted CT assessment
P2 Expansion	Iteration, generalisation, and integrated higher-order constructs are reported more often, or at higher performance levels, at later levels	Iteration and generalisation reported mainly by S6, S8(T), and S9; creativity and critical thinking linked to CT in S17 and S20(T). Tentative	Indicator-by-level assessment across cohorts; growth-curve modelling
P3 Variability	Timing or rate of indicator development differs by gender, grade, culture, or instructional context	Grade-level differences (S10); gender differences small and dimension-specific (S10, S20(T)); culturally meaningful task (S1). Tentative	Multi-group or moderation analyses in large samples
P4 Continuity	Learners who experience coherent tools and pedagogy across a level transition show smoother development than those who do not	Integrated algebraic–computational pathways (S9); varied teacher conceptions (S3(T)). No direct test; tentative	Quasi-experimental comparison of cohorts with and without a bridging curriculum at the transition

Theoretical Contribution and Its Limits

The theoretical contribution of this review is deliberately modest and is stated in proportion to the evidence. First, it specifies what cross-level evidence would be needed before a developmental claim can be made: observations of the same learners across a level transition, a common level-adapted measure, and comparisons of students with students rather than with prospective teachers (Table 6). Second, it shifts the question from how CT develops to how strongly each cross-level claim is supported, and shows that most such claims are currently tentative (Table 5). Third, it offers the spiral-expansive model as four propositions with operational indicators that can be falsified: P1 would be undermined if the same indicators were not observed at later levels, P2 if emphasis did not shift towards integrated constructs, P3 if the timing of development did not vary across groups or contexts, and P4 if cohorts with and without a bridging curriculum did not differ at the transition. Until such tests are carried out, the model remains conceptual and should not be cited as evidence of a developmental trajectory.

Implications for Curriculum and Teacher Education

The implications below are linked to the strength of the supporting evidence in Table 5 and should be read as cautious recommendations. Recommendations resting on moderate evidence (the first and third) are reasonable bases for practice, whereas the second rests on tentative evidence and is best treated as a design principle to be piloted and evaluated. First, CT integration can begin in the early primary grades with an emphasis on concrete, meaningful

activities involving pattern recognition, decomposition, and debugging (moderate evidence); Fang et al. (2025) showed that 5E CT lessons can be implemented on mobile devices with primary learners. Second, curricula should be designed with explicit cross-level continuity (tentative evidence). Teacher reports from Indonesia suggest that the integration of CT in mathematics remains uneven (Djidu et al., 2025), and a coherent framework that defines which CT elements are focal at each level and how they reinforce one another would be a reasonable response, although the benefit of cross-level continuity itself still needs empirical testing. Sarmasági et al. (2025) offer a concrete example of explicit bridges between algebraic and computational thinking in K-12 curricula. Third, teacher professional development should be a priority (Djidu et al., 2025; Humble & Mozelius, 2023; Moon et al., 2023; Dahshan & Galanti, 2024). The moderate evidence on teachers' fragmented conceptions of CT and on the effect of direct CT experience suggests that effective programmes should position teachers as active learners within their own disciplines and levels of education, not as passive recipients.

Limitations and Future Research

This review's conclusions are bounded by several limitations. First, all analysed articles are cross-sectional or short-term, and no study traced the same learners from primary to higher education. The reconstruction of the developmental trajectory is therefore inferential, not directly empirical, and the most important next step for the field is longitudinal research that tracks CT continuity across levels. Second, the search was restricted to one database (Scopus), to open-access English-language journal articles, and to a four-year window. Of 1,155 records identified, 213 remained after the document-type, language, subject-area, and access filters, so relevant paywalled, non-English, or non-Scopus studies, as well as grey literature, may have been missed, and publication bias cannot be ruled out. The size and direction of the resulting selection bias cannot be estimated from this review. Future reviews should search additional databases such as Web of Science and ERIC and should not restrict by access type. Third, no formal quality or risk-of-bias assessment was applied; claims were instead graded by the number, type, and consistency of supporting studies (Table 5), which is a coarser safeguard. Applying an established appraisal instrument such as the MMAT (Hong et al., 2018) with two independent raters would strengthen future reviews. Fourth, the evidence is unevenly distributed: only four articles concern secondary education, and the higher-education evidence is dominated by prospective teachers, so cross-level comparisons mix evidence about students with evidence about teachers and generalise poorly to learners as a whole. Assigning two-level articles to a single level for Table 4 also required judgement, and a different assignment would change some counts. Review-type articles (S12, S14, S17) were not used to grade claims, but whether they synthesised primary studies that are also in the corpus was not checked, so some overlap cannot be excluded. Fifth, no review protocol was prepared, screening and coding were carried out by one author, and co-authors validated only a sample of the included articles without an inter-rater agreement statistic, which limits reproducibility compared with fully independent double screening and coding. Sixth, even the articles that include several countries are biased towards Western and East Asian contexts, which limits generalisability to educational systems in the Global South. Seventh, CT measurement instruments vary between studies and are not yet standardised, which restricts cross-study comparison; developing valid, reliable, and level-sensitive CT assessment tools remains an urgent research need (Ersozlu et al., 2023). Finally, the spiral-expansive model proposed here has not been empirically validated.

CONCLUSION

This Systematic Literature Review synthesised 20 Scopus-indexed, open-access articles published between 2022 and 2025 on CT in mathematics education across educational levels. The conclusions below separate findings supported by the reviewed studies from interpretations that go beyond them, and each is qualified by the evidence grades in Table 5.

Findings supported by the reviewed studies. RQ1: Primary-level studies describe concrete, activity-based CT. Evidence at the secondary level is limited to four articles; the two student-based studies report algorithm design, pattern recognition, and debugging, and teachers' conceptions of CT vary. Higher-education evidence comes mostly from prospective teachers and links CT to creativity, critical thinking, and integrated knowledge. RQ2: No indicator clearly dominated at any level. Debugging, decomposition, abstraction, and algorithm design were each reported by two or more primary-level articles, algorithm design, pattern recognition, and debugging by both student-based secondary studies, and iteration and generalisation mainly by higher-education or cross-level articles. RQ4: Technology type, instructional design, teacher preparation, and affective and demographic factors were the factors most often associated with CT development. Evidence on Scratch-based activities and on teachers' conceptions and preparation is moderate, whereas evidence on instructional design and on gender and other demographic factors is tentative.

Conceptual interpretations. RQ3: The concrete-to-abstract progression, the spiral-expansive pattern, and the possible discontinuity at the primary-to-secondary transition are interpretations inferred by comparing different studies with different participants and instruments. They are hypotheses to be tested in longitudinal research and should not be read as demonstrated developmental trajectories.

Implications and limits. Curriculum developers may use these findings as a cautious basis for designing coherent, cross-level CT frameworks that are embedded in, rather than added on to, mathematics curricula, while testing the continuity assumptions empirically. For teacher educators, the moderate evidence on teachers' conceptions and preparation supports immersive, discipline-specific CT professional learning that gives teachers direct, hands-on CT experience. For researchers, the most pressing needs are longitudinal studies that follow the same learners across levels, studies at the secondary level and in culturally diverse contexts, and standardised, level-sensitive CT assessment. The conclusions are limited by the single-database, open-access, English-language search, by the absence of a formal quality appraisal, and by evidence that is dominated by teacher samples at higher-education level.

REFERENCE

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, *368*, Article l6890. https://doi.org/10.1136/bmj.l6890

Dahshan, M., & Galanti, T. (2024). Teachers in the loop: Integrating computational thinking and mathematics to build early place value understanding. *Education Sciences*, *14*(2), Article 201. https://doi.org/10.3390/educsci14020201

Djidu, H., Retnawati, H., & Haryanto. (2025). Mathematics teachers' perspectives on computational thinking: Insights into knowledge, strategies, and challenges. *The Qualitative Report*, *30*(7), 4019–4040. https://doi.org/10.46743/2160-3715/2025.6494

Dos Santos Dos Santos, J. M., Almeida Pereira Abar, C. A., Reis D'Azevedo Breda, A. M., & Lavicza, Z. (2025). Automatic feedback in mathematics education: A pathway to robotics and computational thinking. *Acta Scientiae*, *27*(2). https://doi.org/10.17648/acta.scientiae.8240

Ersozlu, Z., Swartz, M., & Skourdoumbis, A. (2023). Developing computational thinking through mathematics: An evaluative scientific mapping. *Education Sciences*, *13*(4), Article 422. https://doi.org/10.3390/educsci13040422

Fang, X., Ng, D. T. K., Tam, W. T., & Yuen, M. (2023). Integrating computational thinking into primary mathematics: A case study of fraction lessons with Scratch programming activities. *Asian Journal for Mathematics Education*, *2*(2), 220–239. https://doi.org/10.1177/27527263231181963

Fang, X., Ng, D. T. K., Tam, W. T., & Yuen, M. (2025). Design mobile computational thinking-integrated mathematics lessons based on the 5E instructional model for primary students. *Educational Technology Research and Development*, *73*(4), 2809–2838. https://doi.org/10.1007/s11423-025-10513-x

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher*, *42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Humble, N., & Mozelius, P. (2023). Grades 7–12 teachers' perception of computational thinking for mathematics and technology. *Frontiers in Education*, *8*, Article 956618. https://doi.org/10.3389/feduc.2023.956618

ISTE & CSTA. (2011). *Operational definition of computational thinking for K–12 education*. International Society for Technology in Education & Computer Science Teachers Association. https://cdn.iste.org/www-root/Computational_Thinking_Operational_Definition_ISTE.pdf

Kaup, C. F., Pedersen, P. L., & Tvedebrink, T. (2023). Integrating computational thinking to enhance students' mathematical understanding. *Journal of Pedagogical Research*, *7*(2), 127–142. https://doi.org/10.33902/JPR.202318531

Kaya, D., Yaşar, A. Ö., Çetin, İ., & Kutluca, T. (2025). The relationship between the 21st-century skills and computational thinking skills of prospective mathematics and science teachers. *Journal of Pedagogical Research*, *9*(1), 73–95. https://doi.org/10.33902/JPR.202531498

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage Publications.

Kynigos, C. (2007). Half-baked Logo microworlds as boundary objects in integrated design. *Informatics in Education*, *6*(2), 335–358. https://doi.org/10.15388/infedu.2007.22

Molina-Ayuso, Á., Adamuz-Povedano, N., Bracho-López, R., & Torralbo-Rodríguez, M. (2022). Introduction to computational thinking with Scratch for teacher training for Spanish primary school teachers in mathematics. *Education Sciences*, *12*(12), Article 899. https://doi.org/10.3390/educsci12120899

Molina-Ayuso, Á., Adamuz-Povedano, N., Bracho-López, R., & Torralbo-Rodríguez, M. (2024). Computational thinking with Scratch: A tool to work on geometry in the fifth grade of primary education. *Sustainability*, *16*(1), Article 110. https://doi.org/10.3390/su16010110

Moon, P. F., Himmelsbach, J., Weintrop, D., & Walkoe, J. (2023). Developing preservice teachers' intuitions about computational thinking in a mathematics and science methods course. *Journal of Pedagogical Research*, *7*(2), 5–20. https://doi.org/10.33902/JPR.202318599

Mumcu, F., Kıdıman, E., & Özdinç, F. (2023). Integrating computational thinking into mathematics education through an unplugged computer science activity. *Journal of Pedagogical Research*, *7*(2), 72–92. https://doi.org/10.33902/JPR.202318528

Nordby, S. K., Mifsud, L., & Bjerke, A. H. (2024). Computational thinking in primary mathematics classroom activities. *Frontiers in Education*, *9*, Article 1414081. https://doi.org/10.3389/feduc.2024.1414081

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

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

Park, W., & Kwon, H. (2022). Research trends and issues including computational thinking in science education and mathematics education in the Republic of Korea. *Journal of Baltic Science Education*, *21*(5), 875–887. https://doi.org/10.33225/jbse/22.21.875

Sala-Sebastià, G., Breda, A., Seckel, M. J., Farsani, D., & Alsina, À. (2023). Didactic–mathematical–computational knowledge of future teachers when solving and designing robotics problems. *Axioms*, *12*(2), Article 119. https://doi.org/10.3390/axioms12020119

Sarmasági, P., Rumbus, A., Bilbao, J., Margitay-Becht, A., Pluhár, Z., Rebollar, C., & Dagienė, V. (2025). Bridging algebraic and computational thinking: Impacts on student development in K–12 education. *Informatics in Education*, *24*(2), 343–376. https://doi.org/10.15388/infedu.2025.13

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, *22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Suparman, S., Juandi, D., Turmudi, T., Martadiputra, B. A. P., Helsa, Y., Masniladevi, M., & Suherman, D. S. (2025). Computational thinking in mathematics instruction integrated to STEAM education: A systematic review and meta-analysis. *TEM Journal*, *14*(1), 949–963. https://doi.org/10.18421/TEM141-84

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

Ye, H., Liang, B., Ng, O. L., & Chai, C. S. (2023). Integration of computational thinking in K-12 mathematics education: A systematic review on CT-based mathematics instruction and student learning. *International Journal of STEM Education*, *10*(1). https://doi.org/10.1186/s40594-023-00396-w

Yunianto, W., Jarvis, D., Lavicza, Z., Putra, Z. H., & El-Bedewy, S. (2025). Primary school students' problem-solving strategies in creating artworks with GeoGebra: Integrating computational thinking skills into mathematics and visual arts lessons. *LUMAT*, *13*(2). https://doi.org/10.31129/LUMAT.13.2.2547

Yunianto, W., Prodromou, T., Lavicza, Z., & Sabitzer, B. (2025). Examining students' computational thinking skills in enhanced-technology mathematics lessons: The effects of gender and grade levels. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00734-w

AUTHOR CONTRIBUTION STATEMENT

S: Conceptualization, methodology, formal analysis, writing original draft preparation, writing review and editing. KY: Conceptualization, supervision, writing review and editing, validation. JAD: Supervision, writing review and editing. TP: Methodology, writing review and editing, validation. All authors have read and agreed to the published version of the manuscript.

AI DISCLOSURE STATEMENT

The authors used AI-assisted tools (Claude by Anthropic) during the preparation of this work for language refinement, literature synthesis structuring, and draft writing assistance, and again during the revision of the manuscript in response to reviewers' comments (restructuring of sections and tables, revision of text, and checking of selected source

abstracts). After using the tool, the authors reviewed and edited the content as needed and take full responsibility for the content of this publication.

Sarwan
Department of Mathematics Education Faculty of Mathematics and Science Education,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi, Bandung, Jawa Barat, 40154, Indonesia
Email: sarwan@upi.edu

***Kartika Yulianti (Corresponding Author)**
Department of Mathematics Education Faculty of Mathematics and Science Education,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi, Bandung, Jawa Barat, 40154, Indonesia
Email: kartikayulianti@upi.edu

Jarnawi Afghani Dahlan
Department of Mathematics Education Faculty of Mathematics and Science Education,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi, Bandung, Jawa Barat, 40154, Indonesia
Email: jarnawi@upi.edu

Tia Purniati
Department of Mathematics Education Faculty of Mathematics and Science Education,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi, Bandung, Jawa Barat, 40154, Indonesia
Email: tpurniati@upi.edu