Co-funded by
the European Union

Review of existing CT and AT assessment instruments

The results are based on the work within the project "Computational Thinking and Mathematical Problem Solving, an Analytics Based Learning Environment" (CT&MathABLE). Coordination: Prof. Valentina Dagienė, Vilnius University (Lithuania). Partners: Ankara University (Türkiye), Eötvös Loránd University (Hungary), Gedminų Progymnasium (Lithuania), KTH Royal Institute of Technology (Sweden), Özkent Akbilek Middle School (Türkiye), University of Basque Country (Spain), University of Turku (Finland),. The project has received co-funding by the Erasmus+ Programme KA220.

These results are developed by Daranee Lehtonen and Kim Erola under WP3.

Reviewers: Javier Bilbao, University of Basque Country, Spain, Vaida-Masiulionytė Dagienė, Vilnius University, Lithuania.

General overview

A systematic literature review of existing computational thinking (CT) and algebraic thinking (AT) assessment instruments was conducted at the beginning of the project. The review aimed to define CT and AT learning objectives and contents for the development of COMATH, CT and AT assessment instruments.

Target groups

This document is intended to provide an overview of existing CT and AT assessment instruments to help (1) teachers to select suitable instruments for assessing students' CT and AT skills, (2) researchers to select suitable instruments for their studies, and (3) test developers to develop instruments.

Keywords
algebraic thinking; computational thinking; assessment instrument; K–9 education; systematic review

Table of Contents

1. A systematic literature review of existing CT assessment instruments

This systematic literature review (Erola & Mirel, 2023) aimed to examine the psychometric properties of computational thinking measures in primary education. The review followed Fink's model (2005), and Fink's model consists of seven stages. The first three stages involve formulating the research questions, selecting appropriate databases and search terms, and consulting experts to assess the databases and search terms. The following research questions guided the review:

1. What instruments were used to assess CT in primary school?
2. What were the psychometric properties of the instruments were reported?

1.1 Materials and methods

An electronic search of peer-reviewed articles on primary education computational thinking (CT) assessments published during the last decade in EBSCOHost (Teacher Reference Center, Academic Search Premier, eBook Collection, ERIC, and Education Source) were conducted. The search terms were, “computational thinking” AND (assessment OR measure OR evaluation) AND (“primary school” OR “elementary education” OR “primary education” OR “elementary school” OR “basic education”). In total, 185 articles were identified. After duplicate removal, 166 article remained.

The fourth and fifth stages involve screening and selecting articles based on predefined criteria (see Table 1). This includes narrowing down the articles based on language, time, and content relevance. The goal is to obtain a high-quality material for the review. In this phase, 125 articles were excluded, and 22 articles remained based on their title and abstracts. Both researchers independently read the selected articles and assessed them by answering three questions using responses of NO, MAYBE, or YES. At the end of the full-text screening, nine articles were excluded, and nine articles (see Table 2) with 11 CT assessment instruments were included in the review. The questions used for the full-text screening were as follows:

(a) Does the article address CT?
This question aimed to determine whether the research addressed CT in general, or, for example, programming, robotics, or problem-solving. The key was to identify articles in which CT was the focal point.

(b) Does the article include a measure that specifically assesses CT?
The criterion aimed to identify studies in which CT was measured, and which included a clear measure for assessing CT. At this stage, the names of the CT measures were also recorded for further use.

(c) Does the article discuss the psychometric properties of the assessment tool?
This question aimed to ensure that the included studies in the dataset evaluated the psychometric properties of the assessment tool used. Psychometric properties are used to check the quality of the measure and its effectiveness in fulfilling its intended purpose.

Table 1. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
• Peer-reviewed articles	• Not peer-reviewed article
• Published between 2013 and 2022	• Published before 2013
• Available in English or Finnish	• Not Available in English or Finnish
• About CT	• About robotics, programming, or problem-solving, but not CT
• Including a measure that assesses CT	• Not including a measure that assesses CT
• Reporting psychometric properties of the measure	• Not reporting psychometric properties of the measure

Table 2. Nine articles included in Erola and Mirel's (2023) systematic literature review

Authors	Publication Year	Article Title
Basu, Rutstein, Xu, Wang & Shear	2021	A Principled approach to designing computational thinking concepts and practices assessments for upper elementary grades
Chen, Shen, Barth-Cohen, Jiang, Huang & Eltoukhy	2017	Assessing elementary students' computational thinking in everyday reasoning and robotics programming
Gane, Israel, Elagha, Yan, Luo & Pellegrino	2021	Design and validation of learning trajectory-based assessments for computational thinking in upper elementary grades
Kong & Wang	2021	Item response analysis of computational thinking practices: Test characteristics and students' learning abilities in visual programming contexts
Li, Xu & Liu	2021	Development and validation of computational thinking assessment of Chinese elementary school students
Relkin, de Ruiter & Bers	2020	"TechCheck": Development and validation of an unplugged assessment of computational thinking in early childhood education
Tsarava, Moeller, Román-González, Golle, Leifheit, Butz & Ninaus	2022	A cognitive definition of computational thinking in primary education.
Zapata-Cáceres, Martín-Barroso & Román-González	2021	Collaborative game-based environment and assessment tool for learning computational thinking in primary school: A Case study

Zhong, Wang, Chen & Li	2016	An Exploration of three-dimensional integrated assessment for computational thinking

The last two stages involve conducting the review and synthesis of the literature: testing the process, ensuring the review's reliability, and analysing the gathered information. In this study the articles were analysed at the content level. Initially, the target age groups/grade levels, tasks, task types used in the measures, the number of tasks, and formats of the assessment implementation (e.g., computer or pen and paper), were examined. After a general overview, the focus shifted to the psychometric properties reported for the measures, namely reliability and validity.

The validity of the measures was analysed from the perspectives of content validity and construct validity based on the Cosmin taxonomy (Prinsen et al., 2018). Aspects such as test piloting and expert panel discussions were categorized under content validity, as they are used to strengthen the measure's ability to adequately reflect the construct being measured, thereby improving its content validity. Various factor analyses and their results, as well as correlation analyses supporting concurrent or criterion validity, were classified under construct validity.

The main objective of the data analysis was to examine the reported measurements and results of the psychometric properties of the measures according to the reliability and validity components of the Cosmin taxonomy. The reliability of the measures was assessed through internal consistency and reliability according to the Cosmin taxonomy. Internal consistency was examined and reported using measures such as Cronbach's alpha coefficients and other similar indicators of the consistency between different sections of the measure. The analysis of reliability aimed to identify the proportion of measurement errors in the total variance.

Beside nine articles from Erola and Mirel's, (2023) systematic literature review, 26 publications (journal articles or conference proceedings) on CT assessment from a recent systematic review of Kampylis et al. (2023, see Table 3) were also included in our content analysis. Altogether 35 full-text publications were analysed in terms of CT definition and categorization as well as assessment instruments including instrument names, target groups, test types and items used for assessing students CT and how those are rated as well as psychometric properties (reliability and validity) of the instruments.

Table 3. 26 publications included in Kampylis et al.'s (2023) systematic literature review

Authors	Publication Year	Publication Title
Asbell-Clarke, Rowe, Almeda, Edwards, Bardar, Gasca, Baker & Scruggs	2021	The development of students' computational thinking practices in elementary- and middle-school classes using the learning game, Zoombinis
Caeli & Yadav	2020	Unplugged Approaches to Computational Thinking: A Historical Perspective
Cutumisu, Adams & Lu	2019	A scoping review of empirical research on recent computational thinking assessments
Dagli & Sancar Tokmak	2021	Exploring high school computer science course teachers' instructional design processes for improving students' "computational thinking" skills
de Araujo, Andrade & Guerrero	2016	A systematic mapping study on assessing computational thinking abilities
Djambong, Freiman, Gauvin, Paquet, & Chiasson	2018	Measurement of Computational Thinking in K-12 Education: The Need for Innovative Practices
Fagerlund, Häkkinen, Vesisenaho & Viiri	2020	Assessing 4th Grade Students' Computational Thinking through Scratch Programming Projects
Grgurina, Barendsen, Suhre, Zwaneveld & Van Veen	2018	Assessment of modeling and simulation in secondary computing science education
Guggemos	2021	On the predictors of computational thinking and its growth at the high-school level
Guggemos, Seufert & Román-González	2022	Computational Thinking Assessment–Towards More Vivid Interpretations
Hazzan, Ragonis & Lapidot	2020	Computational Thinking
Hooshyar, Pedaste, Yang, Malva, Hwang, Wang, Lim & Delev	2020	From Gaming to Computational Thinking: An Adaptive Educational Computer Game-Based Learning Approach
Lu, Macdonald, Odell, Kokhan, Demmans Epp, & Cutumisu	2022	A scoping review of computational thinking assessments in higher education

Authors	Publication Year	Publication Title

Niemela, Partanen, Harsu, Leppänen & Ihantola	2017	Computational thinking as an emergent learning trajectory of mathematics
Pasternak	2016	Contextualized teaching in the lower secondary education long-Term evaluation of a cs course from grade 6 to 10
Palts & Pedaste	2020	A model for developing computational thinking skills
Román-González, Pérez-González & Jiménez-Fernández	2017	Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test
Román-González, Moreno-León & Robles	2019	Combining Assessment Tools for a Comprehensive Evaluation of Computational Thinking Interventions
Sun, Hu, & Zhou	2021	Which way of design programming activities is more effective to promote K-12 students' computational thinking skills? A meta-analysis
Tang, Yin, Lin, Hadad & Zhai	2020	Assessing computational thinking: A systematic review of empirical studies
Taslibeyaz, Kursun & Karaman	2020	How to Develop Computational Thinking: A Systematic Review of Empirical Studies. Informatics in Education
Tikva & Tambouris	2021	Mapping computational thinking through programming in K-12 education: A conceptual model based on a systematic literature Review
Tsai, Liang & Hsu	2020	The Computational Thinking Scale for Computer Literacy Education
Wiebe, London, Aksit, Mott, Boyer & Lester	2019	Development of a lean computational thinking abilities assessment for middle grades students
Yağcı	2019	A valid and reliable tool for examining computational thinking skills
Zapata-Cáceres, Martín-Barroso & Román-González	2020	Computational thinking test for beginners: Design and content validation

1.2 Results

Based on the content analysis, we found that:

- About 77% of the 35 publications (see Table 2) were published after 2019.
- All of the CT assessment instruments on systematic review focused on elementary school students (5-12 years old).
- 6 of 11 assessment instruments contained multiple choice questions, whereas the rest used open-ended questions, different kinds of problem-solving, and coding questions.
- Computers were mostly used to execute the tests.
- Validity of instruments on publications was moderate, content validity was mostly secured through expert panels, pilots, and different assessment frameworks.
- Construct validity was a weaker part. Although 8 of 11 instruments ran factor analyses to support that, hypothesis testing was conducted only in 4 instruments.
- In terms of internal consistency, 8 of 11 instruments reported Cronbach alphas and they were considerably strong, apart from one instrument with Cronbach alpha of 0.48.
- Lastly, reliability was the weakest part of all of the instruments, only two instruments in one publication reported Cohen's Kappa.

Regarding the CT definition and categorisation, we concluded that CT's definition is a broad and complex concept, which contains multiple sub-concepts that affect one another. Among the most recent literature (Shin et al., 2022; Ezeamuzie & Leung, 2022), we chose Shute, et al.'s (2017) definition, which portrays CT as a broader cognitive construct and not just a practical skill that is relevant for specific computing-related contexts (Armoni, 2016). Therefore, their definition is appropriate for our purpose, annotating tasks for the development of the COMATH assessment instrument. According to Shute et al. (2017, pp.153), CT consists of six key components:

1. **Decomposition -** Breaking down a complex problem or system into smaller, manageable parts. These parts are functional elements that work together to form the whole.
2. **Abstraction** – Identifying the core aspects of a system. This includes:
    - **Data Collection & Analysis** – Gathering relevant data from various sources and understanding their relationships.
    - **Pattern Recognition** – Detecting patterns or underlying rules within the data.
    - **Modelling** – Creating simulations or models to represent a system's behavior or predict future outcomes.
3. **Algorithms** – Developing structured and logical steps to solve problems, which can be executed by humans or computers. This involves:
    - **Algorithm Design** – Creating step-by-step solutions.
    - **Parallelism** – Performing multiple steps simultaneously.
    - **Efficiency** – Optimizing the process by eliminating unnecessary steps.
    - **Automation** – Enabling solutions to run automatically for repeated tasks.
4. **Debugging** – Identifying and fixing errors when a solution does not function correctly.
5. **Iteration** – Refining solutions through repeated testing and improvement until the desired outcome is achieved.
6. **Generalisation** – Applying CT skills across different situations and domains to solve various problems efficiently.

2. A systematic literature review of existing AT assessment instruments

We conducted a systematic literature review (Lehtonen et al., 2025) to systematically review empirical studies assessing K–9 students' algebraic thinking (AT) between 1950–2023 and provide the state of the art in AT assessment instruments for instrument selection and development. The following research questions (RQs) guide the review:

RQ1. 1.1 How is AT defined?
1.2 What AT competencies were assessed?
1.3 Is AT a unidimensional or multidimensional construct?
RQ2. What are the study contexts in which AT competencies were assessed?
RQ3. 3.1 How was the assessment implemented?
3.2 What are the characteristics of the test instruments used?

2.1 Materials and methods

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). The review process comprised three major steps: identification, screening, and data coding.

First, we conducted an electronic search in three major databases in educational and social sciences: EBSCOhost (ERIC, APA PsycArticles, APA PsycInfo, Teacher Reference Center, and Education Source), ProQuest (Education, Psychology, and Science Database), and Web of Science (Social Sciences Citation Index and Social Citation Index Expanded). A Boolean operation of the search terms including "AT", "assessment instruments", "K–9 education", and their synonyms were used for a title, abstract, and keyword search. In total, 743 articles were identified. After duplicate removal, 587 article remained. Second, we undertook 2-step screening process (titles and abstracts, and full texts) using predefined criteria (see Table 4). A total of 344 and 141 articles were excluded during the title and abstract screening and full-text screening, respectively. As a result, 102 articles were included in the full-text review. Then, we read and coded the reviewed articles to answer the research questions (see Table 5).

Table 4. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
• Peer-reviewed articles • Published from January 1950 to February 2023 • Available in English • Assessing AT of K–9 students with heterogeneous attainments • Using a test as an assessment instrument • Sufficient indicators supporting quality of the studies and the assessment instruments used	• Not peer-reviewed article • Published before 1950 and after February 2023 • Not Available in English • Not about K–9 students' AT assessment • No test use as an assessment instrument • Insufficient indicators supporting quality of the studies and the assessment instruments used, for example, less than 30 participants

Table 5. Coding scheme

Code	Subcode	Coding type
AT definitions		open
AT competencies measured		open
AT dimensionality	Unidimensional construct (The AT assessment was analysed and reported as a whole; for example, the reliability coefficient indicators were reported for the whole test.)	pre-defined
	Multidimensional construct (The assessment of each AT competency was analysed and reported separately; for example, separate reliability coefficient indicators were provided for the assessment of each competency.)	pre-defined
Study contexts	Publication year	pre-defined
	Research location	open
	Researchers	
	Number	pre-defined
	Name	open
	Participants	
	Grade level	pre-defined
	Sampling	pre-defined
	Sample size	pre-defined
Assessment implementation	Format	open
	Duration	pre-defined
Test instrument characteristics	Instrument development	pre-defined
	Self-developed/customised	
	Existing	
	Test items	
	Number	pre-defined
	Type	pre-defined

2.2 Results

We investigated the AT definition, what skills construct it, the study demographics (publication years, study location, and participants), assessment implementation (implementation format and test duration), and the instruments used (instrument development and test items).

Regarding the definition of AT, 10 studies presented a definition of AT proposed by different researchers. Most of the cited researchers, including Al-Shehri (2020), Kieran (2011), and Lins (1992), agree that in contrast to arithmetic thinking, which focuses on the answer to a mathematical problem, AT is a mental process of making meaning of algebra. Therefore, AT involves (1) identifying and generalising mathematical structures and relationships (e.g., Kaput et al., 2008; Radford, 2014), (2) representing generalisation with alphanumeric and other symbols, such as drawings and graphs (e.g., Kaput et al., 2008; Kieren, 2011), and (3) reasoning and modelling with the symbolised generalisation (e.g., Kaput et al., 2008; Kieren, 2004).

Although according to the AT definition, AT should be considered a multicomponential construct, this idea is invisible in most assessment instruments used in the articles. Equations and inequations competency (n = 80) was the most assessed among the reviewed articles, followed by functional thinking (n = 33) and representation (n = 41). Based on the contents of the assessment instruments used in the reviewed articles, we propose that AT comprises multiple subcompetencies that can be grouped into seven main skills:

1. **Generalised arithmetic**: The skill of generalising arithmetic relationships (fundamental properties of number and operation, e.g., the Commutative Property of Addition) and reasoning about the structure of arithmetic expressions rather than their computational value (Blanton et al., 2015)
2. **Equations and inequation**: A relational understanding of the equal sign and the skill of representing and reasoning with expressions and equations in their symbolic form, and describing relationships between and among generalized quantities that may or may not be equivalent (Blanton et al., 2015)
3. **Functional thinking**: The skill of generalising (numerical) patterns to describe functional relationships/relationships between co-varying quantities: their similarities and differences, causality, growth and continuous joint variation (Kaput, 1998)
4. **Variables**: The ability to work with variables (i.e., symbols, usually letters representing generalised or unknown values of mathematical relationships) as if they were numbers. This competency is necessary for other AT competencies, such as solving equations.
5. **Representation**: The skill of using multiple representations, e.g., drawings, tables, letters, and symbols that serve to facilitate gaining information and a deep understanding of the content of the material (Dindyal, 2003)
6. **Transformation**: The skill of changing the form of an expression/equation to maintain equivalence, for instance, collecting like terms, factoring, expanding, substituting, adding and multiplying polynomial expressions, exponentiation with polynomials, and simplifying expressions (Kieran, 1996)
7. **Transversal skills**: Including reasoning, generalising, justifying/proving, modelling/predicting, validating, and solving problems

Although algebra teaching and learning have long been studied, an interest in AT assessment instruments is a relatively recent phenomenon. Our search extended back to 1950, but we found no studies published between 1950–1992; nearly half (n = 46) of the studies published in 2019 or later.

Almost all of studies (n = 97) were conducted in single countries. The assessments were implemented across continents, particularly in America (n = 61), followed by Asia (n = 32) and Europe (n = 14). Of the 102 studies conducted in 25 different countries, more than half were in the US (n = 59), followed by Turkey (n = 8), Singapore (n = 4), and Finland (n = 4).

Altogether 60 studies were conducted at a single grade level, while 42 were conducted at multiple grade levels. More than half of the studies (n = 57) involved lower secondary school students (Grades 7–9), followed by Grades 5–6 (n = 44) and Grades 3–4 (n = 22). Only 14 studies involved K–2 students.

Only 45 studies reported sampling methods. Most of them (n = 35) employed non-probability sampling, particularly convenience sampling, purposive sampling, and voluntary response sampling. The remaining 10 studies employed probability sampling. The sample size of students in the reviewed studies varied from 48 to nearly 25,500 students. A total of 71 studies reported statistical results of specific subsamples, particularly treatment-control conditions and/or other demographics (e.g., grade levels, sex, mathematics achievement, and study location). Most of the studies were small-scale (30–60 students, n = 26; 61–100 students, n = 20; 101–200 students, n = 19), and the smallest subsample was 30.

Less than half of the studies (n = 41) reported how the assessment was implemented. Most of these studies (n = 33) used paper-and-pencil tests. A small number used other implementation formats: digital tests (n = 8), oral and paper-and-pencil tests (n = 5), and paper-and-pencil tests, in which items were physically or digitally presented (n = 5).

Less than half of the studies (n = 47) reported the test duration. If a test comprised multiple sections, we combined the durations of all sections for the analysis. The test duration varied across the studies: 1–10 minutes (n = 6), 11–20 minutes (n = 1), 21–30 minutes (n = 8), 31–40 minutes (n = 6), 41–50 minutes (n = 12), 51–60 minutes (n = 6), and more than 60 minutes (n = 4). Additionally, four studies implemented the assessment without a time limit.

In total, 135 different tests were used in the 102 reviewed studies. Of these 135 tests, 91 were originally developed by the researcher(s) or adapted from existing tests. The remaining 44 tests were constructed in different ways: the researcher(s) took tests/test items from their own previous studies or others' studies or used unpublished tests, standardised tests/test items (e.g., Chelsea Diagnostic Algebra Test, Hart et al., 1985), or national/international assessments (e.g., TIMSS 2007 Mathematics Assessment, International Association for the Evaluation of Educational Achievement, 2007).

Most of the studies (n = 98) reported the number of items used. If a test was composed of multiple sections, we summed the number of items in all sections for the analysis. The number of items in one assessment ranged from 4 to 93. In most of the studies (n = 73), the tests contained no more than 30 items. A total of 84 studies reported the item types used: open-response (n = 72), multiple-choice (n = 50), and dual-choice (true/false or yes/no, n = 13). Of the 84 studies, 39 used one item type, while the remaining combined two (n = 39) or three types (n = 6).

As a result of the CT and AT assessment instrument review, 35 publications mentioning or using CT assessment instruments and 102 empirical studies using AT assessment instruments have been found. The systematic review results will be used for developing COMATH assessment instruments.

References

Al-Shehri, M. A. M. (2020). The effectiveness of the collaborative investigation strategy in the achievement and development of algebraic thinking skills for first intermediate grade students. *Journal of the College of Basic Education for Educational and Human Sciences, 46*(1), 259–272.

Armoni, M. (2016). Computer science, computational thinking, programming, coding: the anomalies of transitivity in K-12 computer science education. In *ACM Inroads*, 7(4), 24–27. https://doi.org/10.1145/3011071

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J.-S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education*, *46*(1), 39–87. https://doi.org/10.5951/jresematheduc.46.1.0039

Dindyal, J. (2003). *Algebraic thinking in geometry at high school level: Students' use of variables and unknowns* [Unpublished doctoral dissertation]. Illinois State University.

Erola, K. & Mirel, J. (2023). Ohjelmoinnillisen ajattelun mittarit: Systemaattinen kirjallisuuskatsaus [Computational thinking assessment instruments: A systematic literature review, Master's thesis, University of Turku]. UTUPUB. https://www.utupub.fi/handle/10024/175021

Ezeamuzie, N. O., & Leung, J. S. (2022). Computational thinking through an empirical lens: A systematic review of literature. *Journal of Educational Computing Research*, *60*(2), 481–511.

Fink, A. (2019). *Conducting research literature reviews: From the internet to paper*. Sage publications.

Hart, K., Brown, M., Kerslake, D., Küchemann, D., & Ruddock, G. (1985). *Chelsea diagnostic mathematics tests* [Student assessment]. Centre for Mathematical Education, University of Nottingham.

International Association for the Evaluation of Educational Achievement. (2007). *TIMSS 2007 Mathematics Assessment.*

Kampylis, P., Dagienė, V., Bocconi, S., Chioccariello, A., Engelhardt, K., Stupurienė, G., Masiulionytė-Dagienė, V., Jasutė, E., Malagoli, C., Horvath, M., & Earp, J. (2023). Integrating Computational Thinking into Primary and Lower Secondary Education: A Systematic Review. *Educational Technology & Society*, *26*(2), 99–117. https://doi.org/10.30191/ETS.202304_26(2).0008

Kaput, J. (1998). Transforming algebra from an engine of inequity to an engine of mathematical power by "algebrafying" the K–12 curriculum. In National Council of Teachers of Mathematics, Mathematical Sciences Education Board, & National Research Council (Ed.), *The nature and role of algebra in the K–14 curriculum: Proceedings of a National Symposium* (pp. 25–26). National Academies Press.

Kaput, J., Blanton, M., & Moreno, L. (2008). Algebra from a symbolization point of view. In J. Kaput, D. Carraher, & M. Blanton (Eds.), *Algebra in the early grades* (pp. 19–55). Lawrence Erlbaum Associates.

Kieran, C. (1996). The changing face of school algebra. In C. Alsina, J. Alvarez, B. Hodgson, C. Laborde, & A. Pérez (Eds.), *The 8th International Congress on Mathematical Education: Selected lectures* (pp. 271–290). Thales.

Kieran, C. (2004). Algebraic thinking in the early grades: What is it? *The Mathematics Educator, 8*(1), 139–151.

Kieran, C. (2011). Overall commentary on early algebraization: Perspectives for research and teaching. In J. Cai & E. Knuth (Eds.), *Early algebraization. A global dialogue from multiple perspectives* (pp. 579–593). Springer-Verlag.

Lehtonen, D., Laakso, M.-J., & Räsänen, P. (12025). A systematic review of assessment instruments for algebraic thinking in K–9 over the past 70 years. Manuscript submitted for publication.

Lins, R. C. (1992). *A framework for understanding what algebraic thinking is* [Doctoral dissertation, University of Nottingham].

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., Hrobjartsson, 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 (Online)*, *372*, n71–n71. https://doi.org/10.1136/bmj.n71

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research, 27*, 1147–1157.

Radford, L. (2014). The progressive development of early embodied algebraic thinking. *Mathematics Education Research Journal, 26*(2), 257–277. https://doi.org/10.1007/s13394-013-0087-2

Shin, N., Bowers, J., Roderick, S., McIntyre, C., Stephens, A. L., Eidin, E., ... & Damelin, D. (2022). A framework for supporting systems thinking and computational thinking through constructing models. *Instructional Science, 50*(6), 933–960.

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142–158.