Studies in Bibliometrics of Education and Technology
Homepage: https://scie-journal.net/index.php/SiBET
Email:

p-ISSN:XXXXX; e-ISSN: XXXXX
SiBET, Vol. 1, No. 3, July 2026
Page 27-43
©2026 SiBET

Research Trends in Robotics-Based Learning and Students' Conceptual Understanding in STEM Education: A Bibliometric and Systematic Review

***Riski Ramadani[1], Afiyah Nikmah[1], Arum Vonie Rachmawati[1], Rohim Aminullah Firdaus[1], Noer Risky Ramadhani[2]**

[1]Postgraduate Physics, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya, Indonesia
[2]Department for Higher Education Research, Faculty of Education, and Architecture, University for Continuing Education Krems, Austria

Article Info

Article history:

Received: 30 June 2026
Revised: 15 July 2026
Accepted: 29 July 2026
Available Online: 31 July 2026

Keywords:

Bibliometric Analysis
Conceptual Understanding
Educational Robotics
Robotics-Based Learning
STEM Education

ABSTRACT

Educational robotics is increasingly used to support active and integrated STEM learning, yet the extent to which it develops students' conceptual understanding remains unclear. This study maps the development of robotics-based STEM learning research and examines how conceptual understanding has been assessed and reported. An integrated bibliometric and systematic-review design was applied to Scopus documents published from 2001 to 2025. Of 59 retrieved records, eight false positives were excluded, resulting in a bibliometric corpus of 51 documents and an abstract-stage empirical evidence subset of 34 studies. Performance indicators, contributors, geographic collaboration, normalized author-keyword co-occurrence, educational characteristics, conceptual-evidence strength, and direction of findings were analyzed. The corpus received 693 citations and achieved an h-index of 13; 56.9% of the publications appeared during 2023–2025. The United States, Taiwan, and China led publication output, while international collaboration remained limited. Four keyword clusters represented the core educational-robotics field, digital and computational tools, constructionist pedagogy, and an emerging artificial-intelligence and equity frontier. Within the empirical subset, computing or robotics concepts accounted for 50.0% of studies, and the robot most often functioned as a construction or programming object. Only ten studies provided direct conceptual evidence, compared with 19 proximal and five indirect studies. Although 24 abstracts reported positive findings, these direction counts are not equivalent to pooled causal effects. The field is expanding rapidly, but its conceptual evidence remains methodologically uneven and concentrated in broad knowledge-performance outcomes. Full-text screening, quality appraisal, and construct-specific effect synthesis are therefore required.

Check for updates OPEN ACCESS

INTRODUCTION

STEM education is intended to help learners connect scientific ideas, mathematical representations, technological tools, and engineering design when addressing authentic problems. Its educational value therefore depends on more than participation in an activity: students must construct coherent explanations, relate multiple representations, predict the behavior of a system, and transfer principles to unfamiliar situations. An integrated STEM framework likewise requires purposeful connections among disciplinary knowledge, learning contexts, engineering design, and twenty-first-century competencies (Kelley & Knowles, 2016). For this reason, students' conceptual understanding is a central outcome for evaluating whether an innovative STEM activity produces meaningful learning rather than short-term engagement alone.

Robotics-based learning offers a particularly visible and manipulable context for such integration. A robot combines sensing, actuation, control, programming, mechanics, electronics, energy, measurement, and mathematical reasoning in a single system. Learners can design a mechanism, translate an idea into code, observe the robot's response, compare that response with a prediction, diagnose failure, and revise

the design. Reviews of educational robotics platforms show that robots are used not only to teach robotics and programming but also to support problem solving, collaboration, creativity, and disciplinary learning across STEM contexts (Anwar et al., 2019; Evripidou et al., 2020). Consequently, the robot can function as a learning object, a cognitive tool, or an experimental model, depending on the intended curriculum and pedagogy.

The pedagogical promise of robotics is commonly associated with constructionist, inquiry-based, design-based, and project-based learning. In these approaches, conceptual knowledge is externalized through artifacts and tested through the build–program–observe–debug cycle. Studies with young learners indicate that developmentally appropriate robotics curricula can support mastery of programming and computational concepts, although mastery varies with the difficulty of the concept and the design of the curriculum (Bers et al., 2014; Sullivan & Bers, 2016). Other empirical work demonstrates that robotics tasks can reveal computational thinking in both everyday reasoning and programming performance, which underscores the importance of aligning the learning objective with an appropriate assessment (Atmatzidou & Demetriadis, 2016; G. Chen et al., 2017).

However, positive engagement with a robot does not automatically indicate conceptual change. A learner may successfully copy a program, assemble a model, or complete a competition challenge while retaining fragmented explanations of force, motion, energy, circuits, proportionality, feedback, or algorithms. Earlier systematic reviews found broad educational potential but also noted short interventions, small samples, frequent use of LEGO platforms, non-experimental designs, and varied instruments such as observation, questionnaires, interviews, and artifact assessment (Benitti, 2012; Xia & Zhong, 2018). A broad review of 147 K–12 studies further organized reported benefits into general effectiveness, learning and transfer, creativity and motivation, diversity and participation, and teacher professional development, illustrating that the outcome literature is extensive but conceptually heterogeneous (Anwar et al., 2019).

Table 1. Analytical Boundary of Robotics-Based Learning and Conceptual Understanding

Construct	Working Definition	Observable Evidence	Review Coding Rule
Robotics-based learning	Learning in which learners build, program, control, investigate, or meaningfully interact with a robot.	Design logs; code; robot behavior; experiments; troubleshooting; explanations.	Include empirical learning activities; exclude industrial robotics papers without an educational process.
Learning about robotics	Robotics itself is the curriculum content, such as sensors, actuators, control, mechanics, or programming.	Mastery of robotics concepts and system behavior.	Code as robotics-content learning; report separately from learning with robotics.
Learning with robotics	A robot is a medium for learning another STEM concept or an integrated STEM problem.	Understanding of physics, mathematics, engineering, computing, or interdisciplinary concepts.	Code the target discipline, concept, robot role, and degree of STEM integration.
Conceptual understanding	Ability to explain, connect representations, predict, transfer, and revise ideas—not merely recall facts.	Concept inventories; diagnostic items; concept maps; interviews; constructed responses; transfer tasks.	Prioritize direct evidence of concepts; distinguish it from total achievement and self-reported perception.
Supporting outcomes	Cognitive, affective, and social outcomes that may enable or accompany conceptual learning.	Computational thinking; problem solving; motivation; collaboration; self-efficacy.	Map as secondary outcomes and examine whether they are linked to measured conceptual gains.

Recent synthesis studies reinforce both the value of the field and the need for greater analytical precision. A systematic review of 39 robotics-based STEM studies published from 2012 to 2021 found that technology and programming were dominant, LEGO Mindstorms was the most frequently used platform, and conceptual knowledge appeared as one among several educational targets (Darmawansah et al., 2023). A multilevel meta-analysis of 21 studies published from 2010 to 2022 reported an overall positive role for educational robotics, while showing that effects vary across disciplines and implementation conditions (Ouyang & Xu, 2024). Thus, aggregate achievement or effect-size estimates alone cannot show which STEM concepts are being understood, how that understanding is measured, or which combinations of robot,

pedagogy, duration, and learner level support conceptual change. At the same time, bibliometric studies have begun to describe the intellectual and social structure of educational robotics research. A mapping of 1,382 Web of Science records from 1975 to 2021 identified the United States, Computers & Education, Marina Umaschi Bers, and Tufts University as influential contributors, while computational thinking, STEM, coding, programming, social robots, and communication emerged as recent themes (Akgün & Atıcı, 2023). Other scientific mapping has charted the growth of robotics-in-education research in Web of Science (López-Belmonte et al., 2021), whereas a more recent analysis of 50 Scopus-indexed studies examined robotics in science education between 2012 and 2024 (Aleksandrova et al., 2024). These studies establish the expansion and thematic diversity of the field, but their units of analysis are publications, authors, sources, and keywords rather than the quality and form of students' conceptual understanding.

Table 2. Positioning of Prior Reviews and the Proposed Study

Study	Scope and Method	Main Contribution	Unresolved Issue
Benitti (2012)	Systematic review; school studies, mainly 2000–2009	Established the educational potential of robotics for science and mathematics.	Predates the recent expansion of the field; no science mapping; conceptual evidence not isolated.
Anwar et al. (2019)	Systematic review; 147 K–12 studies, 2000–2018	Synthesized effectiveness, transfer, motivation, participation, and teacher development.	Broad outcome categories; no bibliometric map or concept-specific synthesis.
Akgün & Atıcı (2023)	Bibliometric review; 1,382 WoS records, 1975–2021	Mapped productivity, influence, collaboration, and emerging themes.	Did not systematically code concepts, pedagogy, instruments, or learning evidence.
Darmawansah et al. (2023)	Systematic review; 39 R-STEM articles, 2012–2021	Compared environments, learners, robots, disciplines, pedagogy, and objectives.	Conceptual understanding was one outcome among many; no bibliometric network analysis.
Ouyang & Xu (2024)	Multilevel meta-analysis; 21 STEM studies, 2010–2022	Estimated overall effects and tested implementation moderators.	Aggregated learning outcomes; no map of intellectual structure or concept-assessment practices.
Aleksandrova et al. (2024)	Bibliometric review; 50 Scopus studies, 2012–2024	Mapped growth and interdisciplinary links in robotics science education.	Science-specific and limited systematic examination of conceptual learning evidence.
Proposed study	Bibliometric mapping plus systematic review of empirical studies	Connects macro research trends with micro evidence about students' STEM conceptual understanding.	Codes target concept, STEM domain, learner level, robot/platform, pedagogy, duration, design, instrument, and finding.

The unresolved problem is therefore a disconnect between macro-level knowledge of research growth and micro-level evidence of learning. Bibliometric analysis can reveal publication trajectories, influential contributors, collaboration networks, citation structure, and the evolution of keywords. It cannot, by itself, determine whether a study assessed explanation, prediction, transfer, misconception change, or only a general achievement score. Conversely, a systematic review can code the learning design and assessment evidence in detail, but may not show how the conceptual-understanding literature is organized as a scientific field. Combining the two approaches creates a stronger account of both where the field is moving and what its empirical claims about conceptual learning actually mean. Accordingly, this study aims to map the development and intellectual structure of robotics-based learning research related to students' conceptual understanding in STEM education and to systematically synthesize how conceptual understanding has been designed, assessed, and reported. The bibliometric component will examine performance and science-mapping indicators, whereas the systematic component will code the empirical studies at the level of learners, STEM concepts, robotic technologies, pedagogical strategies, research designs, assessment instruments, and reported outcomes. This dual design responds to calls for closer alignment among robot roles, pedagogy, STEM content, and learning outcomes.

Proposed Research Questions

RQ1. How has the volume and citation impact of research on robotics-based learning and students' conceptual understanding in STEM education evolved over time?
RQ2. Which authors, institutions, countries, journals, and documents are the most influential, and how are their collaboration and citation networks structured?
RQ3. What themes and conceptual clusters dominate the field, and how have keywords and research fronts evolved?
RQ4. Which STEM concepts, learner levels, robotic platforms, pedagogical approaches, intervention durations, research designs, and assessment instruments are represented in the empirical literature?
RQ5. What patterns of positive, mixed, or null evidence are reported for students' conceptual understanding, and what methodological gaps should guide future research?

RESEARCH METHOD

Research Design

This study uses an integrated bibliometric-systematic review design. The bibliometric component provides a macro-level account of the field through performance analysis and science mapping, whereas the systematic component examines the educational and methodological characteristics of empirical studies in sufficient depth to evaluate claims about students' conceptual understanding. This combination is appropriate because bibliometric indicators can identify productivity, influence, collaboration, and conceptual structure, but they cannot determine how a learning construct was operationalized or measured (Donthu et al., 2021; Zupic & Čater, 2015). The review is reported in accordance with PRISMA 2020 and the PRISMA-S extension for reproducible reporting of literature searches (Page et al., 2021; Rethlefsen et al., 2021). Two analytically linked corpora are constructed from the same search. Dataset A, the bibliometric corpus, contains all publications that satisfy the topical, bibliographic, language, document-type, and period criteria. Dataset B, the systematic evidence subset, contains only empirical educational studies that directly or proximally assess learners' STEM conceptual knowledge. Reviews may remain in Dataset A because they contribute to citation and intellectual structure, but they are excluded from Dataset B because they do not report primary learner-level evidence. The two datasets are integrated only after each analysis has been completed, thereby preventing citation prominence from being treated as evidence of educational effectiveness.

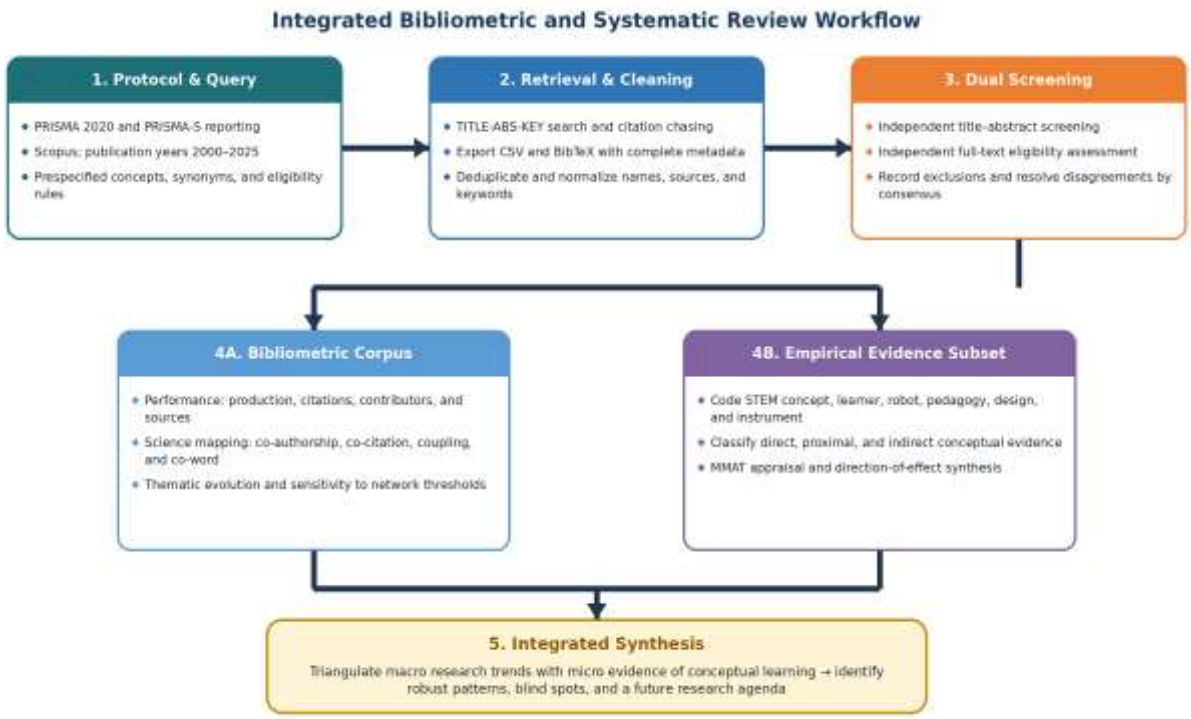


Figure 1. Integrated Workflow for the Bibliometric and Systematic Review Information Source, Coverage Period, and Search Strategy

Scopus is used as the primary database because the study requires a single, internally consistent source of citation, affiliation, author-keyword, and reference metadata for network construction. The coverage period is restricted to complete publication years from 2000 through 2025. The lower boundary captures the expansion of programmable educational robotics, while the upper boundary avoids partial-year bias. The final retrieval date, complete query, applied filters, and number of records returned must be recorded verbatim in the final manuscript and supplementary search log. The search strategy combines three concept blocks: educational robotics, STEM education, and conceptual or content learning. Broad learning terms such as "learning achievement" and "content knowledge" are retained to reduce false negatives because some empirical articles assess conceptual knowledge without using the phrase "conceptual understanding" in the title or abstract. Precision is subsequently restored during full-text screening by applying the operational definition in Table 1. Search reporting follows PRISMA-S, including the database platform, searchable fields, limits, date of search, complete Boolean syntax, and any supplementary citation searching (Rethlefsen et al., 2021).

Table 3. Search Concepts and Retrieval Logic

Concept Block	Purpose	Indicative Terms
Educational robotics	Identify learning activities centered on building, programming, controlling, or using robots.	"educational robot*"; "education* robotic*"; "robotics-based learning"; "robot-assisted learning"; "robot-based learning"; "robotics education"; "learning with robot*"; "programmable robot*"; "robotic kit*"
STEM context	Restrict the corpus to integrated STEM or a constituent STEM discipline.	STEM; "science education"; "mathematics education"; "engineering education"; "technology education"; "computer science education"; "physics education"
Conceptual learning	Capture direct conceptual evidence and studies using broader outcome terminology.	"conceptual understanding"; "conceptual knowledge"; "concept learning"; "concept attainment"; "conceptual change"; misconception*; "content knowledge"; "knowledge acquisition"; "learning achievement"; "learning performance"
Limits	Reduce incomparability and partial-year bias while retaining major educational-robotics outlets.	Publication years 2000–2025; English; article, conference paper, or review for Dataset A; primary empirical studies only for Dataset B.

The proposed string must be pilot-tested against a benchmark set consisting of highly relevant studies already identified in the background review. A query is accepted only when it retrieves every benchmark article that is indexed in Scopus. If a benchmark record is missed, the team revises the synonym block and documents the change before final retrieval. Backward reference checking and forward citation searching are then conducted for the studies admitted to Dataset B. Records discovered through supplementary searching are labeled separately so that the database search remains reproducible.

Eligibility Criteria and Construction of the Two Corpora

When several publications report the same intervention and participant sample, they are linked as a study family. The most complete report is used as the primary source for outcome coding, while companion reports supply missing methodological details. Each publication remains a distinct bibliographic item in Dataset A, but the study family is counted once in the systematic synthesis to avoid double counting educational evidence.

Table 4. Search Concepts and Retrieval Logic

Criterion	Include	Exclude
Topic	A robot or programmable robotic system is a central learning object, learning tool, tutor, peer, or experimental platform.	Robotics is mentioned only as context; industrial automation, surgical robotics, rehabilitation, or robot control research without a learning process.
Educational domain	Integrated STEM or science, mathematics, engineering, technology, computer science, or physics education.	Language learning, therapy, social assistance, or special education without an explicit STEM concept or STEM learning objective.
Population and setting	Learners in early childhood, primary, secondary, vocational, higher, teacher education, or structured informal STEM programs.	Technical evaluations with no learners; purely expert or developer studies.

Outcome	Conceptual understanding, content knowledge, conceptual change, misconception, explanation, prediction, representation, transfer, or a curriculum-aligned knowledge outcome.	Only satisfaction, acceptance, motivation, attitude, engagement, usability, or intention to use, with no knowledge evidence.
Publication	English-language article, conference paper, or review published in 2000–2025 with complete Scopus metadata for Dataset A.	Editorial, note, book review, erratum, abstract-only item, non-English record, or publication outside the period.
Empirical subset	Primary quantitative, qualitative, or mixed-method study reporting learner-level data and sufficient methodological information.	Review, bibliometric paper, conceptual essay, curriculum description without evaluation, protocol, or duplicate report of the same participant sample.

Data Management, Screening, and Audit Trail

Scopus records are exported on the same day in both CSV and BibTeX formats with complete citation information, references, affiliations, abstracts, author keywords, index keywords, funding details, and document identifiers. A read-only raw export is retained. Cleaning is performed on a working copy using DOI as the primary duplicate key, followed by Scopus EID and a normalized combination of title, first author, and year. Author names, institutions, countries, journal titles, and keywords are standardized through a documented thesaurus; raw and cleaned values are both preserved. Two reviewers independently screen titles and abstracts using a calibrated eligibility form. After a pilot of at least 30 records, disagreements are discussed and ambiguous definitions are clarified without changing the review question. Potentially eligible records then undergo independent full-text assessment. A single primary reason for exclusion is recorded for every full-text exclusion using a prespecified hierarchy: wrong topic, wrong educational domain, no learner sample, no eligible outcome, non-empirical publication, duplicate study, or insufficient report. Disagreements are resolved by consensus and, if necessary, a third reviewer. The final PRISMA diagram reports the number of records at each stage (Page et al., 2021).

Data Extraction and Alignment with the Research Questions

A structured coding form is developed before full extraction and pilot-tested on at least five diverse studies. Bibliometric fields are imported directly where possible; educational variables are coded manually from the full text. The coding manual defines every category, decision rule, permissible value, and example. Multi-component interventions are coded at the level of the reported robotics condition, and multiple conceptual outcomes are retained rather than collapsed into a single favorable result.

Table 5. Alignment of Research Questions, Data, Analysis, and Outputs

RQ	Core Data	Analytical Procedure	Primary Output
RQ1	Year, document type, source, citations, publication age.	Annual production; growth rate; citations per year; source productivity.	Temporal development and publication-impact profile.
RQ2	Authors, affiliations, countries, cited references, citing documents.	Productivity and impact; co-authorship; co-citation; bibliographic coupling.	Influential contributors, collaboration structure, and intellectual base.
RQ3	Author/index keywords, titles, abstracts, publication year.	Co-word networks; thematic map; trend topics; overlay and thematic evolution.	Dominant, emerging, declining, and bridging themes.
RQ4	Learner level, sample, country, STEM concept, robot role/platform, pedagogy, duration, design, and instrument.	Frequency tables; cross-tabulation; evidence matrices; qualitative content coding.	Profile of how conceptual learning is designed and assessed.
RQ5	Direction and magnitude of findings, quality criteria, limitations, and implementation conditions.	Direction-of-effect synthesis; quality-stratified comparison; gap analysis.	Robust patterns, mixed/null findings, methodological gaps, and research agenda.

Performance analysis covers annual scientific production, compound annual growth when the initial annual count is nonzero, document and source productivity, total citations, citations per publication, and citations normalized by publication age. Rankings based on raw citations and citations-per-year are reported together because older publications have a longer citation window. Contributor tables report both publication counts and citation influence; country output is distinguished as single-country and multiple-country publication to avoid equating international collaboration with domestic productivity. Science mapping comprises four complementary networks: co-authorship for the social structure of collaboration; reference and source co-citation for the intellectual base; document bibliographic coupling for current research fronts; and keyword co-occurrence for the conceptual structure. Networks are generated with bibliometrix in R and VOSviewer (Aria & Cuccurullo, 2017; van Eck & Waltman, 2010). Fractional counting is used for co-authorship and country collaboration so that papers with many contributors do not dominate the network; the rationale and consequences of fractional versus full counting are reported (Perianes-Rodriguez et al., 2016). VOSviewer association-strength normalization is applied, and node size, link strength, cluster membership, and average publication year are interpreted jointly rather than as isolated visual features.

Minimum-occurrence thresholds are not selected solely to produce visually attractive maps. The main threshold is chosen to retain an interpretable network while preserving at least 70% of the weighted occurrence mass; one lower and one higher threshold are analyzed as sensitivity checks. Keyword cleaning merges spelling variants, singular and plural forms, abbreviations, and clear synonyms but does not merge theoretically distinct constructs such as conceptual understanding, computational thinking, problem solving, and academic achievement. The thesaurus and threshold decisions are archived with the analysis files. Thematic evolution is examined across three prespecified phases: emergence (2000–2010), expansion (2011–2017), and consolidation (2018–2025). For each phase, themes are positioned using centrality and density, while Sankey-style links trace the persistence, merging, or fragmentation of keyword clusters. Overlay maps based on average publication year are used as a complementary continuous view. Interpretations are checked against titles and abstracts of the most representative documents in each cluster to reduce the risk of naming a theme from isolated keywords.

Systematic Coding of Conceptual Understanding

Conceptual understanding is coded as a multidimensional outcome rather than a synonym for total achievement. Evidence is classified into three levels. Level A (direct conceptual evidence) requires explanation, prediction, representation translation, application to a novel context, misconception diagnosis, or transfer. Level B (proximal knowledge evidence) consists of curriculum-aligned tests that contain identifiable conceptual items but do not explicitly assess transfer or explanation. Level C (indirect evidence) consists of composite achievement scores, robot-task completion, code or artifact performance, or teacher judgments from which conceptual understanding cannot be isolated. Perception-only measures are not classified as conceptual evidence. For every empirical study, the review records the target STEM discipline and specific concept; learner level and prior experience; sample size; country and setting; physical, virtual, or hybrid robot; platform and robot role; degree of learner control; pedagogy; group structure; intervention duration and intensity; comparison condition; research design; instrument type and evidence of instrument validity; timing of measurement; conceptual-evidence level; and direction of the finding. Reported effects are categorized as positive, mixed, null, or negative using the authors' statistical results and qualitative evidence, while preserving effect sizes and uncertainty estimates when available.

Synthesis proceeds from descriptive distributions to structured cross-study comparison. Evidence matrices cross-tabulate STEM concept, robot role, pedagogy, learner level, intervention duration, instrument, and finding. A direction-of-effect display is used when outcome metrics are too heterogeneous for statistical pooling. A meta-analysis is undertaken only if at least five independent studies evaluate sufficiently comparable concepts, contrasts, and outcomes; otherwise, effect sizes are reported without a pooled estimate. This decision is made from methodological comparability, not from the statistical significance of individual studies.

Methodological Quality, Reliability, and Bias Control

The Mixed Methods Appraisal Tool (MMAT), version 2018, is used because the empirical corpus is expected to contain qualitative, randomized, nonrandomized, descriptive quantitative, and mixed-method studies (Hong et al., 2018). Two reviewers independently appraise the relevant MMAT criteria for each design. Criterion-level judgments and supporting quotations are retained. Studies are not excluded solely

because of a low summary score, and no overall percentage score is used; instead, quality profiles inform interpretation and sensitivity analyses. Interrater agreement is calculated separately for title-abstract screening, full-text eligibility, key categorical variables, conceptual-evidence level, and MMAT judgments. Percent agreement and Cohen's kappa are reported, with confidence intervals where feasible (McHugh, 2012). A target kappa of at least 0.80 is set for the pilot phase. Values below 0.80 trigger clarification of the coding manual and a second pilot; they are not improved by deleting difficult records. After calibration, disagreements remain visible in the audit file and are resolved through consensus.

Several bias checks are incorporated. Citation-age bias is addressed through citations-per-year; self-citation sensitivity is reported when supported by the exported data; database bias is acknowledged because Scopus does not index every regional outlet; publication-type sensitivity compares results with and without conference papers and reviews; network robustness is tested across thresholds; and systematic conclusions are compared across MMAT profiles and conceptual-evidence levels. These checks prevent highly cited, low-quality, or indirectly measured studies from dominating the educational conclusions.

Integration, Transparency, and Reproducibility

Integration occurs through a joint display that links each bibliometric theme to the empirical studies located within that theme and summarizes their STEM concepts, pedagogies, evidence levels, quality profiles, and findings. Agreement between macro and micro evidence is examined explicitly. For example, a rapidly growing keyword cluster is interpreted as an educationally mature research front only when it is supported by multiple methodologically credible studies with direct conceptual evidence. A prominent cluster supported mainly by perception surveys is instead identified as a visibility-evidence gap. The supplementary package should contain the complete Scopus query and retrieval date; PRISMA screening log; raw and cleaned exports where licensing permits; deduplication rules; thesaurus file; inclusion decisions and exclusion reasons; coding manual; anonymized extraction sheet; MMAT judgments; software names and versions; network parameters; and analysis scripts. Only bibliographic and published study data are analyzed, so formal human-participant ethics approval is not required. Nevertheless, the review protocol, analytic decisions, and any deviations from the planned method are documented to support transparency and replication.

RESULTS AND DISCUSSION

Results

Corpus Screening and Analytical Subsets

The Scopus export retrieved on July 29, 2026 contained 59 records published between 2001 and 2025. Title–abstract screening removed eight false-positive records in which robotics was peripheral or absent from the reported intervention. The principal bibliometric corpus therefore comprised 51 documents. Within that corpus, 34 primary empirical studies reported a direct, proximal, or indirect measure that could inform the conceptual-understanding question. Seven further empirical records focused on perceptions, self-efficacy, motivation, or implementation without a usable conceptual outcome; these records were retained for field mapping but excluded from the evidence profile. The remaining ten records consisted of reviews, frameworks, design descriptions, or other publications without primary learner-level conceptual evidence.

Table 6. Construction of the Bibliometric and Evidence Corpora

Screening Stage or Subset	n	Analytical Treatment
Records retrieved from Scopus	59	Initial export; publication years 2001–2025
Topically excluded after title–abstract screening	8	Robotics was absent, incidental, or not the central learning intervention
Dataset A: screened bibliometric corpus	51	Used for performance analysis, contributors, countries, and keyword mapping
Dataset B: empirical conceptual-evidence subset	34	Used for population, domain, pedagogy, robot-role, and evidence coding
Empirical perception/non-concept studies	7	Retained in Dataset A; excluded from conceptual-evidence synthesis
Other non-primary or non-concept documents	10	Retained in Dataset A where topically relevant; not synthesized as effects

Interpretive Boundary

The systematic component reported below is an abstract-stage evidence profile. The export did not contain full texts; consequently, full-text exclusions, instrument-level validity checks, effect-size extraction, and MMAT judgments have not yet been completed. Counts describe the available abstracts and must not be read as a final causal or quality-adjusted synthesis.

Temporal Development and Citation Structure (RQ1)

The 51 publications received 693 Scopus citations. Citation impact was strongly skewed: the mean was 13.59 citations per document, whereas the median was 5; 13 documents were uncited, the corpus h-index was 13, and the citation Gini coefficient was 0.727. This inequality means that the arithmetic mean alone overstates the visibility of a typical publication. The field was also highly recent. Twenty-nine documents (56.9%) appeared during 2023–2025, and 39 (76.5%) appeared during 2020–2025. Annual production remained at no more than three publications per year through 2019, except for isolated gaps, before increasing to six documents in 2020, ten in 2024, and sixteen in 2025 (Figure 2). The 2024 citation peak was driven by a small group of rapidly cited publications, especially the multilevel meta-analysis by Ouyang and Xu (Ouyang & Xu, 2024). The low citation total for 2025 should therefore be interpreted as a citation-window effect rather than evidence of reduced influence. The publication portfolio was almost evenly divided between journal articles (25; 49.0%) and conference papers (24; 47.1%), with only two reviews (3.9%), indicating that research dissemination remains closely connected to technical and computing-education conferences.

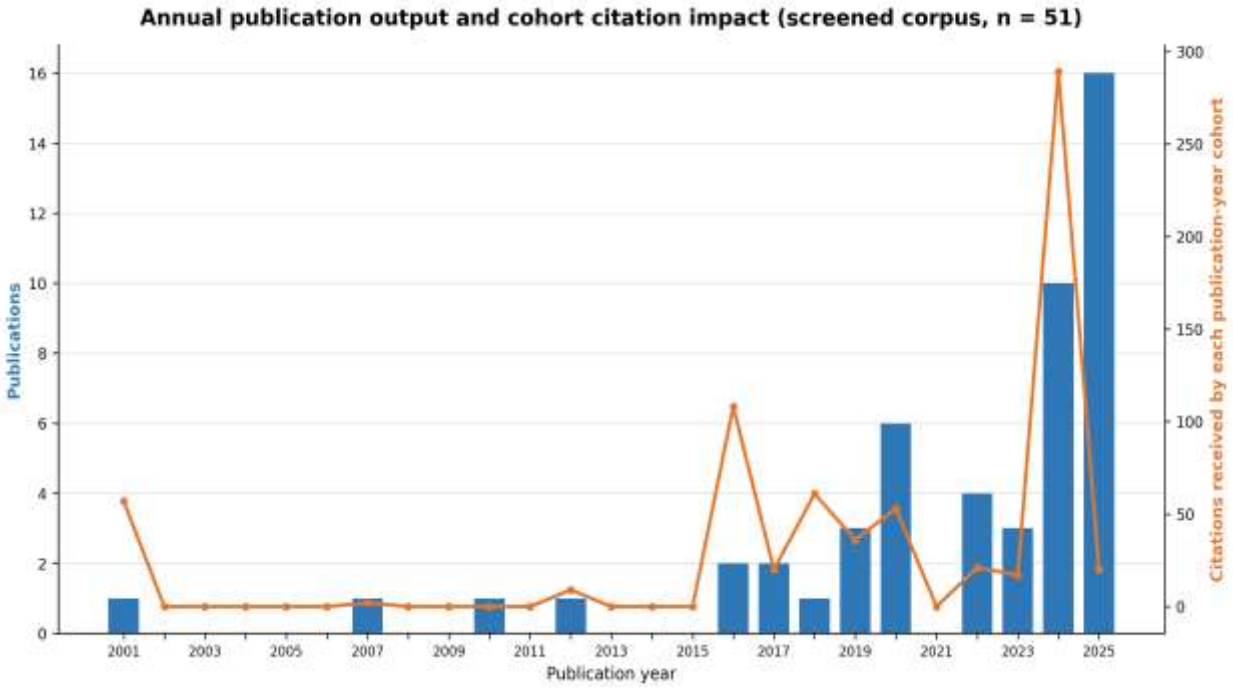


Figure 2. Annual Scientific Production and Citations in the Screened Corpus

Influential Documents, Contributors, Sources, and Geographies (RQ2)

Citation influence was concentrated in a compact intellectual core. The five most-cited documents accumulated 404 citations, or 58.3% of all citations in the corpus. Ouyang and Xu's multilevel meta-analysis led both total and annualized impact, followed in total citations by the US robotics project reported by Nugent et al., Arduino-based maker education, the ICRE framework, and Nagai's early practical-robotics article (Alam & Mohanty, 2024; Chou, 2018; Nagai, 2001; Nugent et al., 2016; Ouyang & Xu, 2024). Because several highly ranked records were recent, citations per year were reported alongside raw counts to reduce age bias.

Table 7. Most Influential Documents in the Screened Corpus

Author(s), year	Document	Citations	Citations/Year
Ouyang & Xu, 2024	Effects of educational robotics: multilevel meta-analysis	148	74.00
Nugent et al., 2016	Robotics camps, clubs, and competitions	79	7.90
Chou, 2018	Maker education with Arduino-based educational robotics	61	7.63
Alam & Mohanty, 2024	Integrated constructive robotics in education model	59	29.50
Nagai, 2001	Learning while doing: practical robotics education	57	2.28
Hu et al., 2024	Educational robot and low-cost self-made toys	30	15.00
Štuikys et al., 2016	Robot-based generative learning objects	29	2.90
Chiu et al., 2024	Robots in science education: review and bibliometric analysis	22	11.00

Authorship and source output were fragmented. The corpus contained 174 unique authors, with a mean of 3.71 authors per document. Xiaowen Gong, Daniela Marghitu, and Chih-Chien Hu were the most productive authors, but each contributed only three publications. Auburn University and Tatung University led the institutional count with three documents each. National Taiwan Normal University, Zhejiang University, South China Normal University, the Technion, the University of Nebraska–Lincoln, and National Taiwan University of Science and Technology each contributed two. Forty-three sources published the 51 documents; the ASEE Annual Conference and Exposition proceedings contributed six, while every journal published no more than two corpus documents. Thus, no single journal served as a stable disciplinary home for the topic.

The United States produced the largest number of documents under full country counting (14), followed by Taiwan (10) and China (8). These three locations accounted for 32 country-document occurrences and dominated both output and citation visibility. Only ten documents involved more than one country, yielding an international collaboration rate of 19.6%; China–United States was the only bilateral pair appearing twice. The field is therefore internationally visible but still organized mainly through domestic teams, a pattern that restricts direct claims about transfer across curricula, school systems, resources, and cultural settings.

Table 8. Most Influential Documents in the Screened Corpus

Dimension	Leading Unit(s)	Value	Interpretation
Productive authors	Gong; Marghitu; Hu	3 documents each	No dominant author group; dispersed expertise
Most productive source	ASEE Annual Conference proceedings	6 documents	Conference-oriented dissemination
Leading institutions	Auburn University; Tatung University	3 documents each	Institutional leadership was modest and distributed
Source diversity	All indexed outlets	43 sources	Highly fragmented publication base
Leading countries	United States; Taiwan; China	14; 10; 8	Geographic concentration in three systems
International documents	Multicountry affiliations	10 (19.6%)	Limited cross-national evidence production
Strongest country link	China–United States	2 documents	All other bilateral links occurred once

Conceptual Structure and Thematic Evolution (RQ3)

Author keywords were available for 46 of the 51 documents. After controlled normalization of spelling, plurality, acronyms, and near-synonyms, educational robotics was the dominant keyword (32 documents), followed by STEM education (16). Four interpretable co-occurrence clusters emerged (Figure 3). The first represented the field's educational-robotics and STEM core. The second connected computational thinking with augmented reality, digital twins, and educational technology. The third captured constructionist pedagogy through project-based, hands-on, collaborative, maker, and K–12

learning. The fourth linked artificial intelligence and machine learning with broadening participation and STEM persistence.

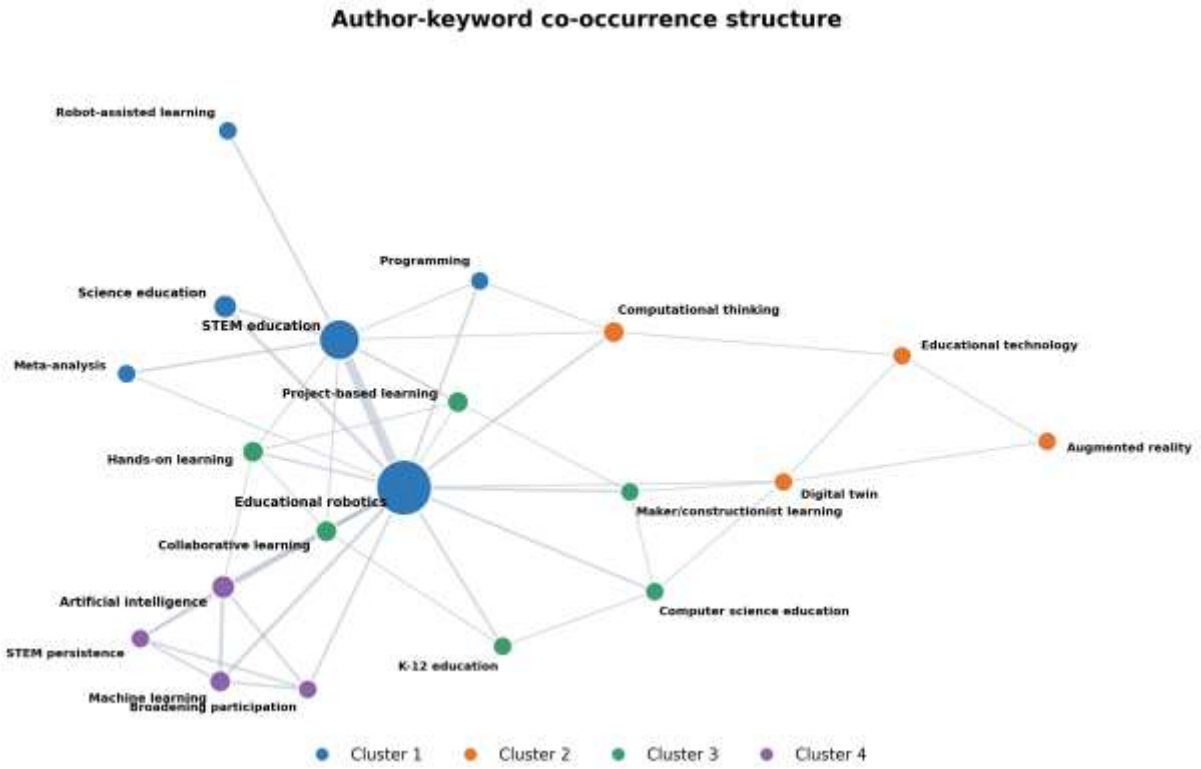


Figure 3. Co-occurrence Network of Normalized Author Keywords

Table 9. Most Influential Documents in the Screened Corpus

Cluster	Leading Normalized Keywords	Substantive Interpretation
1. Core field	Educational robotics (32); STEM education (16); science education (4); programming (2); robot-assisted learning (2)	Stable conceptual center connecting robotics with formal STEM learning
2. Digital and computational tools	Computational thinking (3); augmented reality (2); digital twin (2); educational technology (2)	Hybridization of physical robotics with digital representations and tools
3. Constructionist pedagogy	Project-based, hands-on, and collaborative learning (3 each); K–12, computer science, and maker learning (2 each)	Learning through making, programming, testing, and peer interaction
4. AI and equity frontier	Artificial intelligence (4); machine learning (3); broadening participation (2); STEM persistence (2)	Emerging emphasis on intelligent systems, participation, and retention

The temporal keyword profile suggests three stages of development. During the emergence phase (2001–2010), isolated publications emphasized practical robotics, system integration, manipulators, and early disciplinary concepts such as angle. The expansion phase (2011–2017) introduced computer-science education, undergraduate learning, generative learning objects, and STEM interest. The consolidation phase (2018–2025) established educational robotics and STEM education as the central vocabulary while adding artificial intelligence, machine learning, computational thinking, project-based learning, hands-on learning, and collaboration. The shift is therefore not simply numerical growth; it represents a movement from robotics instruction toward pedagogically framed and increasingly hybrid STEM-learning environments.

Educational and Methodological Profile of the Evidence Subset (RQ4)

The 34 empirical studies were evenly distributed across primary, secondary, and higher education, with nine studies (26.5%) at each level. Teacher education and informal or mixed K–12 settings contributed three studies each. Disciplinary coverage was less balanced: computing or robotics knowledge was assessed in 17 studies (50.0%), integrated STEM knowledge in ten (29.4%), and mathematics and physics in only

three studies each (8.8%). Consequently, the literature's broad STEM label masks a substantial concentration on computing and robotics concepts.

In 22 studies (64.7%), the robot functioned primarily as an object to be built, programmed, tested, or debugged. Only four studies each treated the robot as a teaching agent or peer, as an experimental or representational tool, or in a hybrid/unspecified role. Project-based learning (10 studies) and hands-on learning (9) were the most frequent pedagogies, followed by collaborative learning (6) and robot-assisted instruction (4). Because pedagogy was coded as a multilabel variable, one study could contribute to several categories.

Table 10. Educational Profile of the 34 Empirical Evidence Studies

Dimension	Distribution	Main Reading
Population	Primary 9 (26.5%); secondary 9 (26.5%); higher education 9 (26.5%); teacher education 3 (8.8%); informal/mixed K–12 3 (8.8%); unreported 1	Age-level coverage was comparatively balanced
STEM domain	Computing/robotics 17 (50.0%); integrated STEM 10 (29.4%); mathematics 3 (8.8%); physics 3 (8.8%); engineering/design 1 (2.9%)	Content-specific evidence outside computing remained thin
Robot role	Construction/programming object 22 (64.7%); teaching agent/peer 4 (11.8%); experimental/representation tool 4 (11.8%); hybrid/unspecified 4 (11.8%)	Robots were usually things students made or programmed
Leading pedagogies	Project-based 10 (29.4%); hands-on 9 (26.5%); collaborative 6 (17.6%); robot-assisted instruction 4 (11.8%)	Active construction and iteration defined the pedagogical core
Additional pedagogies	Professional development 3; programming practice 3; inquiry 2; experiential 2; pair learning 2; other named approaches 1 each	A long tail of designs had not yet accumulated replication

Patterns of Conceptual Evidence (RQ5)

Only ten of the 34 empirical studies (29.4%) provided Level A evidence: a direct measure of conceptual understanding, such as explanations, misconception change, conceptual transfer, or a clearly identified disciplinary concept. Nineteen studies (55.9%) provided Level B evidence through proximal knowledge or learning-performance tests, and five (14.7%) provided Level C evidence through indirect achievement-related indicators. The evidence base was therefore broader than it was conceptually deep. At the abstract-reporting level, 24 studies (70.6%) described positive findings, eight (23.5%) reported mixed findings, one (2.9%) reported a null result, and one (2.9%) was unclear (Figure 4). These counts are a direction-of-effect inventory, not a pooled estimate: study designs, comparison conditions, instruments, durations, and outcome constructs differed substantially. Positive abstract language also cannot rule out selective reporting or publication bias.

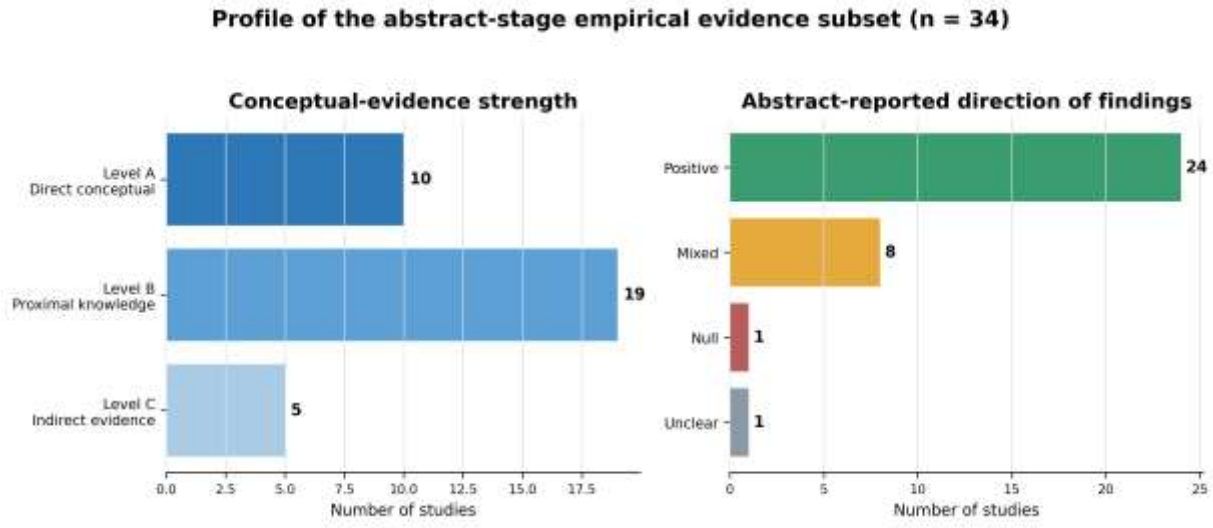


Figure 4. Abstract-stage Profile of Conceptual Evidence

The direct studies nevertheless illuminate several plausible learning mechanisms. A robot peer's observable failures can support productive failure and classroom learning (L. Chen et al., 2025); gestural interaction with a remotely controlled robot can embody the abstract notion of friction (Merkouris et al., 2019); and low-cost robots or maker systems can connect design decisions with STEM knowledge and performance (Chou, 2018; Hu et al., 2024). A 6E learning cycle combined with a robot teaching assistant further illustrates how the robot's role and the surrounding pedagogy can be designed together rather than treated as interchangeable features (Hsiao et al., 2024). These examples support mechanism-oriented interpretation, but the small number of direct studies prevents a general claim that every robotics activity produces conceptual change.

DISCUSSION

The Field Is Expanding Faster Than Its Evidential Consolidation

The bibliometric results show a rapidly expanding but still fragmented research field. More than half of the publications appeared in the final three years of the observation period, yet output was dispersed across 43 sources and nearly half of the documents were conference papers. This configuration is typical of an active technology-education domain in which prototypes, curriculum designs, and classroom implementations develop quickly, but shared measurement standards and cumulative journal-based programs mature more slowly. The pattern extends prior reviews that identified growing interest, varied robot roles, and diverse pedagogical designs in robotics-based STEM and science education (Chiu et al., 2024; Darmawansah et al., 2023). Citation concentration reinforces this interpretation. A small number of syntheses and framework papers have become intellectual anchors, while 13 documents remain uncited. High citation visibility is valuable for identifying organizing ideas, but it is not equivalent to learner-level evidence. For example, a review or conceptual framework may shape the field without testing a specific conceptual outcome. The joint design used here therefore prevents the most-cited documents from automatically determining conclusions about teaching effectiveness.

What Robotics Appears to Contribute to Conceptual Learning

The evidence suggests that robotics is most educationally productive when it makes a disciplinary relationship observable and revisable. Building and programming a robot externalizes a learner's model: a prediction is translated into code or mechanism, the physical system returns feedback, and a mismatch creates a reason to diagnose and revise the underlying idea. This mechanism is consistent with the dominance of project-based and hands-on learning in the evidence subset and with maker-oriented reports linking design activity to knowledge acquisition (Chou, 2018; Hu et al., 2024). It also explains why a robot can function as more than a motivational device: the artifact becomes a testable representation of the learner's reasoning. However, the dominant construction/programming role also narrows what has been demonstrated. Half of the empirical subset assessed computing or robotics knowledge, whereas only six studies focused specifically on mathematics or physics. Direct conceptual studies on friction, angle, or mechanics show that robots can embody disciplinary concepts, but these applications remain comparatively sparse. Future studies should specify the target concept before selecting the robot and then align the robot's affordance with a known conceptual difficulty—for example, using motion, sensing, and feedback to contrast competing models rather than adding a robot to an otherwise unchanged activity.

The direction-of-effect profile is encouraging but deliberately weaker than a meta-analytic conclusion. Ouyang and Xu reported an overall moderate effect of educational robotics with important variation by discipline, while Bai and Tian reported medium effects for conceptual knowledge ($g = 0.636$) and applied skills ($g = 0.663$), and a smaller effect for learning attitudes ($g = 0.422$) (Bai & Tian, 2025; Ouyang & Xu, 2024). The present review does not reproduce those pooled estimates. Instead, it shows why further construct-level synthesis is necessary: only ten included empirical studies measured conceptual understanding directly, while most relied on broader knowledge or performance tests.

Visibility Evidence Gaps in Emerging Themes

The keyword network positions artificial intelligence, machine learning, augmented reality, and digital twins as visible research fronts. These themes are important because they can extend physical robotics with adaptive feedback, simulation, analytics, and multiple representations. Yet keyword visibility does not establish conceptual effectiveness. The emerging cluster contains fewer direct conceptual studies than its contemporary prominence might imply, creating a visibility–evidence gap. Research on these technologies

should therefore compare what learners can explain or transfer, not only whether they complete a task, enjoy the system, or perceive it as useful.

A second gap concerns equity and contextual transfer. Broadening participation and STEM persistence appear in the keyword map, but the bibliometric production system is concentrated in the United States, Taiwan, and China, and fewer than one in five documents involve international collaboration. Claims about access and persistence should be tested across lower-resource schools, different curricula, linguistic contexts, and varying levels of teacher expertise. Low-cost self-made tools provide one promising design direction (Hu et al., 2024), but affordability, maintenance, teacher preparation, and accessibility must be reported as implementation variables rather than treated as background details.

Implications for STEM Teaching and Research Design

For teaching, the central implication is to design the learning sequence before choosing the technology. A concept-focused robotics unit should identify the target model or misconception, elicit an initial prediction, make the relevant variables visible through robot behavior or sensor data, require learners to explain a mismatch, and assess transfer to a non-robotic context. Structured cycles such as 6E can help connect engagement, exploration, explanation, engineering, enrichment, and evaluation when each phase is explicitly tied to the target concept (Hsiao et al., 2024). For research, conceptual understanding should be operationalized separately from motivation, self-efficacy, computational skill, and task completion. Stronger studies would use validated concept inventories where available, open-ended explanations or misconception coding, delayed retention, and transfer tasks; report intervention duration and teacher support; describe the comparison condition precisely; and publish sufficient statistics for effect-size calculation. Multisite and cross-country studies are also needed to move the field beyond isolated demonstrations.

Table 11. Evidence-Based Agenda for Robotics-Based STEM Learning

Observed issue	Educational implication	Research recommendation
Most studies used proximal knowledge tests	State the target concept and misconception before the robotics activity	Use direct explanation, transfer, retention, or misconception measures
Robots usually served as construction/programming objects	Make robot behavior a representation of disciplinary reasoning	Compare robot roles while holding pedagogy and content constant
Project-based and hands-on learning dominated	Preserve prediction, testing, explanation, and revision cycles	Report implementation fidelity and the active pedagogical ingredients
Computing/robotics content dominated	Develop concept-specific applications in physics, mathematics, and science	Accumulate comparable studies within the same disciplinary construct
AI and hybrid technologies are emerging	Use adaptive or simulated features only when they clarify a learning mechanism	Test conceptual gain rather than usability or engagement alone
International collaboration was limited	Adapt designs to curriculum, resources, language, and teacher capacity	Conduct multisite, cross-cultural, and lower-resource replications

Limitations and Next Analytical Steps

Several limitations qualify the results. First, the search used one database, English-language metadata, and a 2001–2025 publication window; regional journals or differently indexed terminology may therefore be underrepresented. Second, citation counts are dynamic and favor older documents, although citations per year were added as a partial correction. Third, five records lacked author keywords, and network structure depends on normalization and inclusion thresholds. Fourth, country and collaboration results used full counting, so a multicountry paper contributed once to every participating country.

Most importantly, the systematic synthesis is provisional because the supplied export contains abstracts rather than full texts. The present coding cannot verify the exact conceptual instrument, sample attrition, intervention fidelity, comparison condition, statistical precision, or risk of bias. The next stage should retrieve all candidate full texts, complete dual eligibility screening, apply MMAT by study design, extract effect sizes where outcomes are sufficiently comparable, and rerun the synthesis by evidence level and disciplinary concept. Until that stage is complete, the defensible conclusion is that robotics-based STEM learning is a fast-growing and often positively reported field whose direct evidence for conceptual change remains promising but limited and unevenly distributed.

CONCLUSION

This integrated review shows that robotics-based learning and conceptual understanding have developed into a rapidly expanding but still fragmented area of STEM education research. The screened corpus comprised 51 documents published between 2001 and 2025, accumulated 693 citations, and was dominated by recent growth, with 56.9% of publications appearing during 2023–2025. Educational robotics and STEM education formed the conceptual core of the field, while artificial intelligence, machine learning, augmented reality, and digital twins emerged as newer research fronts. Nevertheless, production was dispersed across 43 sources, international collaboration was limited to 19.6% of documents, and research activity remained geographically concentrated in the United States, Taiwan, and China. The evidence synthesis further indicates that the field's growth has not yet been matched by equivalent depth in the measurement of conceptual change. Among 34 empirical studies, only ten assessed conceptual understanding directly, while 19 used proximal knowledge measures and five relied on indirect indicators. Robots were most frequently used as objects for construction or programming, and the empirical literature concentrated on computing, robotics, project-based learning, and hands-on activity. Accordingly, robotics should not be treated as an inherently effective instructional technology; its contribution depends on whether robot behavior is deliberately connected to a target concept, prediction, explanation, revision, and transfer. Future studies should prioritize validated conceptual measures, comparable disciplinary outcomes, delayed assessment, transparent comparison conditions, and multicountry replication. Because the present systematic synthesis is based on title–abstract evidence, these conclusions remain provisional until full-text screening, MMAT appraisal, and effect-size extraction are completed.

REFERENCES

Akgün, M., & Atıcı, B. (2023). Bibliometric map of educational robotics studies. *International Journal of Technology in Education*, *6*(2), 283–309. https://doi.org/10.46328/ijte.402

Alam, A., & Mohanty, A. (2024). Integrated constructive robotics in education (ICRE) model: A paradigmatic framework for transformative learning in educational ecosystem. *Cogent Education*, *11*(1), 2324487. https://doi.org/10.1080/2331186X.2024.2324487

Aleksandrova, N. S., Sokolova, E. G., Kosarenko, N. N., Ibragimova, E. M., Budkevich, R. L., & Yagudina, R. I. (2024). Exploring the frontiers: A comprehensive bibliometric analysis of robotics in science education. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(10), em2520. https://doi.org/10.29333/ejmste/15432

Anwar, S., Bascou, N. A., Menekse, M., & Kardgar, A. (2019). A systematic review of studies on educational robotics. *Journal of Pre-College Engineering Education Research*, *9*(2), Article 2. https://doi.org/10.7771/2157-9288.1223

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems*, *75*, 661–670. https://doi.org/10.1016/j.robot.2015.10.008

Bai, S., & Tian, P. (2025). Educational robotics may enhance students' conceptual knowledge, applied skills, and learning attitude in STEM education: A meta-analysis. *Educational Technology and Society*, *28*(4), 274–303. https://doi.org/10.30191/ETS.202510_28(4).SP06

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, *58*(3), 978–988. https://doi.org/10.1016/j.compedu.2011.10.006

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education*, *72*, 145–157. https://doi.org/10.1016/j.compedu.2013.10.020

Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., & Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & Education*, *109*, 162–175. https://doi.org/10.1016/j.compedu.2017.03.001

Chen, L., Cai, Y., Fang, Y., Yang, Z., Xia, D., You, J., Xiao, S., Song, Y., Zhan, L., Chen, J., & Sun, L. (2025). Observing a robot peer's failures facilitates students' classroom learning. *Science Robotics*, *10*(106), eadu5257. https://doi.org/10.1126/scirobotics.adu5257

Chiu, M.-C., Hwang, G.-J., & Tu, Y.-F. (2024). Roles, applications, and research designs of robots in science education: A systematic review and bibliometric analysis of journal publications from 1996 to 2020. *Interactive Learning Environments*, *32*(5), 1834–1859. https://doi.org/10.1080/10494820.2022.2129392

Chou, P.-N. (2018). Skill development and knowledge acquisition cultivated by maker education: Evidence from Arduino-based educational robotics. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(10), 1–15. https://doi.org/10.29333/ejmste/93483

Darmawansah, D., Hwang, G.-J., Chen, M.-R. A., & Liang, J.-C. (2023). Trends and research foci of robotics-based STEM education: A systematic review from diverse angles based on the technology-based learning model. *International Journal of STEM Education*, *10*(1), 12. https://doi.org/10.1186/s40594-023-00400-3

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Evripidou, S., Georgiou, K., Doitsidis, L., Amanatiadis, A. A., Zinonos, Z., & Chatzichristofis, S. A. (2020). Educational robotics: Platforms, competitions and expected learning outcomes. *IEEE Access*, *8*, 219534–219562. https://doi.org/10.1109/ACCESS.2020.3042555

Hong, Q. N., Pluye, P., 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. (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

Hsiao, H.-S., Chen, J.-H., Chang, T.-L., Li, P.-H., & Chung, G.-H. (2024). A study on the effects of using the 6E model and a robot teaching assistant on junior high school students' STEM knowledge, learning motivation, and hands-on performance. *Journal of Science Education and Technology*, *33*(5), 759–778. https://doi.org/10.1007/s10956-024-10119-7

Hu, C.-C., Yang, Y.-F., Cheng, Y.-W., & Chen, N.-S. (2024). Integrating educational robot and low-cost self-made toys to enhance STEM learning performance for primary school students. *Behaviour and Information Technology*, *43*(8), 1614–1635. https://doi.org/10.1080/0144929X.2023.2222308

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*, 11. https://doi.org/10.1186/s40594-016-0046-z

López-Belmonte, J., Segura-Robles, A., Moreno-Guerrero, A.-J., & Parra-González, M.-E. (2021). Robotics in education: A scientific mapping of the literature in Web of Science. *Electronics*, *10*(3), 291. https://doi.org/10.3390/electronics10030291

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Merkouris, A., Chorianopoulou, B., Chorianopoulos, K., & Chrissikopoulos, V. (2019). Understanding the notion of friction through gestural interaction with a remotely controlled robot. *Journal of Science Education and Technology*, *28*(3), 209–221. https://doi.org/10.1007/s10956-018-9760-2

Nagai, K. (2001). Learning while doing: Practical robotics education. *IEEE Robotics and Automation Magazine*, *8*(2), 39–43. https://doi.org/10.1109/100.932756

Nugent, G., Barker, B., Grandgenett, N., & Welch, G. (2016). Robotics camps, clubs, and competitions: Results from a US robotics project. *Robotics and Autonomous Systems*, *75*, 686–691. https://doi.org/10.1016/j.robot.2015.07.011

Ouyang, F., & Xu, W. (2024). The effects of educational robotics in STEM education: A multilevel meta-analysis. *International Journal of STEM Education*, *11*(1), 7. https://doi.org/10.1186/s40594-024-00469-4

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*, n71. https://doi.org/10.1136/bmj.n71

Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA Statement for reporting literature searches in systematic reviews. *Systematic Reviews*, *10*(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Sullivan, A., & Bers, M. U. (2016). Robotics in the early childhood classroom: Learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. *International Journal of Technology and Design Education*, *26*(1), 3–20. https://doi.org/10.1007/s10798-015-9304-5

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Xia, L., & Zhong, B. (2018). A systematic review on teaching and learning robotics content knowledge in K-12. *Computers & Education*, *127*, 267–282. https://doi.org/10.1016/j.compedu.2018.09.007

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629

Author (s):

*Riski Ramadani (corresponding author)
Postgraduate Physics, Faculty of Mathematics and Natural Science,
Universitas Negeri Surabaya, Indonesia
Email: riski.20056@mhs.unesa.ac.id

Afiyah Nikmah
Postgraduate Physics, Faculty of Mathematics and Natural Science,
Universitas Negeri Surabaya, Indonesia
Email: afiyahnikmah08@gmail.com

Arum Vonie Rachmawati
Postgraduate Physics, Faculty of Mathematics and Natural Science,
Universitas Negeri Surabaya, Indonesia
Email: arum.20018@mhs.unesa.ac.id

Rohim Aminullah Firdaus
Postgraduate Physics, Faculty of Mathematics and Natural Science,
Universitas Negeri Surabaya, Indonesia
Email: rohimfirdaus@unesa.ac.id

Noer Risky Ramadhani
Department for Higher Education Research, Faculty of Education, and Architecture,
University for Continuing Education Krems, Austria
Email: ramadhanirisky52@gmail.com