frontiers | Frontiers in Education

TYPE Systematic Review
PUBLISHED 10 December 2024
DOI 10.3389/feduc.2024.1480908

Check for updates

OPEN ACCESS

EDITED BY
Dina Tavares,
Polytechnic University of Leiria, Portugal

REVIEWED BY
Anastasia Misirli,
University of Patras, Greece
Georg Jäggle,
Pädagogische Hochschule Wien, Austria

*CORRESPONDENCE
José-María Romero-Rodríguez
✉ romejo@ugr.es

RECEIVED 14 August 2024
ACCEPTED 08 November 2024
PUBLISHED 10 December 2024

CITATION
Trapero-González I, Hinojo-Lucena FJ, Romero-Rodríguez J-M and Martínez-Menéndez A (2024) Didactic impact of educational robotics on the development of STEM competence in primary education: a systematic review and meta-analysis.
Front. Educ. 9:1480908.
doi: 10.3389/feduc.2024.1480908

Didactic impact of educational robotics on the development of STEM competence in primary education: a systematic review and meta-analysis

Irene Trapero-González, Francisco Javier Hinojo-Lucena, José-María Romero-Rodríguez* and Alejandro Martínez-Menéndez

Department of Didactics and School Organization, University of Granada, Granada, Spain

As robotics become increasingly present in about every area of the human daily life scheme, their presence in the educational world has become increasingly common, especially with regard to earlier stages and in relation to disciplines framed within the Science, Technology, Engineering, and Mathematics (STEM) concept, given its innate links to these programmable companions. Consequentially, the main objective of this study is to analyze the evolution in time of robotic learning mediation of STEM-based teaching and learning in the Primary Education stage. To achieve this objective, a systematic review of the literature and a meta-analysis were conducted to retrieve experiences of interest (n = 13; n = 8). The present study indicates that, despite certain hiatuses, this educational research field has been of global increasing interest, implementing a variety of robotic-related products that, in the end, have been reported to cause a moderate benefit regarding the acquisition and strengthening of STEM competences. Future research lines are discussed, pointing to the urgent need to establish a framework of reference for didactic planning around these resources in order to extend their use to every potential teacher in the target stage.

KEYWORDS

primary education, meta-analysis, robotics, STEM, systematic review

1 Introduction

Technological advancements have, undoubtedly, caused unique and irreversible changes in the way we understand, process, and experience a vast, diverse range of areas in our daily lives. These changes range from simply distracting ourselves with content published online to integrating these new tools and gadgets as essential components in our job positions. Garzón et al. (2021) assert that robotics, primarily due to its proximity to human life and tasks, may be one of the most indispensable ones.

Educators, especially those employed at earlier courses and stages, are obligated to initiate training and research regarding the appropriate utilization of these resources and materials. This is a key aspect of pedagogical leadership to remain up-to-date in innovations in the educational domain that enable learners to further push their limits and development (Palomino et al., 2022). According to Martínez et al. (2023a), it is only a matter of time before these robotic products, toys, and commodities are fully integrated into our classrooms and daily teaching, despite initial skepticism regarding their potential utilization and benefits.

These tools, following Papert's (1993) line of reasoning regarding make-believe in classrooms, may be some of the most powerful tools when it comes to boosting abstract thinking in children, therefore removing the notion of being mere toys that has traditionally haunted Educational Robotics and its predecessors.

1.1 Robots and education: a recent yet well-established relationship

Although robotics, as a field external to educational theory and practice, appears to be a well-established field, it is important to establish a precise definition of the idea of a robot that can be appropriate regarding delimitations when addressing different technological assets or products. In this way, according to Kalaitzidou and Pachidis (2023), a robot can be defined as a system of interconnected components, including sensors, actuators, processors, effectors, and the controlling software, or programming environment, itself. This system is controlled through pre-defined, although open to further modifications, programming that can be modified, either directly or indirectly, by a human user (Seckel et al., 2022).

Educational Robotics, hereinafter referred to as ER, as a field of practice and research, could be broadly defined as the implementation of either straight-up robots or any given robotic design part or component in educational settings and proposals with a didactic-pedagogical goal. Nevertheless, potential specifications for how to optimally integrate these elements into classroom-like environments have ended up configuring certain archetypes of methodological designs within this educational scope.

For starters, Martínez et al. (2023b) and Chatzopoulos et al. (2022) establish that using robots as objects of learning themselves, either through their programming with or without prior crafting steps, is usually the main go-to strategy regarding the educational use of robotics. However, Evripidou et al. (2020) argue that implementing robots as prefabricated and already programmed technical tools, as it is common regarding multiple other kinds of technological toys or devices, can be considered a form of ER as well. Krūmiņš et al. (2024) consider online programming platforms as a third possibility for integrating robotics into classrooms, which are digital environments in which the user designs a string of code that will be applied to a simulated robot in a fully digital context or transported to an actual physical entity. The majority of this programming is developed by employing intuitive coding languages, such as block programming software, with *Scratch* being the widely used provider in this regard (Seckel et al., 2022). As suggested by Bers (2021), while having value in and of itself, coding can be considered a language on its own, crucial in the development of children's critical thinking and problem-solving capabilities.

However, it is possible to introduce an external typology-based classification regarding the different existing types of robots themselves, according to their components and/or functionality. Humanoid robots tend to be the most popular utensils with younger students, but robots designed based on the shape of a given structure or animal, commonly modularly built, are also quite intuitive and popular. Nevertheless, floor robots, with built-in wheels that move at ground-level in the absence of leg-like actuators, are usually the preferred model when it comes to Early Childhood Education (Kalaitzidou and Pachidis, 2023).

Another interesting robotic modality can be found in telepresence robots, which, according to Leoste et al. (2022), act as intermediate agents, giving a distant human user a robotic physical manifestation in any given context. They are especially popular in teleteaching situations. Although barely implemented in educational settings, synergetic robots, frequently shaped like hive-mind animals such as swarm robots, can foster quite interesting educational experiences.

Similarly, the possibilities now opened by the recent unprecedented development of software based on Artificial Intelligence may lead to the creation of robotic assistants that are able to adapt to students' needs in real time. This would, therefore, alleviate the workload of human educators (Evripidou et al., 2020), who could act as supervisors and devote themselves to activities that require socioemotional proximity or cognitive processes of a higher order.

1.2 Points of interest regarding the improvement of learning and skill acquisition derived from ER implementation in educational settings

ER, according to Qu and Fok (2022), provides pupils with an actual physical intermediate entity between the analogical reality of the interdisciplinary learning skills, knowledge, and attitudes linked to any given subject and the abstract world of programming and digital software. Therefore, utilizing robots in educational environments is regarded as one of the main ways to achieve a true global and meaningful development of key competences and skills without having to use digital media or expensive and complex software, which are inadequate to the cognitive traits of target students (Moreno-Palma et al., 2024). Following Yang (2024), ER has been proven to be more effective when it comes to mastering both coding and Computational Thinking (CT) than tablet or mobile-based applications. This gives a sense of utility and meaning to the programmed actions.

Despite that, and naturally, Alonso-García et al. (2024) show that this approach has the greatest benefits in curricular areas linked to the design and application of somewhat complex logical and/or cognitive strings, highlighting its effectiveness on the learning of linguistics and mathematics. Interestingly, and establishing a connection between the contemporaneity of these elements and the need for an update regarding the training and development of future generations, Coşkun and Filiz (2023) indicate that using robots is able to favor the development of the so-called XXI Century skills, including flexibility, autonomy, cognitive dexterity, and innovation. This in turn catalyzes the effect of robotics on learners, achieving greater learning benefits.

As Avsec et al. (2016) point out, the introduction of ER into educational programs can be a remarkable emotional learning catalyst, given the innate attraction of learners toward interaction with robotic products, especially at younger ages. According to an intervention with Primary Education learners reported by Fung et al. (2024), the introduction of physical robots in classrooms can further boost the three main categories of learner engagement and intrinsic motivation. Such findings hold great relevance among practitioners, as the social and interactive capabilities of robots are key to enhancing learner motivation. They also hold potential for social skill development and skill generalization with individuals with lower socialization development, such as children diagnosed with Autism Spectrum Disorders (ASD) (Konishi et al., 2024).

Regarding the use of ER in order to transform educational proposals into actions toward inclusion and equality, in an intervention carried out with trainee female Primary Education teachers, Romero-Rodríguez et al. (2023) found out that working with robotic LEGO kits has been effective regarding the transfiguration of negative self-efficacy toward Science, Technology, Engineering, and Mathematics (STEM), which, according to the study conducted by Jung and Lee (2022), is crucial regarding the integration and participation of female employees in peer groups within STEM-related occupations. Additionally, and as Papert (1993) stated, "the construction that takes place in the head often happens especially felicitously when it is supported by construction of a more public sort in the world… the product can be shown, discussed, examined, probed, and admired" (p. 142). Therefore, this combination of meddling with the design and practice testing of ER, while at the same time actually constructing and building the robotic parts by themselves, will surely catalyze the pupils' integration of concepts, skills, and attitudes within their pre-established cognitive schemes, creating interdisciplinary webs of knowledge right from direct experience.

1.3 Risks and challenges of ER in relation to teacher training and practice

Opposing all these benefits, however, Seckel et al. (2022) argue that an optimal introduction of ER requires heavy and difficult teacher training in order to fully grasp the needs and characteristics of a constructivist educational approach involving such a unique and revolutionary kind of resource. This factor can be extremely relevant when it comes to managing and controlling systems of multiple interconnected robots, as failing to properly do so can turn a potentially perception-changing approach into a disappointing and anxiety-provoking ordeal (Bravo et al., 2021). As commented by Andrée et al. (2024), it is quite common for pupils to feel *robot anxiety* when certain parts, mechanisms, or functions, especially the mobile ones, do not operate as required and/or expected.

Failing to acquire the practice and conceptual aspects linked to the educational application of robotics, as Zhao et al. (2024) comment, may lead to learners not mustering nearly enough attention nor effort directed toward the activities at hand, therefore exclusively focusing on the more playful dimension of ER, in detriment of its teaching-learning potential. Related to this potential misleading tendency, Xie et al. (2024) have reported that an excessive expression of humor and/or informal interactive formulas by chatbots may lead to a loss of focus on the task or necessity that initiated the interaction in the first place. This is similar to how background noise and inadequate expression cause a negative and unpleasant effect on the participants of the conversational exchange even in human-only situations (Pourfannan et al., 2024).

It is undoubtedly true that a fully-grown positive attitude toward robotics in classrooms remains to be developed among both teachers and learners alike. This is especially true regarding older generations of each social group, who especially dislike robots that possess either human-like expressive capabilities or appearance (Zhang et al., 2023). On the other hand, conversely, this seems to attract the majority of younger students, who find it an essential trait for robots to have in order to be deemed interesting and appealing (Istenic et al., 2021).

1.4 Previous ER-based experiences in primary and early childhood education

ER, as described by Wang et al. (2024), is the most prevalent tool implemented in Early Childhood Education and the first years of Primary Education when it comes to addressing STEM-related skills, concepts, and attitudes. This could potentially point to a diversity of empirical studies involving the matter. Taking the study developed by Qu and Fok (2022) as a framework of reference, educational interventions based on ER can be designed under three main schemes of action, including programming and computing, non-participating observation, and direct interaction between learners and robots.

As an ideal example of the first and second approaches, Bravo et al. (2021) developed a proposal based on programming line-following robots as if they were characters in a theater play, with groups of Colombian teenagers between 11 and 13 years old in charge of them as directors. They observed and overall had a higher engagement with the subject. Using a sample of students between 5 and 8 years old, Liu et al. (2023) tested a modularly crafted robot that needed to be tested in a simulated environment before using it underwater in order to challenge the learners in the creation of a structure that could withstand the conditions of the target environment. Learners showed high interest in phenomena related to water, such as buoyancy, through the *learn by making* approach.

Finally, the majority of the educational proposals associated with one-on-one student-robot interactions tend to focus on improving the sociolinguistic skills and capabilities of functionally and/or linguistically diverse learners. In the study presented by Lorenzo et al. (2024), four Spanish children diagnosed with conditions linked to ASD were able to reinforce their social proactivity and engagement by playing different interactional games with humanoid robots in non-stressful environments. This significantly enhanced their attention span in contrast to previous iterations with human participants. Some researchers have even attempted to consolidate the figure of ER as a permanent and close companion for students who may feel stressed or anxious in determined social situations, designing pocket-size relieving utilities based on this very same principle (Frederiksen et al., 2024).

However, actions regarding the use of these tools aimed at their implementation by educators employed in both of these stages should not, nonetheless, go unnoticed. An intervention developed by Sun and Liu (2024) involving 56 science teachers in Primary Education demonstrated that attitudes toward programming activities significantly affect their development of CT when acting as receivers of activities involving ER. As a result, one could only assume that fomenting the training and preparation of practitioners regarding proposals of this kind is a key step in order to consolidate the implementation of these educational resources in classrooms based on an actual compromise and believe in the effectiveness of the approach itself.

1.5 Competence concept and the STEM framework

In a society where conceptual and theoretical learning alone are not able to aid students in solving real-world issues anymore, the idea of configuring a new educational paradigm based on developing various and interdisciplinary competences, i.e., being able to apply, in any given context, conceptual knowledge and practical skills, all mediated by psychosocial factors such as attitude, beliefs, and motivation, in order to successfully carry out any given action or solve a problematic situation (Vitello et al., 2021), has been established and expanded over the last few lustrums.

Among these, and given their undeniable linkage and connection, developing competences linked to STEM areas tend to appear and be addressed as a linked package that can act as the base and soil for numerous didactic proposals to grow within an interdisciplinary and global curricular scope. Recent conceptual debates have occasionally led to even questioning the existence of the limited STEM field itself, having attempted to reshape the concept in order to integrate the Arts, STEAM, or even Reading and Writing, STREAM (Raycheva, 2024). Establishing limits regarding this set of abilities has experienced its own barriers and controversial standpoints as time and educational sciences have gone by. In this particular scenario, a vast proportion of researchers tend to be undecisive about the role of Computational Thinking (CT), a skill closely connected to programming and robotics, within the STEM competence framework.

Even though its main dimensions, stated by Bento et al. (2024) to be "(i) identification and decomposition, (ii) abstraction, (iii) pattern recognition, (iv) design of algorithms, and (v) testing and debugging" (p. 6), are deeply connected to the addressed subject areas in STEM, especially with Technology and Mathematics, authors like Sung et al. (2017) affirm that CT should necessarily be regarded as an interdisciplinary skill, subject to be used and developed in the didactic planning of any curricular area at hand. However, as these authors point out, it is undeniable that lessons mainly devoted to developing STEM aspects are the main and most reliable context for favoring the acquisition of such abilities. Similarly, Misirli and Komis (2023) describe CT as an extension of algorithmic thinking whose development plays a pivotal role in the effective and efficient management of any given kind of situation, when applied to both STEM-related areas and knowledge fields that are external to them.

For the purposes of this study, CT will be considered a skill outside of the STEM competence framework. However, given the obvious connection between both entities, such as the well-established significant link between learner attitude toward STEM and the development of CT (Küçükaydın et al., 2024), reports of results regarding this skill or any of its dimensions will be considered related to STEM competences, as developing one necessarily ends up in promoting an improvement in the other.

1.6 The present study: background and objectives

A previous systematic literature review developed by Ortuño and Serrano (2024), centered around the most common practices regarding the addressing of computational thinking in Primary Education, found that ER is one of the most prominent approaches regarding the matter in the aforementioned educational stage. These authors, nevertheless, point out that few registers have been found for its earlier grades, which could lead to a gap in the general use of robotics with the youngest students of Primary Education. This would further confirm the commonly reported lack of quality studies focused on ER-related skills such as CT in stages different from higher education, as pointed out by Misirli and Komis (2023).

A review of the exact same topic and studying experiences developed both in Primary and Early Childhood Education developed by Ching and Hsu (2024) confirms the previously mentioned results and details that there is a relatively stable balance regarding the time extent of the analyzed interventions. Half of the addressed sample in the aforementioned study lasts for about 10 h, while the other half tends to accommodate programs for more than 11 h. It was observed that *LEGO Mindstorms* robotic kits were the predominant humanoid ER resource, whereas *BeeBot* and *KIBO* were the sole representatives of the floor robot category.

In order to fit the main research topic of the present study, Darmawansah et al. (2023) developed a literature review on the use of ER in the development of STEM-based courses across all educational stages, determining that Primary Education, with significant differences, is the stage where ER has been implemented and researched the most. They also describe a general rising tendency in the publishing volume of papers addressing the matter over the last decade.

A review on the use of ER regarding the development of STEM competences in Primary and Early Childhood Education developed by Tselegkaridis and Sapounidis (2022) unveiled that, contrary to Ortuño and Serrano's (2024) findings, the majority of robotics-based educational interventions, up to 2020, have been focused around the first grades of the former stage, following an international distribution structured in a south-European group and a core located in the United States of America as an only country publishing block.

An updated version of this last review undergone by the same authorial team (Sapounidis et al., 2024) states that more than three quarters of the studied sample had been published during the last 2 years of the work's temporal span, comprising between 2016 and 2021. This indicates an exponential increase in the publication volume regarding the subject matter during the last lustrum and supports the publishing tendency claims stated in Darmawansah et al. (2023).

Given the ample relevance of ER in Primary Education and the lack of studies that solely focus on the needs and traits of the student body at this stage, as well as the reported ever-increasing publishing tendency involving robotics-based STEM education, the main goal of the present study is none other than to analyze the state of integration of ER-based programs in the Primary Education stage, determining and critically appraising the existent evidence in relation to their effects on the students' learning results regarding the development of STEM competences. As to structuring the research process, several Research Questions (RQ) were posed:

- RQ1. What are the main editorial and contextual features of the analyzed studies regarding their geographical and temporal distribution, type, and language of publication?
- RQ2. How were the selected interventions designed in terms of target sample characteristics, including grade, sex, and age?
- RQ3. Which study design characteristics, including research methodology and design, sample selection, group configuration, statistical, and standardized tests, were introduced in the reviewed experiences?
- RQ4. What didactic aspects, including strategies, resources, and activities, were implemented during the conduct of the examined programs?

- RQ5. What are the reported results of the addressed experiences regarding the development of STEM competencies through ER-based interventions?

2 Methodology

In order to properly answer the previously detailed research questions and, by extension, fulfill the goal of the present research, a systematic review of the existing literature regarding the development of STEM competences through ER in Primary Education was developed. As Higgins and Green (2008) point out, this methodology entails the compilation, collation, and subsequent analysis, organized and managed based on pre-established objectives, of the existing evidence regarding a particular subject matter, following a protocol that guarantees the rigor, fidelity, validity, and replicability of the detailed research process.

Additionally, a meta-analysis was conducted based on all the effect sizes reported in the selected areas of the final study sample. This research methodology, which is based on the statistical combination of results originated in multiple publications, aims at precisely solving conflicting aspects and situations that cannot possibly be resolved by individual studies (Higgins and Green, 2008).

The combination of both of these research approaches aims to produce a synthesis of quality and rigor from which conclusions of interest and relevance to the target research field may be retrieved, avoiding any kind of bias on the part of the researchers throughout the process. Furthermore, in order to assure the replicability of the research process presented in this paper, this systematic review of the literature complies with the guidelines established in the PRISMA 2020, *Preferred Reporting Items for Systematic Reviews and Meta-Analyses, 2020 edition* statement (Page et al., 2021), an update to the original, worldwide, and acknowledged multidisciplinary framework of reference regarding review quality (Moher et al., 2009).

Therefore, an initial planification stage took place, in which the main objective for this research as well as specific research questions were established, determining the databases that would be consulted and the specific search strategy that was to be applied during the course of the research as well as the inclusion and exclusion criteria that would be used to configure the studies to analyze. After that, an action stage was developed, starting with a preliminary search in the selected databases and proceeding to filter and refine the retrieved registers through the application of the aforementioned criteria, to determine the final study sample that supports the present review.

2.1 Search strategy

The search revolved around the Scopus and Web of Science, hereinafter WoS, databases, given their ample acknowledgment in the Social Sciences research field and the quality of their indexed research papers, determined through the Journal Citation Reports (JCR) and Scimago Journal & Country Rank (SJR) indicators.

The definitive search string that was to be used was based around the key terms Robotics, Primary Education, and STEM, determining the educational resource of interest to this research, the stage where experiences are to be analyzed, and the competence framework to be developed through the indicated methodological approach.

In order to configure a search string capable of reaching every register related to the topic of interest to the present study, several thesauruses were consulted, including the Educational Resources Information Center (ERIC), United Nations Educational, Scientific, and Cultural Organization (UNESCO), and European Education ones, retrieving synonyms and related terms linked to every aforementioned key term as well as the term intervention, which was implemented in order to refine the retrieved results and avoid the integration of non-empirically elaborated works. It is worth mentioning that the term Robotics, given its already specific and limited range, found no registered or associated synonyms in the reviewed thesauruses.

The search string implemented in the present research can be divided into four blocks, including the one linked to the independent variable of the present review, robotics, and descriptors regarding the target educational stage, Primary Education. Descriptors regarding the wanted methodological typology, empirical studies based on interventions, and finally the dependent variable, STEM competencies.

Therefore, the search string to be used in this systematic literature review is: (Robotics OR Robots) AND (intervention* OR program* OR practice* OR train* OR initiative* OR action* OR project*) AND (STEM OR "Science, Technology, Engineering, and Mathematics" OR Science OR Technology OR Engineering OR Mathematics) AND ("primary education" OR "primary school" OR "elementary education" OR "elementary school" OR "preadolescents" OR "middle school" OR K-12 OR first-grad* OR second-grad* OR third-grad* OR fourth-grad* OR fifth-grad* OR sixth-grad*).

With the goal of further limiting and narrowing down the retrieved registers so as to build up a rigorous and specific sample of the highest possible quality, several inclusion criteria along with their counterparts in the form of exclusion criteria were established. These criteria can be retrieved from Table 1, along with their corresponding reason for implementation present in Table 2.

2.2 Data extraction and analysis

The information compilation and screening process took place during July 2024, being structured in four different phases: preliminary search, accompanied by elimination of duplicates; first level of screening, by reading the title and abstract of retrieved studies; second level of screening, including retrieval of remaining registers as well as exclusion based on full text reading in case registers could be retrieved; and, finally, selection of remaining registers to be included in the systematic review of the literature, as well as assessing the eligibility of said registers in order to be included in the subsequent meta-analysis.

During the initial phase, or *identification* phase, the previously established search string was implemented in the selected databases in order to retrieve an initial study pool of $n = 1,712$ studies, out of which 425 were deemed to be duplicates. It is also worth mentioning that bibliographical registers referring to whole books, including chapters that had already been considered within the preliminary study sample ($n = 71$), were also considered as duplicate registers as well.

TABLE 1 Inclusion and exclusion criteria.

Inclusion criteria (IN)	Exclusion criteria (EX)
IN1. Language must be either English or Spanish	EX1: Language different from English and Spanish
IN2. Empirical, quantitative, or mixed-methods studies	EX2. Empirical qualitative or theoretical studies
IN3. Sample must be students between 6 and 12 years old, both included	EX3. Sample including teachers, younger than 6 or older than 12 years old, or sample of unspecified age
IN4. ER experiences	EX4. Experiences either not implementing ER or in which ER is not the main approach
IN5. Experiences aimed at developing STEM competences	EX5. Experiences not aiming to develop STEM competencies or in which doing so is not the main objective
IN6. Studies must be based on intervention programs	EX6. Studies are not based on programs, or the intervention of the program is not specified
IN7. Studies must include comparable control and experimental groups based on the introduction, or not, of ER	EX7. Studies either not including control groups or in which both control and experimental groups use ER
IN8. The study fulfills basic quality criteria	EX8. The study fails to fulfill basic quality criteria

Authors' own work.

The second phase, or *screening*, phase was divided into two different instances. In order to avoid any potential kind of researcher bias in the evaluation of the registers at hand, all researchers appearing as authors of the present manuscript individually analyzed every single one of the remaining studies after the elimination of duplicates ($n = 1,216$) based on reading their title and abstract.

In order to guarantee the efficiency and efficacy of this evaluation process, the authors waited until every researcher had finished analyzing the preliminary study pool in order to share their individual results and present any particular concerns or ambiguous results. The inclusion and exclusion criteria that were to be applied during this initial reviewing instance involved all those that could be assessed without needing to access the full text, that is, IN2-EX2, IN3-EX3, IN4-EX4, and IN5-EX5. Nevertheless, these criteria would be fully implemented in the second screening instance once again, as it was likely that abstracts and titles would potentially be misleading regarding the content of each actual paper.

The authors agreed on the vast majority of the decisions taken in relation to the rejection or acceptance for further consideration of the studies at hand, solving disputes through common debate and argument in relation to 12 registers in which the researchers initially disagreed about their inclusion or not. As to guarantee the quality of the present research paper, the 12 studies for which an initial agreement was not reached were left for evaluation during the next screening instance.

After establishing an initial pool of accepted studies, including those whose authors had not previously agreed, each author tried to retrieve the full text of the remaining registers in order to analyze them in whole, applying both the previously mentioned criteria as well as those linked to publishing language (IN1-EX1) and academic quality (IN8-EX8). At the conclusion of this phase, studies whose full text could not be obtained were automatically discarded. An additional meeting regarding the definitive acceptance or rejection of the remaining studies was held place by the end of this screening instance, during which the authors shared the results of their eligibility assessment process.

Although there was complete agreement regarding the application of most exclusion criteria and the derived results, some researchers shared different opinions about the quality assessment of the final study sample. In order to resolve these disputes, a fifth subject, belonging to the research group of the aforementioned authors, participated as a mediator and external evaluator during an additional gathering of the authors.

This situation involved two studies that were lacking critical information, as the data provided were sufficient to accept their inclusion within the application of the established criteria, yet characteristically lacking to answer several of the established RQ.

Finally, after having reached a consensus regarding the exclusion of these two studies that had returned different quality evaluation results, the final study sample for the present systematic literature review was configured ($n = 13$), having to exclude $n = 5$ studies from it in order to conform the sample to be used in the meta-analysis ($n = 8$) as these registers were missing full outcome data.

Once both final study samples had been configured and given the objective nature of the process, the first two authors of this study proceeded to extract the requested information from each study, which was then given to the third and fourth authors in order to prepare them for presentation and visualization. The final results were then reviewed by the aforementioned external researcher.

The results derived from conducting the quality assessment linked to IN8-EX8, including the registers that were not considered for the systematic literature review as well as those excluded from the meta-analysis, can be accessed in Table 3.

3 Results

A flow diagram representing every phase of the data collection process, including the number of registers included or excluded at each instance of the process, can be seen in Figure 1. As previously outlined, and in order to fully comply with essential quality guidelines established in Page et al. (2021), every author, as well as an additional support researcher, was directly involved in the screening and selection processes in order to mitigate any potential risk of researcher bias. Additionally, and with the goal of avoiding any kind of subjective influence emanating from the agreement of the authorial team, every included study was evaluated under a critical appraisal tool, which led to the exclusion of two additional studies and the objective configuration of the two final study samples for both the systematic literature review and the subsequent meta-analysis.

In order to adequately answer and tackle the previously established research questions, the present epigraph will be structured into several subsections regarding each and every one of the research subgoals. The graphics presented in this epigraph were created using the software Tableau (Tableau Software, Seattle, WA, United States, version 24.1.1062) and OriginLab (OriginLab Corp., Northampton, MA, USA, version 10.15.132). The meta-analysis was carried out in the Jamovi Statistical package (The Jamovi Project, Sydney, NSW, Australia, version 2.5).

TABLE 2 Justification for each inclusion–exclusion criterion pair.

Criteria	Reason for implementation
IN1–EX1	English and Spanish, as some of the most widely spoken languages worldwide, are essential languages when it comes to retrieving both the highest possible number of registers as well as results of the highest quality.
IN2–EX2	One of the research questions of the present study involves determining the results retrieved during the review through a meta-analysis, which necessarily requires quantitative data. Every study in which it is not the main outcome measure is not suitable for inclusion.
IN3–EX3	Primary Education is an educational stage that usually starts "not below 5 years old nor above 7 years old […] typically lasts six years" (UNESCO Institute for Statistics, 2012, p. 30). Therefore, in an attempt to configure a study sample with international relevance, the starting age shall be deemed six, as the exact middle point between five and seven, culminating at twelve, with a difference of six years.
IN4–EX4	No study in which the final outcomes are not directly caused, mainly influenced, or determined by the use of robotics for educational purposes is of interest for the present review.
IN5–EX5	No study in which the introduction of ER directly or indirectly relates to the development, or not, of STEM competences, understanding the ample width of the field as previously stated, is of interest for the present review.
IN6–EX6	According to the UNESCO Institute for Statistics (2012), a program is a preplanned intervention, ordered under a given methodological approach, translated into certain didactic strategies, and developed into specific activities that are extended over a period of time. Therefore, only interventions lasting for a duration exceeding two lessons, with at least a session between presentation and finalization, will be considered as programs.
IN7–EX7	A comparison group is fundamental in order to actually determine if the results of any given educational intervention actually derive from the resources, techniques, methodologies, and/or strategies that were implemented (Hunter et al., 2014). Consequently, only studies in which the control group has adhered to a traditional, i.e., non-active receptive-expositive, methodological approach will be included to ensure a proper comparison of treatment outcomes.
IN8–EX8	Preselected studies will undergo assessment under the *Mixed Methods Appraisal Tool* (Hong et al., 2018), one of the most commonly implemented interdisciplinary quality instruments in educational research. Studies obtaining a minimum score of 4 out of 5 when answering the essential screening questions will be considered for review. Studies obtaining a perfect score, including all outcome measures, will be included in the meta-analysis.

Authors' own work.

3.1 What are the main editorial and contextual features of the analyzed studies regarding their geographical and temporal distribution and type and language of publication?

The final 13 included studies correspond to eight research articles and five conference proceedings, published within the temporal threshold established between the years 2011 and 2023. As it can be seen in Figure 2, the year with the highest publishing concentrations was 2023, with a total of three registers that were included within the final study sample. The rest of the addressed, namely 2011, 2012, 2013, 2020, and 2022, contain a single study per year, whereas 2019 and 2021 are associated with the publication of two different registers each.

It is, however, worth noting that there is a characteristic hiatus established between the years 2014 and 2018, both included, in which no new study that was included in the final study sample of the present study was published. The majority of the registers were written in English ($n = 11$), while only 2 studies used Spanish as their publication language.

This, as shown in Figure 3, serves as a representation of the international existing interest in the subject matter at hand, as three publishing blocks can be established when examining the final study sample under a geographical scope. The most prominent one, namely the European, pertains to experiences developed in Spain ($n = 5$), the Netherlands ($n = 1$), Denmark ($n = 1$), Greece ($n = 1$), and Italy ($n = 1$), whereas the United States of America acts as an only-country publishing block with two included registers. Finally, it is worth noting that Taiwan was the only Asian country to be included in the final study sample, offering a total of two works.

3.2 How were the selected interventions designed in matters of target sample characteristics, including grade, sex, and age?

Examining the final study sample as a whole, a total of $n = 1,082$ students enrolled in the analyzed ER intervention programs, therefore resulting in a mean of 83.23 learners per accepted study. The sample sizes of each individual study ranged between 26 as the minimum mark and a maximum of 260 learners.

Regarding the characteristics of the participant sample, the ratio of boys and girls was fairly balanced, as the total number of involved males escalates up to 466 students (43.07%), while a total of 418 female students (38.63%) were involved in the total sum of the included studies. It is worth noting, nonetheless, that four registers did not specify the sex ratio existing within their target samples, therefore being unable to determine the belonging to one group or the other of 198 learners (18.30%).

The present review addressed results derived from a vast age range of learners (M = 10.37; SD = 2.89), including all six grades of Primary Education as previously described and limited. As shown in Figure 4, the experiences mainly revolved around the third cycle of Primary Education, including students between 10 and 12 years old ($n = 594$; 54.89%). Nevertheless, a significant proportion of students present in the reviewed interventions were still assigned to one of the grades included within the second cycle, fourth grade, involving an amount of 288 learners (26.62%). The first cycle, including learners between 6 and 7 years old, was the one participating the least in the final study sample, with only 14.23% of the sample ($n = 154$).

10.3389/feduc.2024.1480908

TABLE 3 Results of the quality assessment process (IN8-EX8).

Work	Method	S1	S2	1	2	3	4	5	Included in
Caballero-González and García-Valcárcel (2020)	Mixed methods	Y	Y	Y	Y	Y	Y	Y	Meta-analysis
Casad and Jawaharlal (2012)	Quantitative non-randomized	Y	Y	Y	Y	Y	Y	Y	Meta-analysis
Chiazzese et al. (2019)	Quantitative non-randomized	Y	Y	Y	Y	Y	Y	Y	Meta-analysis
Chou (2018)	Quantitative randomized control trials	Y	Y	Y	Y	N	Y	Y	Literature review
Finsterbach et al. (2023)	Quantitative non-randomized	Y	Y	Y	Y	N	Y	Y	Literature review
Ferrada et al. (2023)	Quantitative non-randomized	Y	Y	Y	Y	Y	Y	Y	Meta-analysis
Huang et al. (2013)	Quantitative randomized control trials	Y	Y	Y	Y	N	Y	Y	Literature review
Merino-Armero et al. (2022)	Quantitative non-randomized	Y	Y	Y	Y	Y	Y	Y	Meta-analysis
Meza et al. (2012)	Quantitative non-randomized	N	N	Y	Y	N	N	Y	Excluded
Ortiz (2011)	Quantitative randomized control trials	Y	Y	Y	Y	N	Y	Y	Literature review
Papadakis et al. (2024)	Quantitative non-randomized	Y	Y	Y	Y	N	Y	Y	Literature review
Sáez-López et al. (2021)	Quantitative non-randomized	Y	Y	Y	Y	Y	Y	Y	Meta-analysis
Ponce et al. (2019)	Quantitative randomized control trials	Y	Y	Y	Y	Y	Y	Y	Meta-analysis
Sáez-López et al. (2019)	Quantitative non-randomized	N	N	CT	CT	N	N	Y	Excluded
Smakman et al. (2021)	Quantitative non-randomized	Y	Y	Y	Y	Y	Y	N	Literature review

Authors' own work, based on the template and tool provided by Hong et al. (2018). S1 and S2 stand for the screening questions of the tool. C1, C2, C3, C4, and C5 refer to the applied criteria, varying based on the method of the referred works. Y, Yes; N, No; CT, Cannot Tell.

3.3 What study design characteristics, including research methodology and design, sample selection, group configuration, statistical and standardized tests, were introduced in the reviewed experiences?

All of the included studies selected their research sample following non-probabilistic methods; however, the majority of them followed non-random sampling based on convenience (*n* = 9; 69.23%), while the rest opted for purposive sampling (*n* = 4; 30.77%). Now, regarding the followed research methodology, 46.15% of the included studies opted for utilizing a mixed methods research approach (*n* = 6; 46.15%), being the rest of the sample developed under quantitative approaches (*n* = 7; 53.85%). It is, however, worth noting that only three studies (23.08%) opted for a true experimental design, configuring the divide between control and experimental groups at random, being the rest of the included studies conceived under the quasiexperimental design approach.

There is some diversity in regard to the statistical tests that were conducted in order to analyze the retrieved data within the accepted and reviewed studies, being t-tests the most common kind (*n* = 8; 30.77%), followed by one-way Analysis of Variance (ANOVA; *n* = 4; 15.38%), Mann–Whitney U test (*n* = 11.54%), Kolmogorov–Smirnov's test, Levene's test, Shapiro–Wilk's test, Pearson's Chi square test (*n* = 2; 7.69%), Wilcoxon's signed rank test, ordinal regression, and Multivariate Analysis of Covariance (MANCOVA; *n* = 1; 3.85%).

This, nonetheless, was the opposite case to what could be seen regarding the use of standardized tests in order to assess the developed competences throughout the experiences, as majority instruments were of an *ad-hoc* design, including properly validated and designed tests (*n* = 5; 27.78%) and *ad-hoc* rubrics and other kinds of exams (*n* = 6; 33.33%). It is worth noting, however, that some instruments, though initially externally validated, were adapted and modified regarding their usability in the reviewed experiences (*n* = 2; 11.11%).

Finally, the implemented standardized tests that were used in the accepted studies include the Bareka test, the Test of Visual Blocks and Robotics (TVBR), Three-Dimensions of Students Attitude Towards Science (TDSAS), Escala de Actitud hacia las Matemáticas (Attitude towards Mathematics Scale, EAM), and Matematikprofilen (Mathematic Profile).

3.4 What didactic aspects, including strategies, resources, and activities, were implemented during the conduct of the studied programs?

In order to present the characteristics of each program along with their most prominent activities and strategies, Table 4 summarizes the design bases of the interventions reviewed as part of the final study sample.

In observance of these results, it is quite worth noting that Mathematics appears to be the most frequently addressed area in the reviewed studies (*n* = 10), followed by Technology (*n* = 6) and Science (*n* = 4), while quite distant from Engineering (*n* = 3). A diagram showcasing the connection between these elements can be seen in Figure 5.

3.5 What are the reported results of the addressed experiences in relation to the development of STEM competences through ER-based interventions?

To guarantee a non-biased visualization of the retrieved results of the selected ER-based STEM interventions, a meta-analysis using the Standardized Mean Difference (SMD) between the control and experimental groups as the outcome measure was performed.

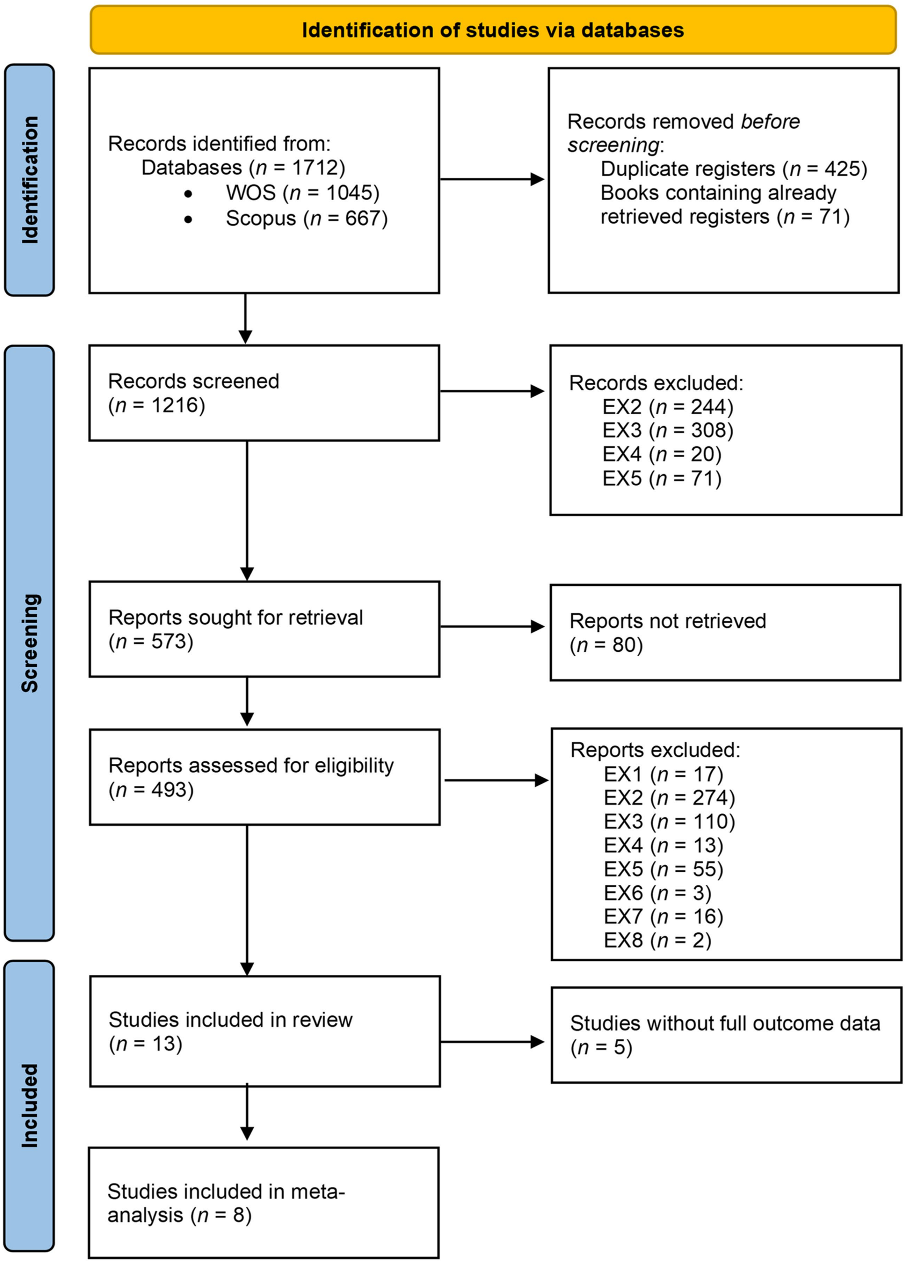

FIGURE 1
Flowchart illustrating the filtering process. Authors' own work, based on Page et al. (2021).

3.5.1 Assessment of publication bias and detection of potential outliers

Given that the majority of the studies, as previously stated, implemented different evaluation procedures, most of them of an *ad hoc* nature, a Random-Effects model was used in order to conduct a heterogeneity analysis on the results retrieved from the final study sample that was included in the meta-analysis. Provided that six different null hypotheses were to be tested in the present epigraph, a Bonferroni-corrected two-sided significance value of $\alpha = 0.008$ was established in order to avoid committing Type I errors in the present work.

As shown in Table 5, including the results of the applied Q-test (Cochran, 1954) and related statistics, given that the Tau^2 value returned positive results, it can be affirmed that a certain degree of heterogeneity is present within the study sample. Nevertheless, it is worth noting that the Q statistic retrieved positive yet relatively low values, which, although effectively points out the existence of heterogeneity between the introduced registers, may be influenced by the relatively low number of registers ($k = 18$, coming from nine different publications) that were addressed in this meta-analysis. This may leave the Q statistic with a relatively low statistical power to establish such heterogeneity.

Regarding the origin of the existing heterogeneity, the I^2 statistic indicates that 70.28% of it is due to the naturally existing heterogeneity between the included studies. There was no need to calculate the value of the R^2 statistic given that no mediators were considered within this

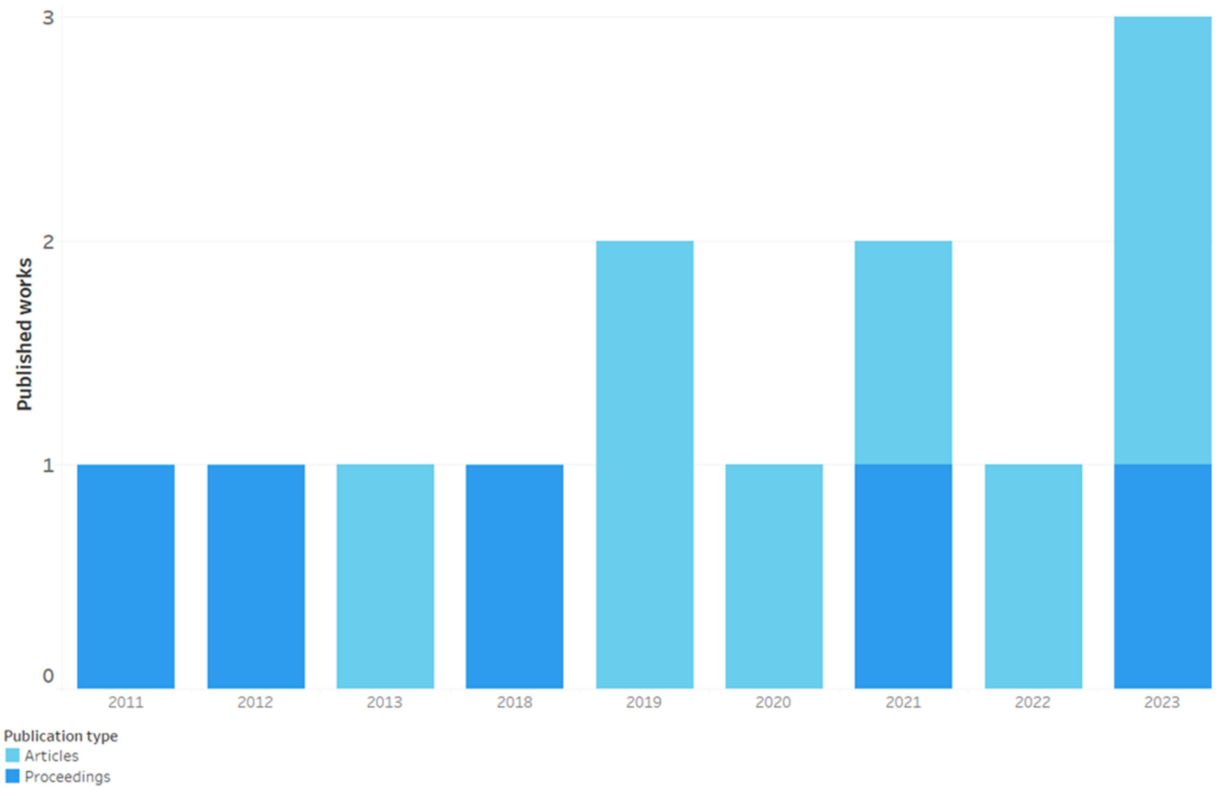

FIGURE 2
Publication of selected studies throughout the years.

meta-analysis model. Consequently, the resulting 28.72% of study heterogeneity could be simply attributed to chance; however, it is worth noting that the studied effect sizes measure different competencies included within the STEM framework, which could explain why such a high percentage of the existing differences between studies cannot be immediately explained by following different experimental conditions or methodological design across retrieved registers.

These conclusions are supported up by a value higher than one for the H^2 statistic, which confirms the absence of perfect homogeneity between the included effect sizes. Additionally, a value of Tau that is closer to zero than to one or minus one, therefore not showing any trace of strong correlation between the presented results.

Once the registers included in the conducted meta-analysis were deemed to present significantly different effect sizes, i.e., results, an evaluation of potential outliers can be conducted within the established model. In order to do so, studentized residuals and Cook's distances were the residual-based measures that were chosen as the main parameters of reference. Every study that was reported to surpass specifically established thresholds, these being higher than the $100 \times (1–0.05) / (2 \times k)$th percentile of a normal distribution, being k the number of analyzed registers for studentized residuals, and the median plus six times the interquartile range for Cook's distances, would be established as potential outliers. No registers were shown to return studentized residuals larger than ±2.9913, which, along with the fact that studying Cook's distances did not highlight any register as overly influential, leads to affirming that no outliers were present in the model.

Finally, regarding the asymmetry of the funnel plot established based on the SMD and SE of each register, both Begg and Mazumdar's (1994) rank correlation test (tau = −0.124; $p = 0.501$) and Egger's regression test (Egger et al., 1997) (intercept = −1.124; $p = 0.261$) were conducted, both pointing out the overall symmetry of the established funnel plot as they retrieved non-significant results and therefore the absence of publication bias in the present study sample, as visually appreciable in Figure 6. This is further supported by the tolerance value obtained through Rosenthal's Fail-Safe N, which returns a value of N = 572 ($p = < 0.001$), which surpasses the recommended threshold of $5\ k + 10$ (Rosenthal, 1979), equaling 90 in this study. As a result, there was no need to implement the trim and fill method of solving publication bias with regard to the final study sample.

3.5.2 Outcomes derived from the included registers

Proceeding to study the predicted true outcomes of the selected registers, establishing a Confidence Interval (CI) of 95% given the reported existing heterogeneity, it can be affirmed, as shown in Figure 2, that most effect size estimates, around 83% of the final register sample, were of a positive character. Despite this, it is worth mentioning that there were relevant differences regarding the nature of the reported effect sizes, as the lowest one, established in Ferrada et al. (2023), is −0.6264, while the highest one, 1.4009, can be traced back to the study developed by Chiazzese et al. (2019).

In order to fully comprehend the actual effects resulting from the introduction of ER in STEM intervention programs, every effect size related to the subject matter was included in the present meta-analysis,

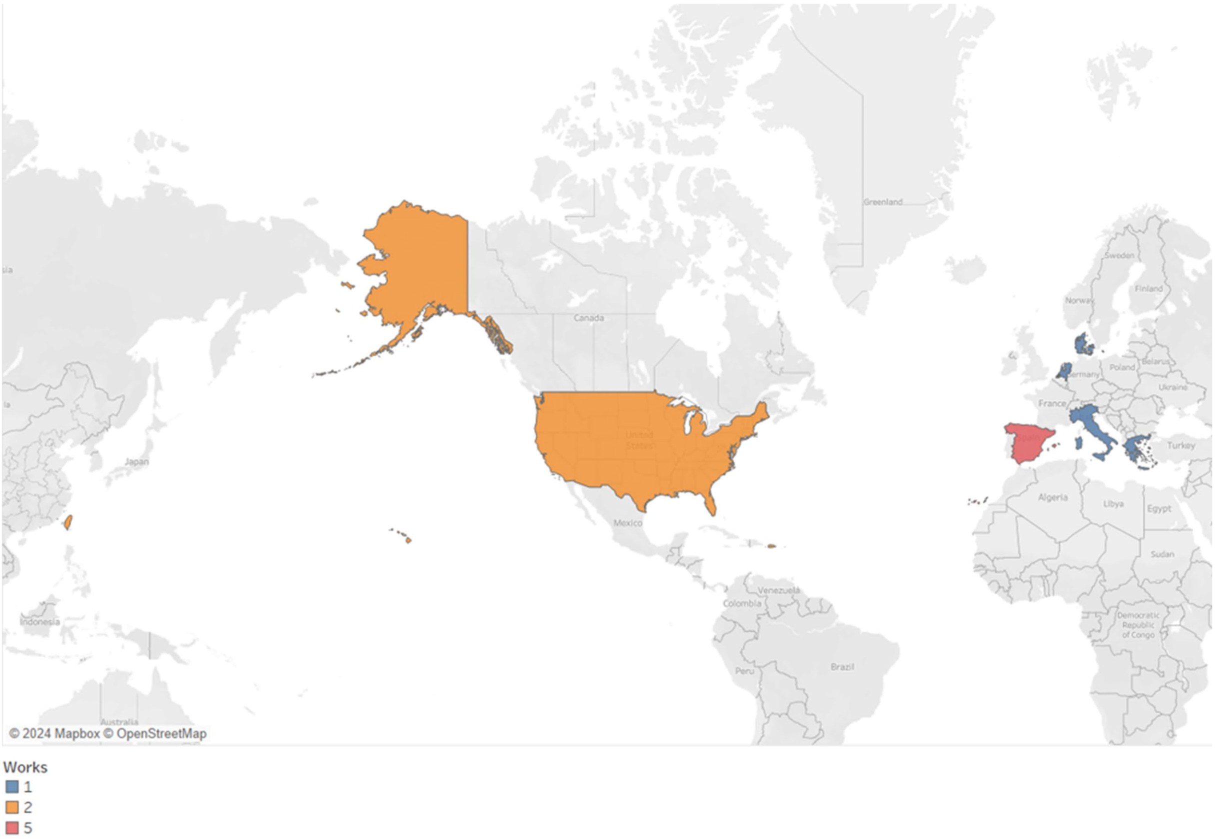

FIGURE 3
Geographical distribution of the final study sample.

which results in certain studies being addressed more than once in the previously mentioned Figure 7.

As shown in Table 6, the estimated pooled effect size of all the included registers, utilizing SMD as the outcome measure, was 0.535 (SE = 0.104; CI: 0.3314 to 0.7393) which, given a significant deviance from zero (z = 5.1453; $p < 0.001$). It is, however, worth noting that the prediction interval established regarding true outcomes (CI: −0.1814 to 1.2521), points out at the possibility of potential negative true outcomes within the contemplated register sample, despite the positive nature of the average outcome as previously addressed. Given that the pooled effect size of the register sample yielded a Cohen's *d* value of 0.54 (CI: 0.33 to 0.74), therefore, the implementation of ER in the development of STEM-related competences returns a medium, positive effect (Chen et al., 2010).

Regarding the validity of the hereby presented results, the conducted meta-analysis, based on $n = 18$ reported effect sizes extracted from the included studies, conforms to the study number standards established by Higgins and Green (2008), who recommend a minimum of 10 works for a meta-analysis approach to be deemed rigorous and potentially representative of the target study population. Similarly, the evaluation of funnel plot asymmetry was developed following the recommendations from Sterne et al.'s (2011) work, who state that in order to properly determine the existence of publication bias, at least 10 registers should be addressed in the overall meta-analysis.

The conduct of the aforementioned forest plot and subsequent determination of the pooled effect size of the addressed final study sample conforms to Jackson and Turner's (2017) general quality guidelines when utilizing a Random-Effects Model approach in order to establish a prediction interval for the true effect size measure of the sample, i.e., inclusion of at least five reported effect sizes using the same measure. It is, however, worth noting that this study additionally complies with the requirements highlighted by Valentine et al. (2010) regarding the conceptualization of the meta-analytical approach to evidence assessment and interpretation, which pinpoints the necessary number of included studies at a minimum of two.

Finally, and in order to fully make sure that the presented analysis holds sufficient statistical power in order to properly and meticulously answer the previously established research questions, a meta-analytical statistical power test was developed using the tool available at Quintana and Tiebel (2018), based on a previously published study by the same authors (Quintana, 2017). Taking into consideration that the present study shows a high level of heterogeneity, i.e., $I^2 > 50\%$, as determined by West et al. (2010), the conducted meta-analysis therefore has a statistical power of 0.9996 at the two-sided $\alpha = 0.05$ level of significance, which allows the established findings and conclusions to be considered of relevance and significance.

4 Discussion

It can be affirmed that the publishing tendency in the field, though rising during the last few years, has seen a relevant hiatus since its birth as a research field within the educational scientific world,

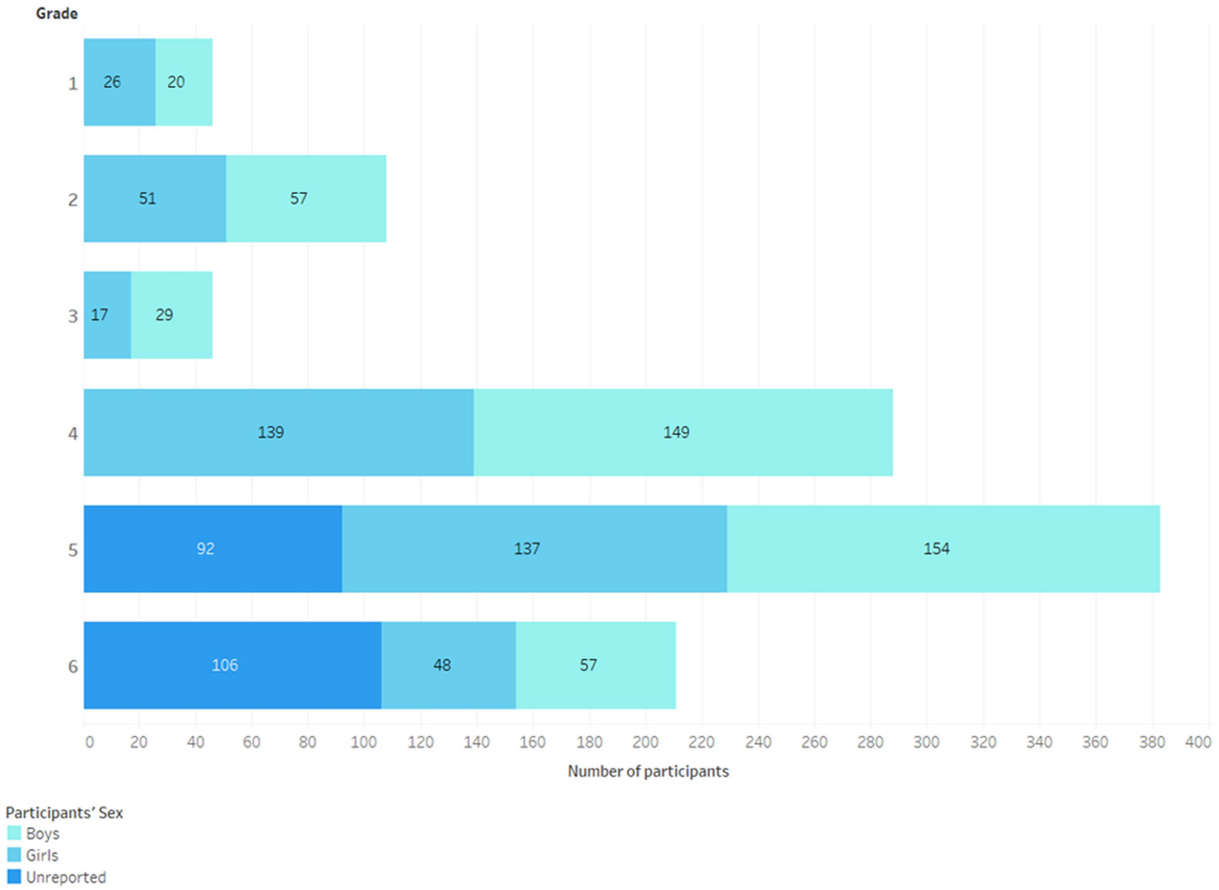

FIGURE 4
Age and grade distribution of the final participant sample.

showing variate publishing rates both in academic events and formal research articles. Within the STEM competence framework, Mathematics, along with Technology, are the most frequently addressed subject areas, in detriment of further strengthening both Science and Engineering Nevertheless, it was still possible to configure a diverse research sample of interest, including both male and female students and representatives of every Primary Education grade, especially focusing on the latter years within the stage. Interestingly enough, it appears that most didactic proposals are either directly designed or adapted by their own conductors, a tendency that has been seen to affect even the use of different evaluation instruments and resources that, even if well-accounted for their quality and relevance, are not observed frequently in multiple educational settings. Finally, an overall positive intervention result, significant yet not extensive, derived from these experiences was reported.

4.1 Summary of main findings

Comparing the results established in this research study with the existing literature, starting with giving meaning to the data linked to RQ1, the most relevant difference regarding previously established results can be found, without a doubt, in the rising publishing tendency in regard to the introduction of ER in STEM-related interventions. As it has been pointed out, both an existing hiatus between the years 2014 and 2018, as well as a lack of robust signs of improvement during the most recent years apart from non-significantly higher numbers, which can be further backed up by the fact that no study published in the present year was accepted in the review, contradict the findings reported in the review conducted by Darmawansah et al. (2023) and Sapounidis et al. (2024), which, simultaneously, appears to negate the apparent irreplaceable character of ER established by Garzón et al. (2021).

Although there is no apparent explanation behind this severe difference between recently conducted reviews, as it was stated, the fact that the aforementioned works focused both on Primary and Early Childhood Education can be of great relevance regarding the matter. As stated by Kalaitzidou and Pachidis (2023), simple floor robots such as the *BeeBot*, *mBot,* or *Ozobot* tend to be especially common in Preschool and not recommendable for older learners.

The fact that all of these appeared in the reviewed experiences, not being only limited to first graders, may uncover a deep need for further training aimed at Primary Education teachers, as reported by Seckel et al. (2022), that capacitates them to use robots adequate for their learners' age, and therefore fomenting the implementation of ER in the stage. Nevertheless, another potential explanation could be found in how the majority of the educators implement *ad-hoc*

TABLE 4 Methodological summary of included experiences.

Program	Duration in weeks	STEM	Resources	Procedure
Caballero-González and García-Valcárcel (2020) *Ad-hoc* program	10	MAT TEC	BeeBot	The second module focused on the use of the robotics resource at hand, dedicating two sessions to working on sequences and algorithmic thinking, two other sessions to developing abstraction patterns, and two final sessions were devoted to depuration. In every module, educational rugs designed to test potential programming sequences on the robot of choice were the primary activity type.
Casad and Jawaharlal (2012) Robotics Education through Active Learning (REAL)	25	MAT ENG	*Ad-hoc*-built robots	The first module lasting two weeks, was based on bringing students closer to robot construction and basic connections between components. The second module lasted two weeks, revolving around designing, building, and experimenting with a three-wheeled robot. The last module, lasting eleven weeks, involved learning about mechanical components and basic engineering aspects in order to build a robot to be presented in a competition.
Chiazzese et al. (2019) *Ad-hoc* program	4	MAT TEC	Lego Education WeDo 2.0 (and its platform)	The program involved four sessions framed within a Project-Based Learning based on a cooperative group methodology in which a gradual approach to robotics and programming was followed. First, students were introduced to the robotic kits at hand and their various hardware parts in order for them to build and program the resulting robots in order to solve various STEM-related challenges, such as representing the life cycle of a frog.
Chou (2018) Robot MakerSpace	16	TEC ENG	Scratch, mBot, mBlock	The intervention can be divided into three different modules. The starter one, involving two weeks, focuses on teaching learners how to program robots using the desired platforms, including practices. The second one, lasting ten weeks, is based on learning how different components, effectors, and actuators of robots look and act, along with gradually building a robot using these resources. Finally, the last four weeks have focused on controlling the robot and engaging in actual competitions with it.
Finsterbach et al. (2023) *Ad-hoc* program	48	MAT	BeeBot, Scratch, Ozobot, Micro:bits, Microsoft MakeCode	The intervention extended over two years, dedicating the first one to activities based on the functioning of the BeeBot in combination with some elemental Mathematics aspects, such as creating paths for the robot based on geometrical tiles. The second year implemented programming activities and sequences in order to develop robotic artifacts that could perform different basic actions, such as making out simple arithmetic.
Ferrada et al. (2023) Ciudad Sostenible Granatensis-Robotics (CISOGRA)	12	MAT SCI	mBot, Scratch	In order to tackle several different activities related to sustainable development and mobility, students have to program the robotics resource to, among other actions, carry out specific routes through a prop recreation of the city where the program was conducted.
Huang et al. (2013) *Ad-hoc* program	10	TEC	LEGO Mindstorms NXT (and its platform), Logo	The first three weeks were dedicated to building a robot using the provided kit and programming it through a visual interface, while the control group was devoted to learning programming only through flowcharts in a traditional manner. The next seven weeks were dedicated to programming a Logo robotic turtle, with no distinction between the control and experimental groups.
Merino-Armero et al. (2022) *Ad-hoc* program	6	SCI MAT	Ozobot	The sessions addressed curricular aspects linked to Social Sciences, mainly the orographic properties of both Spain and Europe. Several activities were introduced to develop Computational Thinking among the student group, using the idea of programming a spatial imaginary robot that had to do terrain mapping as the gimmick for the activities.
Ortiz (2011) Engineering Fusion	5	MAT	Unspecified LEGO robotics kit	Experienced teachers assisted learners in understanding and implementing the concepts of ratio and proportion through the construction of LEGO-based robotic products throughout the sessions.
Papadakis et al. (2024) *Ad-hoc* program	13	ENG	Microsoft MakeCode	The intervention involved two different phases, including initial sessions in which learners were introduced to programming under a block-based approach. A second phase was based on implementing the sequential and control structures previously learned in order to build artifacts that could listen to music and a mechanism able to move.
Sáez-López et al. (2021) *Ad-hoc* program	3	SCI	Scratch	The experimental group developed the identical didactic unit as the control group; however, learners were required to develop various activities related to the addressed content through block programming, whereas the control group did so in a traditional pen and pencil approach.

(Continued)

TABLE 4 (Continued)

Program	Duration in weeks	STEM	Resources	Procedure
Sáez-López et al. (2019) *Ad-hoc* program	6	MAT SCI	mBot, Scratch	Students were required to use knowledge acquired via pre-planned lessons on whole numbers, electricity, and magnetism, in order to program the robotics resource at hand. They were required to develop certain actions, such as moving in one direction or another at varying speeds, while ensuring that it avoids obstacles detected by its ultrasonic sensor.
Smakman et al. (2021) Robotics-supported Bareka method	5	MAT	NAO, Robotsindeklas platform	Each session consisted of exercises pertaining to arithmetic measurement and/or the graphical representation of data.

Authors' own work. SCI, Science; TEC, Technology; ENG, Engineering; MAT, Mathematics. Program length has been standardized in weeks comprised, approximately, of a two-hour intervention session.

intervention programs and even build robotics on their own, without depending on buying external prefabricated products that may result in additional expenses either for educators or their institutions, which would directly contradict Moreno-Palma et al.'s (2024) hypotheses.

It is, however, worth noting that the geographical international distribution mostly presented by Tselegkaridis and Sapounidis (2022) mostly aligns with the one derived from the final study sample of the present research work, with the exception of the appearance of Asian learning contexts, mainly Taiwan, as well as the retrieval of ER reports located in northern European countries. Still, the majority of experiences have been developed both in southern Europe and the United States, as indicated by the aforementioned authors.

Moving on to the information associated with RQ2, it has been found that, in Primary Education, the majority of published reported ER experiences, at least within the STEM competence framework, have been developed in the later grades of the stage, the third cycle, which further supports Ortuño and Serrano's (2024) findings while contradicting the results reported by Tselegkaridis and Sapounidis (2022). It is more than likely that one of the reasons behind this age barrier lies beneath the special learning benefit fostered by ER when applied to content and skills that have a higher cognitive workload and complexity, as stated by Alonso-García et al. (2024). Therefore, teachers in lower grades may feel reticent to use robotics at younger ages due to the apparent complexity behind their control, management, and maintenance, thus backing up Bravo et al.'s (2021) ideas.

Despite the fact that some studies did not specifically report participant learners' sex, the information obtained through analyzing the ones that do so proves that men and women tend to participate with equal interest and enjoyment in ER-based STEM experiences, which further reinforces the ideas presented in Romero-Rodríguez et al. (2023) and Jung and Lee (2022) about how women can be subject to characteristic attitudinal changes toward STEM when working with robotics in educational environments.

Regarding RQ3, it has been observed that the majority of the educators and researchers prefer to establish educational contexts in which, mostly, the participant learner population is pre-chosen and specifically configured in order to provide students with optimal learning conditions within the designed intervention programs. Although this approach, as established by Avsec et al. (2016), may be of great benefit, as it further favors attitudinal and emotional changes within learners derived from the educational experiences, it may be quite detrimental when it comes to purely comparing the learning result between two independent groups in order to actually retrieve data of quality and rigor in relation to the usefulness of ER in stages such as Early Childhood Education and Primary Education (Liu et al., 2023).

In relation to RQ4, being this the most relevant item of the present research study regarding curriculum design and management, ER programs aiming to develop STEM competencies appear to be subject to both longer and shorter intervention duration, therefore supporting Ching and Hsu's (2024) claims, although it can be said that long interventions are more likely to extend themselves in time than short interventions to end up being too concentrated or reduced, which can be a factor of relevance when designing similar didactic plans with younger learners due to their limited attention spans.

Another point of interest to be addressed within this RQ is the high presence of programming as a means to ER instead of simply writing lines of code without further or tangible application. As a consequence, proving Chatzopoulos et al.'s (2022) and Qu and Fok's (2022) claims about the educational use of ER as analogical forms of the abstract programming world, the learn by making approach, presented as innovative and groundbreaking by Lorenzo et al. (2024), has ended up being one of the most frequent and utilized ways in which ER is introduced in the Primary Education classroom.

Conversely, using robots as prefabricated social companions, as explored by Lorenzo et al. (2024) and Evripidou et al. (2020), although present in the study sample, does not hold a significant position within current ER approaches. Similarly, humanoid robots have barely been used in the reviewed interventions, which can be traced back to how most of them were developed in later years during the stage, while robots with expressive and emotional capabilities are seen as more appropriate for earlier grades (Istenic et al., 2021) and generally disapproved by educators (Zhang et al., 2023).

Finally, the information provided in RQ5 shows that ER actually has an intermediate beneficial effect on the development of STEM competences, which, although in line with the conclusions established by Palomino et al. (2022), positions itself against Coşkun and Filiz's (2023) claims related to the vast potential of ER to foment certain behaviors and attitudes in learners, as it was attitude measures that tended to be either negative or non-influential when introduced as part of the conducted meta-analysis. This, although potentially due to a plethora of factors, could be explained by inadequate usage of ER in classrooms that, following Zhao et al. (2024), may have led children to prioritize its playful aspect over learning and attention.

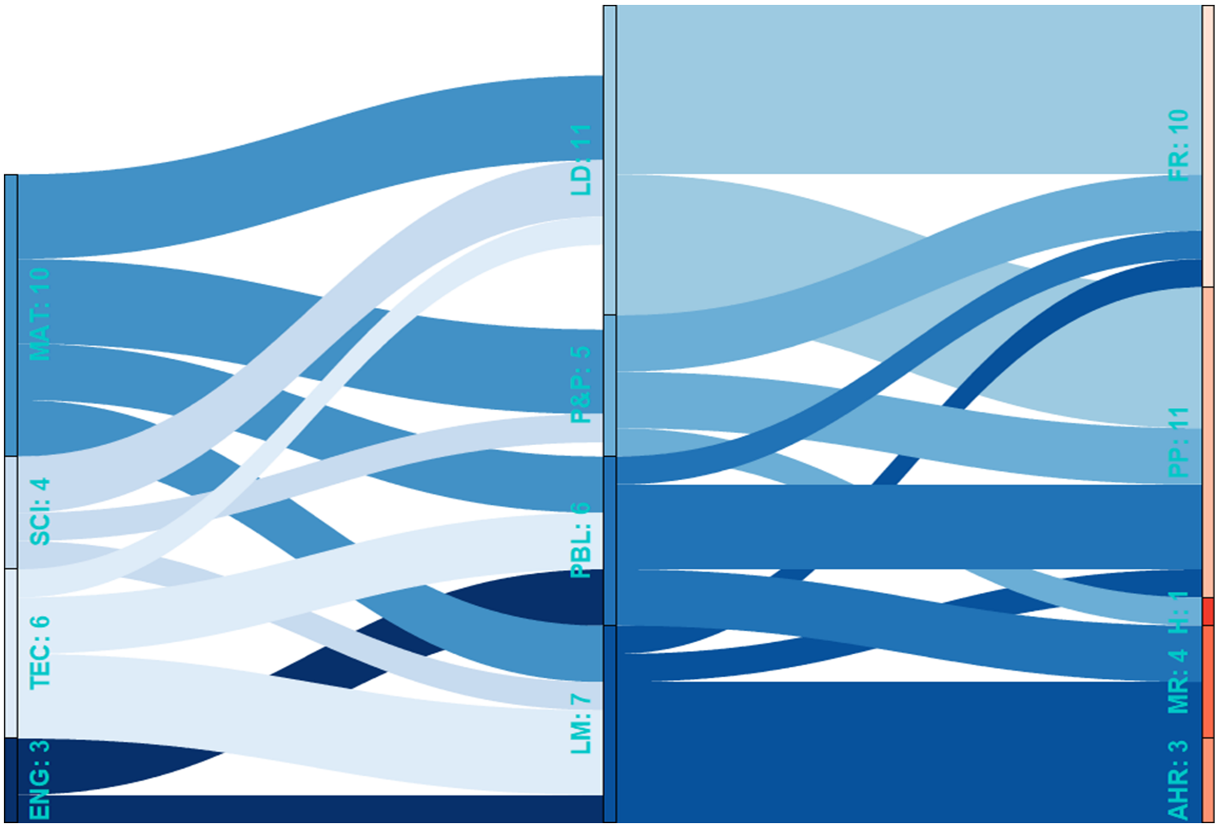

FIGURE 5
Connections between the reviewed methodological aspects. SCI, Science; TEC, Technology; ENG, Engineering; MAT, Mathematics; LM, Learn by Making; PBL, Project-Based Learning; P&P, Pen and Paper Activities; LD, Learn by Doing; AHR, *Ad-hoc* Robots; MR, Modular Robots; PP, Programming Platforms; FR, Floor Robots.

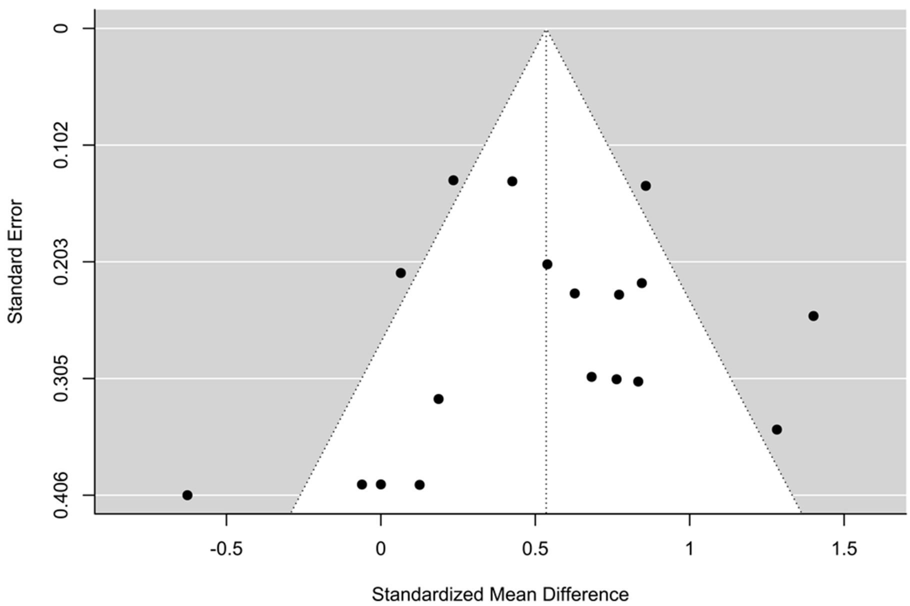

FIGURE 6
Funnel plot including the k = 18 final study sample registers.

10.3389/feduc.2024.1480908

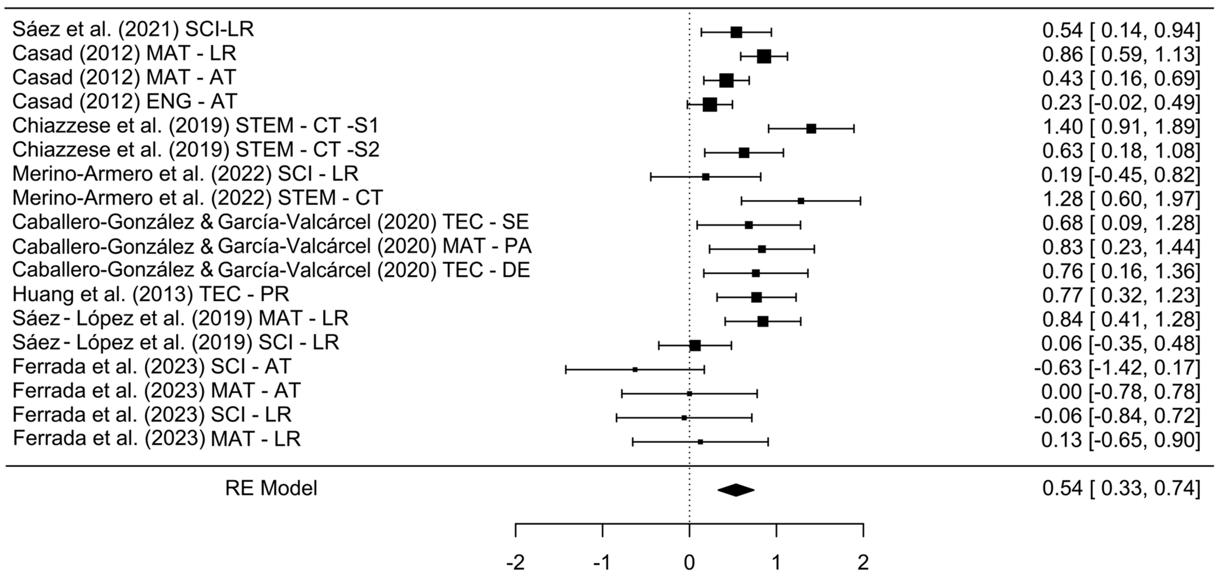

FIGURE 7
Forest plot including the k = 18 final study sample registers. SCI, Science; TEC, Technology; ENG, Engineering; MAT, Mathematics; LR, Learning Result; AT, Attitude; CT, Computational Thinking; S1/S2, Sample 1/Sample 2; SE, Sequencing; PA, Pattern recognition; DE, Debugging; PR, Programming.

TABLE 5 Heterogeneity statistics.

Tau	Tau^2	I^2	H^2	R^2	df	Q	*p*
0.351	0.1229 (SE = 0.0653)	70.28%	3.365	N/A	17.000	52.134	< 0.001

Tau^2 estimated through Restricted Maximum Likelihood. SE, Standard Error; Significance at $p < 0.008$.

TABLE 6 Predicted SMD based on the reported effect sizes of the final study sample.

	Estimate	SE	Z	*p*	CI lower bound	CI upper bound
Intercept	0.535	0.104	5.15	< 0.001	0.3314	0.7393

Regression analysis was conducted under the Random Effects Model approach.

4.2 Limitations and future research lines

This study, naturally, did not go out without limitations. Firstly, given that it was focused on a specific educational stage, some works were eventually discarded due to having taken student samples around the age limits that were initially established in the criteria for the present review, which could have potentially altered the presented results. Additionally, a lack of consensus about the belonging of certain content and skills to the field of STEM competencies may have hindered the potential retrieval of studies of interest due to not having identified their publication with the proper contextual background and/or keywords, similarly to how CT is usually defined as a competence linked to the STEM framework when it is purely interdisciplinary. Finally, some studies not reporting the exact outcome measure of their student sample divided by sex and/or gender have notably limited the results available to the present review, as having meta-analyzed the effect sizes of the ER interventions within the STEM field in relation to empowering future women aiming at their integration within a traditionally considered male-exclusive environment could have resulted in groundbreaking progress toward inclusive education.

As a consequence of all these, future research lines pointing out and further investigating the potential effects of ER not only with regard to STEM competences but also to other kinds of curricular aspects may be of urgent interest, as these advancements are, in our days, in a critical phase of prelaminar acceptance within classrooms, awaiting further validation and support from researchers who, analyzing the existing practices, are able to determine potential flaws and/or benefits of the approach at hand. Therefore, deeper research regarding the implementation of ER both in Primary Education as well as other educational stages is urgently needed in order to establish frameworks of reference able to aid curriculum designers and educators alike in their respective labors.

5 Conclusion

The conduct of this systematic review and meta-analysis has demonstrated that ER is an educational approach with wide baggage in the Primary Education stage. Numerous experiences have shown that ER can introduce a great deal of interdisciplinary contents of interest through active methodologies in which learners, under their innate curiosity and spontaneity, act as true software engineers and mechanics, building all kinds of robots. Additionally, this study has proven how, in situations of lack of formation and/or training toward the use and implementation of any given resource, teachers in this educational stage are able to make these unprecedent tools their own, deconstructing their bases and foundations in order to transform what appears to be an amalgam of theoretical aspects and skills into true and unique learning experiences for their pupils.

The results retrieved from the present literature review determined the publication trend regarding publication type and date of the final study sample, deepening further into the characteristics of the target student sample as well as their distribution throughput Primary Education, while summarizing the primary research methodological aspects that were present in the accepted interventions, including statistical and standardized tests, among others. Finally, both the conducted narrative review and the developed meta-analysis were able to briefly establish some of the most common ER practices applied to STEM educational proposals, followed by a critical analysis of their actual outcomes, determining an overall moderate beneficial effect. Thus, this study was capable of answering every posed research question and, as a consequence, fulfilling its primary objective.

Undoubtedly, ER is one of the most interesting, innovative, and attractive approaches to technology-mediated education in our days. However, as such, it suffers from the very same principle as every other educational resource emerging from technological advancement: misuse due to inadequate training. Although the innate value of *ad-hoc* approaches cannot possibly be denied, as educators are the individuals who best know their learners and their unique interests, likes, and traits. However, the vast heterogeneity of results linked to chance, as well as the absence of genuine standardized tests and intervention programs, irrevocably call for some unification within the field in order to address potential barriers that may hinder the true benefits of these cabled companions.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

IT-G: Conceptualization, Investigation, Methodology, Writing – original draft. FH-L: Resources, Supervision, Validation, Writing – review & editing. J-MR-R: Funding acquisition, Investigation, Project administration, Software, Validation, Writing – review & editing. AM-M: Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was funded by Consejería de Universidad, Investigación e Innovación (Grant Number C-SEJ-009-UGR23) and the ERDF Andalusia Program 2021-2027.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alonso-García, S., Rodríguez, A.-V., Ramos, M., and Victoria-Maldonado, J.-J. (2024). Enhancing computational thinking in early childhood education with educational robotics: a meta-analysis. *Heliyon* 10:e33249. doi: 10.1016/j.heliyon.2024.e33249

Andrée, M., Anderhag, P., Björnhammer, S., and Salomonsson, N. (2024). Aesthetic experience in technology education – the role of aesthetics for learning in lower secondary school robotic programming. *Front. Educ.* 9:1291070. doi: 10.3389/feduc.2024.1291070

Avsec, S., Rihtarič, D., and Kocijancic, S. (2016). The impact of robotics-enhanced approach on students' satisfaction in open learning environment. *Int. J. Eng. Educ.* 32, 804–817.

Begg, C. B., and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50, 1088–1101. doi: 10.2307/2533446

Bento, A. L., Nunes, J. M., Bernades, R. J., and Lopes, T. (2024). Meaningful learning in mathematics: a study on motivation for learning and development of computational thinking using educational robotics. *Educ. Media Int.* 61, 4–15. doi: 10.1080/09523987.2024.2357472

Bers, M. U. (2021). Coding, robotics and socio-emotional learning: developing a palette of virtues. *Pixel-Bit. Rev. Medios Educ.* 62, 309–322. doi: 10.12795/pixelbit.90537

Bravo, F. A., Hurtado, J. A., and González, E. (2021). Using robots with storytelling and Drama activities in science education. *Educ. Sci.* 11:329. doi: 10.3390/educsci11070329

Caballero-González, Y. A., and García-Valcárcel, A. (2020). Learning with robotics in primary education? A means of stimulating computational thinking. *Educ. Knowl. Soc.* 21:10. doi: 10.14201/eks.21443

Casad, B., and Jawaharlal, M. (2012). *Learning through guided discovery: an engaging approach to k-12 stem education [congress paper].* 2012 ASEE Annual Conference & Exposition, San Antonio, Texas, United States of America.

Chatzopoulos, A., Kalogiannakis, M., Papadakis, S., and Papoutsidakis, M. (2022). A novel, modular robot for educational robotics developed using action research evaluated on technology acceptance model. *Educ. Sci.* 12:274. doi: 10.3390/educsci12040274

Chen, H., Cohen, P., and Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun. Stat. Simul. Comput.* 39, 860–864. doi: 10.1080/03610911003650383

Chiazzese, G., Arrigo, M., Chifari, A., Lonati, V., and Tosto, C. (2019). Educational robotics in primary school: measuring the development of computational thinking skills with the Bebras tasks. *Informatics* 6:43. doi: 10.3390/informatics6040043

Ching, Y.-H., and Hsu, Y.-C. (2024). Educational robotics for developing computational thinking in young learners: a systematic review. *TechTrends* 68, 423–434. doi: 10.1007/s11528-023-00841-1

Chou, P.-N. (2018). Skill development and knowledge acquisition cultivated by maker education: evidence from Arduino-based educational robotics. *Eurasia J. Math. Sci. Tech. Ed.* 14, 1–15. doi: 10.29333/ejmste/93483

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* 10, 101–129. doi: 10.2307/3001666

Coşkun, T. K., and Filiz, O. (2023). The impact of twenty-first century skills on university students' robotic achievements. *Educ. Inf. Technol.* 28, 16255–16283. doi: 10.1007/s10639-023-11850-1

Darmawansah, D., Hwang, G.-J., Chen, M.-R. A., and 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. *Int. J. STEM Educ.* 10:12. doi: 10.1186/s40594-023-00400-3

Egger, M. G., Smith, M. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629–634. doi: 10.1136/bmj.315.7109.629

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

Ferrada, C., Carrillo-Rosúa, J., Díaz-Levicoy, D., and Silva-Díaz, F. (2023). Evaluation of a sustainable educational proposal with a STEM approach to improve attitudes towards science or mathematics in 5th and 6th grade primary school students in Spain. *Investig. Ensino Cienc.* 28, 111–126. doi: 10.22600/1518-8795.ienci2023v28n1p111

Finsterbach, C., Pedersen, P. L., and Tvedebrink, T. (2023). Integrating computational thinking to enhance students' mathematical understanding. *J. Pedagog. Res.* 7, 127–142. doi: 10.33902/JPR.202318531

Frederiksen, M. R., Stoy, K., and Matarić, M. (2024) *Toward anxiety-reducing pocket robots for children* [paper presentation]. 21st international conference on ubiquitous robots, New York, NY, USA. doi: 10.1109/UR61395.2024.10597525

Fung, K. Y., Lee, L. H., Sin, K. F., Song, S., and Qu, H. (2024). Humanoid robot-empowered language learning based on self-determination theory. *Educ. Inf. Technol.* 29, 18927–18957. doi: 10.1007/s10639-024-12570-w in press

Garzón, E., Sola, T., Trujillo, J. M., and Rodríguez, A. M. (2021). Competencia digital docente en educación de adultos: un estudio en un contexto español [Digital competence for educators in adult education: a study in the Spanish context]. *Pixel-Bit. Rev. Medios Educ.* 62, 209–234. doi: 10.12795/pixelbit.89510

Higgins, J., and Green, S. (2008). Cochrane handbook for systematic reviews of interventions. Hoboken, NJ: Wiley-Blackwell.

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., et al. (2018). The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *Educ. Inf.* 34, 285–291. doi: 10.3233/EFI-180221

Huang, K., Yang, T., and Cheng, C. (2013). Engineering to see and move: teaching computer programming with flowcharts vs. LEGO robots. *Int. J. Emerg. Technol. Learn.* 8, 23–26. doi: 10.3991/ijet.v8i4.2943

Hunter, J. E., Jensen, J. L., and Rodgers, R. (2014). The control group and meta-analysis. *J. Methods Meas. Soc. Sci.* 5, 3–21.

Istenic, A., Bratko, I., and Rosanda, V. (2021). Are pre-service teachers disinclined to utilise embodied humanoid social robots in the classroom? *Br. J. Educ. Technol.* 52, 2340–2358. doi: 10.1111/bjet.13144

Jackson, D., and Turner, R. (2017). Power analysis for random-effects meta-analysis. *Res. Synth. Methods* 8, 290–302. doi: 10.1002/jrsm.1240

Jung, S. E., and Lee, K. (2022). A girl's gendered engagement in designing and building robots. *Int. J. Technol. Des. Educ.* 32, 2425–2444. doi: 10.1007/s10798-021-09705-2

Kalaitzidou, M., and Pachidis, T. P. (2023). Recent robots in STEAM education. *Educ. Sci.* 13:272. doi: 10.3390/educsci13030272

Konishi, S., Kuwata, M., Matsumoto, Y., Yoshikawa, Y., Takata, K., Haraguchi, H., et al. (2024). Self-administered questionnaires enhance emotion estimation of individuals with autism spectrum disorders in a robotic interview setting. *Front. Psych.* 15:1249000. doi: 10.3389/fpsyt.2024.1249000

Krūmiņš, D., Schumann, S., Vunder, V., Põlluäär, R., Laht, K., Raudmäe, R., et al. (2024). Open remote web lab for learning robotics and ROS with physical and simulated robots in an authentic developer environment. *IEEE Trans. Learn. Technol.* 17, 1325–1338. doi: 10.1109/TLT.2024.3381858

Küçükaydın, M. A., Çite, H., and Ulum, H. (2024). Modelling the relationships between STEM learning attitude, computational thinking, and 21st century skills in primary school. *Educ. Inf. Technol.* 29, 16641–16659. doi: 10.1007/s10639-024-12492-7 in press

Leoste, J., Kikkas, K., Tammemäe, K., Rebane, M., Laugasson, E., and Hakk, K. (2022). "Telepresence robots in higher education – the current state of research" in Robotics in education. eds. W. Lepuschitz, M. Merdan, G. Koppensteiner, R. Balogh and D. Obdržálek (Berlin: Springer), 124–134.

Liu, L., Oh, H., Zhang, L., Fang, T., Huang, M., Hao, Y., et al. (2023). A study of children's learning and play using an underwater robot construction kit. *Int. J. Technol. Des. Educ.* 33, 317–336. doi: 10.1007/s10798-021-09720-3

Lorenzo, G., Lorenzo-Lledó, A., and Gilabert-Cerdá, A. (2024). Application of robotics in autistic students: a pilot study on attention in communication and social interaction. *Technol. Knowl. Learn.* 29, 757–780. doi: 10.1007/s10758-023-09718-x

Martínez, A., De la Cruz-Campos, J.-C., Fernández, J., and Victoria-Maldonado, J.-J. (2023a). El videojuego en las aulas de Educación Primaria. El miedo a lo desconocido [Videogames in the primary education classroom: fearing the unknown]. In Cáceres, M. P., Schmitt, E., García, P. J., and Fuentes, A. (coords.), Experiencias basadas en evidencias en la formación y práctica docente: perspectivas de investigación para la innovación (pp. 33–38). Madrid: Dykinson.

Martínez, A., Victoria-Maldonado, J.-J., and Alcalá, M. J. (2023b). Avances teóricos de la robótica en Educación Infantil [Theoretical advancements of robotics in early childhood education]. In J. M. Trujillo, P. J. García, T. Pertínez and B. Berral (coords.), Transformando la educación a través de la tecnología: innovación, investigación y aprendizaje digital (pp. 205–211). Madrid: Dykinson

Merino-Armero, J., González-Calero, J., Cózar-Gutiérrez, R., and Del Olmo-Muñoz, J. (2022). Unplugged activities in cross-curricular teaching: effect on sixth graders' computational thinking and learning outcomes. *Multimodal Technol. Interact.* 6:13. doi: 10.3390/mti6020013

Meza, J., Ramírez, A., and Gardea, R. (2012). Laboratorio móvil tecno educativo: cursos de robótica de bajo costo para la alfabetización científica y tecnológica [Mobile techno educational laboratory: offering low-cost robotic courses in favor of scientific and technological alphabetization]. *Educ. Knowl. Soc.* 13, 145–161. doi: 10.14201/eks.9135

Misirli, A., and Komis, V. (2023). Computational thinking in early childhood education: the impact of programming a tangible robot on developing debugging knowledge. *Early Child. Res. Q.* 65, 139–158. doi: 10.1016/j.ecresq.2023.05.014

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

Moreno-Palma, N., Hinojo-Lucena, F.-J., Romero-Rodríguez, J.-M., and Cáceres-Reche, M.-P. (2024). Effectiveness of problem-based learning in the unplugged computational thinking of university students. *Educ. Sci.* 14:693. doi: 10.3390/educsci14070693

Ortiz, A. (2011). *Fifth grade students' understanding of ratio and proportion in an engineering robotics program [congress paper].* 2011 ASEE Annual Conference & Exposition, Vancouver, British Columbia, Canada.

Ortuño, G., and Serrano, J. L. (2024). Implementación y formación del profesorado de educación primaria en pensamiento computacional: una revisión sistemática [Implementation and training of primary education teachers in computational thinking: a systematic review]. *Rev. Iberoam. Educ. Distancia* 27, 255–287. doi: 10.5944/ried.27.1.37572

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLoS Med.* 18:e1003583. doi: 10.1371/journal.pmed.1003583

Palomino, J. M., Cáceres-Reche, M. P., Hinojo, F.-J., and Aznar, I. (2022). Hacia un liderazgo pedagógico en la Educación Superior [Towards pedagogical leadership in Higher Education]. In F.-J. Hinojo, O. C. Marichal and M. P. Cáceres-Reche, & Barrientos, C. A. (pp. 65–78), Aportes de investigación derivados de la Red Iberoamericana de Investigación en Liderazgo y Prácticas Educativas (RILPE). Madrid: Dykinson.

Papadakis, S., Tzagkaraki, E., and Kalogiannakis, M. (2024). "Teaching electricity topics with project-based learning and physical computing to enhance primary school students in science education. An educational experiment with BBC Micro:bit board" in Towards a hybrid, flexible and socially engaged higher education proceedings of the 26th international conference on interactive collaborative learning – volume 2. eds. M. E. Auer, U. R. Cukierman, E. Vendrell and E. Tovar (Berlin: Springer), 320–331.

Papert, S. (1993). The children's machine. *Technol. Rev.* 96, 28–36.

Ponce, P., Molina, A., Caudana, E., Reyes, G., and Parra, N. (2019). Improving education in developing countries using robotic platforms. *Int. J. Interact. Des. Manuf.* 13, 1401–1422. doi: 10.1007/s12008-019-00576-5

Pourfannan, H., Mahzoon, H., Yoshikawa, Y., and Ishiguro, H. (2024). Sound masking by a low-pitch speech-shaped noise improves a social robot's talk in noisy environments. *Front. Robot. AI* 10:1205209. doi: 10.3389/frobt.2023.1205209

Qu, J. R., and Fok, P. K. (2022). Cultivating students' computational thinking through student–robot interactions in robotics education. *Int. J. Technol. Des. Educ.* 32, 1983–2002. doi: 10.1007/s10798-021-09677-3

Quintana, D. (2017) *How to calculate statistical power for your meta-analysis*. Towards Data Science. Available at: https://towardsdatascience.com/how-to-calculate-statistical-power-for-your-meta-analysis-e108ee586ae8

Quintana, D., and Tiebel, J. (2018). *Meta power calculator* (version 1.0) [software]. ShinyApps. Available at: https://jtiebel.shinyapps.io/MetaPowerCalculator (Accessed August 8, 2024).

Raycheva, N. (2024). Concept on stem education in accordance with Bulgarian curriculum content of secondary school. *Innov. Educ. Technol.* 96, 7–15. doi: 10.53656/ped2024-2s.01

Romero-Rodríguez, J.-M., Cruz-Campos, J.-C., Ramos-Navas-Parejo, M., and Martínez-Domingo, J. A. (2023). Robótica educativa para el desarrollo de la competencia STEM en maestras en formación [Educational robotics for the development of STEM competencies in pre-service female teachers]. *Bordón Rev. Pedagog.* 75, 75–92. doi: 10.13042/Bordon.2023.97174

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychol. Bull.* 86, 638–641. doi: 10.1037/0033-2909.86.3.638

Sáez-López, J., Otero, R., and García-Cervigón, S. (2021). Introducing robotics and block programming in elementary education. *Rev. Iberoam. Educ. Distancia* 24, 95–113. doi: 10.5944/ried.24.1.27649

Sáez-López, J., Sevillano-García, M., and Vazquez-Cano, E. (2019). The effect of programming on primary school students' mathematical and scientific understanding: educational use of mBot. *Educ. Technol. Res. Dev.* 67, 1405–1425. doi: 10.1007/s11423-019-09648-5

Sapounidis, T., Tselegkaridis, S., and Stamovlasis, D. (2024). Educational robotics and STEM in primary education: a review and a meta-analysis. *J. Res. Technol. Educ.* 56, 462–476. doi: 10.1080/15391523.2022.2160394

Seckel, M. J., Vásquez, C., Samuel, M., and Breda, A. (2022). Errors of programming and ownership of the robot concept made by trainee kindergarten teachers during an induction training. *Educ. Inf. Technol.* 27, 2955–2975. doi: 10.1007/s10639-021-10708-8

Smakman, M., Smit, K., Lan, E., Fermin, T., van Lagen, J., Maas, J., et al. (2021). "Social robots for reducing mathematics hiatuses in primary education, an exploratory field study" in 34th bled eConference digital support from crisis to progressive change – Conference proceedings. eds. A. Pucihar, M. Kljajić, R. Bons, H. Sheombara and D. Vidmar (Slovenia: University of Maribor), 657–674.

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., et al. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343:d4002. doi: 10.1136/bmj.d4002

Sun, L., and Liu, J. (2024). A gender differential analysis of educational robots' effects on primary teachers' computational thinking: mediating effect of programming attitudes. *Educ. Inf. Technol.* 29, 19753–19782. doi: 10.1007/s10639-024-12655-6 in press

Sung, W., Ahn, J., and Black, J. B. (2017). Introducing computational thinking to young learners: practicing computational perspectives through embodiment in mathematics education. *Technol. Knowl. Learn.* 22, 443–463. doi: 10.1007/s10758-017-9328-x

Tselegkaridis, S., and Sapounidis, T. (2022). Exploring the features of educational robotics and STEM research in primary education: a systematic literature review. *Educ. Sci.* 12:305. doi: 10.3390/educsci12050305

UNESCO Institute for Statistics (2012). International standard classification of education: ISCED 2011. Paris: UNESCO Publications.

Valentine, J. C., Pigott, T. D., and Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for META-analysis. *J. Educ. Behav. Stat.* 35, 215–247. doi: 10.3102/1076998609346961

Vitello, S., Greatorex, J., and Shaw, S. (2021). *What is competence? A shared interpretation of competence to support teaching, learning and assessment*. Cambridge University Press & Assessment. Available at: https://www.cambridgeassessment.org.uk/Images/645254-what-is-competence-a-shared-interpretation-of-competence-to-support-teaching-learning-and-assessment.pdf

Wang, X., Chan, K. K., Li, Q., and Leung, S. O. (2024). Do 3–8 years old children benefit from computational thinking development? A meta-analysis. *J. Educ. Comput. Res.* 62, 1182–1208. doi: 10.1177/07356331241236744

West, S. L., Gartlehner, G., Mansfield, A. J., Poole, C., Tant, E., Lenfestey, N., et al. (2010). *Comparative effectiveness review methods: clinical heterogeneity* (report n° 10-EHC070-EF). US Agency for Healthcare Research and Quality. Available at: https://pubmed.ncbi.nlm.nih.gov/21433337/

Xie, Y., Liang, C., Zhou, P., and Zhu, J. (2024). When should chatbots express humor? Exploring different influence mechanisms of humor on service satisfaction. *Comput. Hum. Behav.* 156:108238. doi: 10.1016/j.chb.2024.108238

Yang, W. (2024). Coding with robots or tablets? Effects of technology-enhanced embodied learning on preschoolers' computational thinking and social-emotional competence. *J. Educ. Comput. Res.* 62, 938–960. doi: 10.1177/07356331241226459

Zhang, X., Chen, Y., Bao, Y., and Hu, L. (2023). Robot illustrated: exploring elementary students' perceptions of robots via the draw-a-robot test. *J. Res. Technol. Educ.*, 1–26. doi: 10.1080/15391523.2023.2232058 in press

Zhao, J.-H., Yang, Q.-F., Lian, L.-W., and Wu, X.-Y. (2024). Impact of pre-knowledge and engagement in robot-supported collaborative learning through using the ICAPB model. *Comput. Educ.* 217:105069. doi: 10.1016/j.compedu.2024.105069