Sociedade Brasileira de Educação Matemática

RIPEM

Systematic review in Mathematics Education: what it is and how to do it

Denise Sayuri Oda Nampo
Universidade Federal da Integração Latino-Americana
Foz do Iguaçu, PR — Brasil
✉ denise.nampo@unila.edu.br
0000-0002-8534-160X

Fernando Kenji Nampo
Universidade Federal da Integração Latino-Americana
Foz do Iguaçu, PR — Brasil
✉ fernando.nampo@unila.edu.br
0000-0001-6411-6131

Susimeire Vivien Rosotti de Andrade
Universidade Estadual do Oeste do Paraná
Foz do Iguaçu, PR — Brasil
✉ susimeire.andrade@unioeste.br
0000-0001-9188-8620

2238-0345

10.37001/ripem.v16i1.4544

Received • 06/04/2025
Approved • 09/02/2026
Published • 30/04/2026

Editorial • Edvonete Souza de Alencar
Veridiana Rezende

Abstract: This article presents a methodological study aimed at systematizing the main procedures for conducting an effective systematic review in the field of Mathematics Education. Considering its essential role in organizing scientific knowledge, especially in a context of rapid expansion, we carried out a documentary analysis of consolidated international guidelines, describing fundamental procedures and identifying particularities inherent to the application of this method in investigations within Mathematics Education. In addition to increasing the statistical power of research through data aggregation, this approach helps clarify inconsistencies among individual studies, contributing to evidence-based decision-making. As a result, we developed a synthesis guide that highlights methodological nuances that are still scarcely discussed in the field's literature. The article contributes to strengthening the planning and execution of effective systematic reviews by researchers in Mathematics Education, promoting greater rigor, transparency, and reproducibility.

Keywords: Evidence Synthesis. Systematic Review. Mathematics Education. Research Methodology.

Revisión sistemática en Educación Matemática: qué es y cómo hacerla

Resumen: Este artículo presenta un estudio metodológico cuyo objetivo es sistematizar los principales procedimientos para la realización de una revisión sistemática de efectividad en el campo de la Educación Matemática. Considerando su papel esencial en la organización del conocimiento científico, especialmente en un contexto de rápida expansión, llevamos a cabo un análisis documental de directrices internacionales consolidadas, describiendo procedimientos fundamentales e identificando particularidades inherentes a la aplicación de este método en investigaciones en el ámbito de la Educación Matemática. Además de aumentar el poder estadístico de la investigación mediante la agregación de datos, este enfoque ayuda a aclarar inconsistencias entre estudios individuales, contribuyendo a la toma de decisiones basada en evidencias. Como resultado, elaboramos una guía sintética que pone de relieve matices metodológicos aún poco discutidos en la literatura del área. El artículo contribuye a cualificar la planificación y la ejecución de revisiones sistemáticas de efectividad por parte de investigadores en Educación Matemática, promoviendo un mayor rigor, transparencia y reproducibilidad.

Palabras clave: Síntesis de Evidencia. Revisión Sistemática. Educación Matemática. Metodología de Investigación.

Revisão sistemática em Educação Matemática: o que é e como fazer

Resumo: Este artigo apresenta um estudo metodológico cujo objetivo é sistematizar os principais procedimentos para a condução de revisões sistemáticas de efetividade no campo da Educação Matemática. Considerando o papel essencial dessa abordagem na organização do conhecimento científico, especialmente em um contexto de crescente produção científica, realizamos uma análise documental de diretrizes internacionais consolidadas, a partir da qual foram descritos procedimentos fundamentais e identificadas particularidades inerentes à aplicação desse método em investigações no campo da Educação Matemática. Além de ampliar o poder estatístico da pesquisa pela agregação de dados, as revisões sistemáticas possibilitam esclarecer inconsistências entre estudos individuais, contribuindo para decisões baseadas em evidências. Como resultado, apresentamos um guia síntese que evidencia nuances metodológicas ainda pouco discutidas na literatura da área. Assim, o artigo contribui para qualificar o planejamento e a execução de revisões sistemáticas de efetividade por pesquisadores da área, promovendo maior rigor metodológico, transparência e reprodutibilidade.

Palavras-chave: Síntese de Evidências. Revisão Sistemática. Educação Matemática. Metodologia de Pesquisa.

1 Introduction

Mathematical education is an area of knowledge belonging to the social sciences and humanities, dedicated to the study of the processes of teaching and learning mathematics (Fiorentini & Lorenzato, 2006). In Brazil, its establishment as an academic field dates back to the late 1970s and early 1980s, a period marked by the founding of the Brazilian Society of Mathematics Education (SBEM). Since then, the consolidation of the field has been accompanied by a growing diversification of topics, approaches, and research methods, resulting in a significant increase in scientific production.

This expansion can be observed in the records in the *Scopus* database, which indicate a 648% increase in the total number of articles on Mathematics education published in the five-year period 2020-2024 (16,510 articles) compared to 2001-2005 (2,205 articles) (Elsevier, 2025). Although this dynamic highlights the growing relevance of the topic and contributes to advances in Mathematics education, it also poses challenges in keeping up with the latest developments and remaining at the forefront of knowledge.

In addition to the significant volume of publications, the results of studies on Mathematics education often present inconsistencies. An illustrative example is the case of active teaching methodologies. Although most studies published up to 2014 did not point to statistically significant differences between learning achieved through these methodologies and learning resulting from traditional teaching, a systematic review showed that, when aggregating the results of individual studies, active methodologies improve academic performance and reduce failure rates by approximately 55% when compared to traditional approaches (Freeman et al., 2014). This discrepancy between the results of individual studies and the conclusions obtained through systematic synthesis reveals an important paradox: a mathematics educator who consults the literature in isolation is more likely to find individual studies that do not show clear advantages of active methodologies and, based on this, may choose not to incorporate them into their pedagogical planning – a decision that would be in line with the isolated evidence but at odds with the conclusions of the more comprehensive and relevant synthesis.

In this context of fragmented scientific production, systematic reviews stand out as fundamental methodological tools for: (1) integrating scattered studies; (2) harmonize seemingly contradictory evidence; and (3) provide a critical synthesis of accumulated knowledge, including the identification of gaps in knowledge and limitations of studies already conducted (Lasserson, Thomas & Higgins, 2022; Mulrow, 1994). From this perspective, mastery of systematic review techniques is an essential skill for researchers in the field of education, allowing them not only to critically evaluate the available literature, but also to guide new practices and investigations based on consolidated evidence.

Given this, this article aims to clarify the main steps necessary for conducting systematic reviews of effectiveness, providing methodological support to researchers interested in adopting this approach. In doing so, it also seeks to bring this discussion closer to the specificities of scientific production in Mathematics education, highlighting the potential of this methodology to qualify research in the field. In the following sections, we present detailed information that will guide the reader in understanding what a systematic review is and how to conduct it, highlighting specific nuances to be considered in its application in the field of Mathematics education.

2 What is a systematic review?

Systematic review is a research methodology aimed at synthesizing the best available evidence on a specific issue. To this end, it uses rigorous and transparent procedures to locate, evaluate, and analyze the results of studies relevant to the topic under investigation (The Campbell Collaboration, 2019). This methodological transparency makes the process reproducible, enabling other researchers to replicate the steps of the review with precision.

Systematic reviews should not be confused with narrative reviews of the literature. While narrative reviews originate from broad research questions and less structured search processes, often limited to one or a few databases without an explicit protocol, systematic reviews are based on rigorous and transparent methods, defined *a priori* in a registered protocol. Their overarching goal is to identify, evaluate, and synthesize all available evidence to answer a clearly defined research question. (O'Gorman *et al.*, 2013; Shamseer *et al.*, 2015).

The methodological rigor of systematic reviews manifests itself in three critical dimensions: (1) comprehensive and reproducible search strategies; (2) systematic processes for selecting and evaluating studies, based on explicit inclusion and exclusion criteria; and (3) standardized methods for data extraction and analysis, all designed to minimize bias at each stage of the process. Ideally, a systematic review should be preceded by the registration of a protocol, which allows any reader to verify the adherence between the initial planning and the actual conduct of the study. This robust methodological structure allows systematic reviews to synthesize primary studies more reliably, thus constituting the highest hierarchy of scientific evidence. (Cumpston *et al.*, 2024; Lasserson *et al.*, 2022).

Additionally, when a systematic review finds two or more studies that are comparable to each other – that is, that are sufficiently similar in terms of population, intervention, comparison, and outcomes – it becomes possible to statistically combine their results. This quantitative synthesis, called meta-analysis, consists of integrating the results of individual studies as if they constituted a single study. (Clarke & Chalmers, 2018; Glass, 1976).

3 Brief history of systematic reviews and meta-analysis

The practice of synthesizing research results to inform decision-making predates the formal emergence of systematic reviews by a long time: the first scientific records date back to the early 18th century. In 1720, physician and mathematician James Jurin, secretary of the Royal Society, together with physicians Thomas Nettleton and John Gasper Scheuchzer, combined study results to argue that variolation – the inoculation of infected material into healthy people, a precursor to vaccination – reduced the risk of death from smallpox. Later, in 1753, Scottish physician James Lind reviewed and compiled existing research on the treatment of scurvy in the publication "A Treatise of the Scurvy", identifying the association between ascorbic acid consumption and the prevention of the disease in sailors. (Bhatt, 2010; Clarke & Chalmers, 2018).

The first recognized attempt at quantitative combination of clinical study data through what is now understood as meta-analysis is attributed to Karl Pearson, mathematician and statistician, then director of the Biometry Laboratory at University College London. In 1904, Pearson compiled data from 11 studies comparing infection and mortality rates from typhoid fever among inoculated and non-inoculated volunteer soldiers, identifying a statistically significant association between inoculation and reduced infection and mortality rates from the disease. (O'Rourke, 2007; Pearson, 1904).

Three years later, in 1907, Joseph Goldberger published a summary of research on typhoid fever and bacteriuria in a study recognized for adopting a systematic set of procedures that later became central elements of systematic review: identification of relevant studies, application of specific criteria for study selection, data extraction and tabulation, and statistical analysis. (Chalmers, Hedges & Cooper, 2002). Subsequently, given the lack of clarity in published reviews and the consequent need to standardize such publications, international groups developed guidelines to guide the preparation and reporting of systematic reviews, including the Quality of Reporting of Meta-analysis (QUOROM), which gave rise to the Preferred Reporting Items of Systematic Reviews and Meta-analysis (PRISMA) (Moher *et al*., 1999; Moher, Liberati, Tetzlaff & Altman, 2009; Page *et al*., 2021).

Although initiatives aimed at combining quantitative data from individual studies date back to the early 20th century, the term meta-analysis was first used in 1976 by Gene V. Glass, a statistician and researcher in the field of educational psychology. Glass used the term to refer to the statistical analysis of a broad set of results from individual and methodologically similar studies, with the aim of integrating and synthesizing research findings (Glass, 1976). The conceptual and methodological differentiation between meta-analysis and systematic review, however, was only consolidated in the 1990s with the advent of Evidence-Based Medicine. While systematic reviews aim to verify the consistency of scientific findings and their possibility of generalization through rigorous qualitative synthesis, meta-analyses seek to increase the accuracy of intervention effect estimates by statistically pooling data in a quantitative synthesis and, consequently, narrowing the measures of dispersion by increasing the number of observations.(Mulrow, 1994)

From this brief historical overview, it is possible to observe that efforts to synthesize research results have a consolidated trajectory, directly related to the need to base decisions – especially in the area of health – on robust, high-quality scientific evidence. In this process, it is worth highlighting the work of two leading international organizations in the production and dissemination of high-quality systematic reviews: the Cochrane Collaboration and the Campbell Collaboration. The former is globally recognized for its rigorous methodological standards and its focus on health intervention studies, in addition to providing free manuals and detailed guidelines for conducting systematic reviews (Higgings *et al*., 2024). The latter, in turn, adopts an interdisciplinary approach, covering the areas of education, health, social development, and economics. Both contribute significantly to the production of scientific evidence that supports informed decision-making and choices in various contexts (The Campbell Collaboration, 2019). We therefore recommend that readers deepen their knowledge of the work carried out by these organizations.

4 What should you know before starting a systematic review?

Before addressing the steps necessary to conduct a systematic review, it is necessary for the reader to understand some fundamental definitions and be familiar with the methodological and technological tools that support its preparation.

4.1 What is PRISMA?

It is a set of guidelines designed to guide the transparent and complete reporting of systematic reviews and meta-analyses. Initially developed in 2009 and updated in 2020, PRISMA aims to ensure clarity in presenting the relevance of the review, the methodological procedures adopted, and the main findings of the research (Page *et al*., 2021). The guide consists of 27 items that should be considered when writing a systematic review, and its most recent version can be found at the official PRISMA[1] website. Although it is not a manual for conducting systematic reviews, prior knowledge of its items contributes significantly to the planning and execution of the study. As an example, the guide establishes the obligation to report the date on which the searches were conducted; if the researcher is not aware of this recommendation from the beginning of the process, they may fail to record essential information, making it impossible to include it later in the final draft of the review.

4.2 What is meta-bias?

The term meta-bias is a combination of the words “meta-analysis” and “bias”. Bias, or systematic error, refers to non-random and directional distortions in the results of a study, resulting from structural flaws in the methodological design, data collection, or analysis, which predictably skew estimates away from their true value, reducing the internal validity of the research (Higging *et al*., 2024). Meta-bias, in turn, refers to biases that emerge in the context of conducting a meta-analysis. It is important to note that meta-bias can occur even when the individual studies included are methodologically sound and free of bias. An example of this is related to scientific publication standards. It cannot be said that articles published in English or in high-impact journals are, in themselves, methodologically biased; however, such characteristics contribute to the emergence of meta-bias, since studies with statistically significant (positive) results are more likely to: (1) be submitted to high-visibility journals; (2) be published in English; and (3) achieve wide dissemination. In contrast, studies with negative or null results tend to: (1) remain unpublished (“publication bias”); (2) be published in regional journals with less impact; and (3) be written in other languages (“language bias”). This asymmetry in the availability of studies, in which positive results are easier to find, can significantly distort the results of evidence syntheses (Egger, Smith, Schneider & Minder, 1997). Thus, systematic reviews that restrict their searches to articles in English and easily locate articles tend to favor the inclusion of studies with positive results – thereby reducing the likelihood of finding studies with negative results – which increases the risk of meta-bias.

[1] https://www.prisma-statement.org/.

4.3 What are independent reviewers?

Independent reviewers are researchers who have institutional, financial, or academic conflicts of interest in relation to the studies analyzed and the other members of the review team. Their primary function is to ensure impartiality in three critical stages of the systematic review process: (1) study selection, (2) data extraction, and (3) methodological assessment. Independence – both in relation to the studies and among the reviewers themselves – is essential to ensure the objectivity and validity of the results, such that current recommendations suggest that there should be two independent reviewers to conduct the aforementioned stages (Aromataris, Lockwood, Porritt, Pilla & Jordan, 2024; Higgins *et al*., 2024). In practical terms, each independent reviewer performs their analysis individually and in isolation. After the evaluations are completed, the results are compared to identify discrepancies and seek consensus. When disagreements persist, they can be resolved through discussion, as they often result from textual oversights, or, when necessary, with the participation of a third independent reviewer.

4.4 What is a reference manager and what is it used for?

Reference managers are programs designed to store, organize, and use references in a practical and efficient manner in scientific and academic work. These tools allow for the automatic insertion of citations while writing the text, as well as the automated generation of reference lists. In addition, they usually allow linking to a document, such as the full text file of the article; making notes about the reference; directly importing references from electronic databases; and identifying references that have been imported in duplicate, facilitating the compilation and organization of sources that will be analyzed in the stages of a systematic review. Among the most popular reference managers are EndNote[2], Mendeley[3] and Zotero[4]. While Mendeley and Zotero are free, EndNote offers a free version with limited functionality and a more complete paid version. In the context of systematic reviews, reference managers are particularly useful for importing references from databases, eliminating duplicate references, and managing the sources used in writing the protocol and final review report.

4.5 Other technological resources

Several online platforms have been developed to support the conducting of systematic reviews, such as Covidence, Rayyan, PicoPortal and Sumari (JBI SUMARI, n.d.; Ouzzani, Hammady, Fedorowicz & Elmagarmid, 2016; PICO Portal, 2025; Veritas Health Innovation, 2023). In summary, these portals assist in the management of included studies and in the execution of steps involved in conducting a systematic review, such as the distribution of references among independent reviewers, the identification of discrepancies, or even the generation of figures and graphs related to the synthesis of studies. Other applications, such as Revman and Comprehensive Meta-Analysis, can also provide support for conducting meta-analyses (The Cochrane Collaboration, 2024). In addition, advances have been observed in the use of artificial intelligence-based applications with the potential to automate certain steps of the systematic review; however, to date, human validation has proven necessary (Blaizot *et al*., 2022).

5 Steps for conducting a systematic review in Mathematics Education

In the upcoming sections, we will present the steps necessary to conduct a systematic

[2] https://access.clarivate.com/login?app=endnote.
[3] https://www.mendeley.com/?interaction_required=true.
[4] https://www.zotero.org/.

review, specifically a systematic review of effectiveness. The text guides the reader/researcher from structuring the research question and drafting the protocol to the stages of conducting the review. For educational purposes, we will use the following example research question: "What is the impact of using educational digital games on elementary school students' learning of algebra, compared to traditional teaching without the use of technology?" In addition, the steps for conducting a systematic review have been organized into 10 steps: the first seven result in the construction of the research protocol, which must be recorded or published, while steps eight to ten encompass the implementation of the protocol, from the literature search to the synthesis of evidence.

5.1 Step one: define the research question

The first step is to define the focus of the review by formulating the research question. A well-formulated research question guides the subsequent methodological aspects, which include defining the eligibility criteria, the procedures for extracting data from the selected studies, and organizing the synthesis and results of the investigation.

The research question should be clear, concise, and specific, and should result in a feasible, relevant, interesting, innovative, and ethical investigation (Cummings, Hulley & Browner, 2013). A widely adopted strategy for structuring research questions aimed at evaluating the effectiveness of interventions is the use of the acronym PICO, which systematizes the main elements necessary for its proper design (The Campbell Collaboration, 2019; Thomas, Kneale, McKenzie, Brennan & Bhaumik, 2024). Although originally developed for clinical research, the use of PICO has also proven appropriate in social science studies. Each letter of the acronym corresponds to an essential element of the research question:

(P) Population: refers to the group of individuals under investigation. Example: elementary school students;

(I) Intervention: refers to the action applied to the population. Example: teaching mediated by educational digital games;

(C) Comparison or control: consists of the group or condition used as a reference to evaluate the effects of the intervention. Example: traditional teaching, without the use of technology;

(O) Outcome: refers to the results or indicators that will be measured to compare the intervention and control groups. Depending on the study design, it is possible to remove the control group from the research question. Example: learning algebra.

After filling in all the elements of the PICO acronym, as shown in the examples above, the following research question can be formulated: What is the impact of using educational digital games on elementary school students' learning of algebra, compared to traditional teaching without the use of technology?

The formulation of the question must be conducted with care and deep reflection. Once you have an initial research question, we recommend conducting a preliminary search on the topic in the main electronic databases. This preliminary search is particularly important when researchers are beginning research in areas in which they do not yet have extensive knowledge, as it allows for an initial and brief exploration of the existing literature, which will deepen their understanding of the topic, identifying existing reviews, and assess the volume of potentially relevant studies. This preliminary analysis contributes to the refinement of the research question and to a more precise delimitation of the scope of the review. Thus, after this exploratory search, authors can validate, adjust, or refine their research questions.

5.2 Step two: define the eligibility criteria

Eligibility criteria correspond to the set of parameters used to determine whether a study is eligible for inclusion in the systematic review, with the aim of supporting the answer to the research question. This section of the protocol should clearly explain which characteristics of the studies constitute inclusion or exclusion criteria. These characteristics may be related to the elements of the PICO acronym, the study design, the context or setting, the period investigated, as well as the language, year of publication, and publication status (Moher *et al*., 2015).

As previously highlighted, a well-defined research question facilitates the definition of eligibility criteria. At this stage, it is important to detail the characteristics of each element of the PICO acronym, as well as other criteria considered relevant by the researcher. (McKenzie *et al*., 2019).

In the example presented, we defined elementary school students as the population. However, this definition requires further specification: does it refer to students in any grade at this level of education? Of both gender and different socioeconomic status? With average or below-average academic performance? Will students with special educational needs be included? It is at this point that the researcher must stipulate the inclusion and exclusion criteria that will guide the selection of studies of interest.

The same reasoning should be applied to the other components of the research. What types of digital games will be considered eligible? Are there any digital games that should be excluded from the study? What type of intervention should the control group receive? How should learning in algebra be measured? Will the review consider studies that used tests administered before and after the interventions? Will studies that did not conduct pre-tests in the intervention group and control group be excluded? What types of studies will be included? Studies published in which years and languages? Will experimental, quasi-experimental, qualitative, or non-comparison group studies be considered? These are countless details that are not explicit in the research question and that the researcher should reflect on.

When establishing eligibility criteria, researchers should keep the population of the review as a central reference, that is, the group to which the study results are intended to be applied. In the field of mathematics education, as in the example adopted, it seems reasonable to include studies with students of different academic performance levels, since this diversity usually reflects the reality of classrooms and, therefore, the results of the systematic review will be useful for most educational settings. An illustrative summary of the eligibility criteria applicable to the case presented is shown in Table 1.

Table 1: Eligibility criteria guided by the acronym PICO

Element	Definition	Eligibility criteria
P	Population	Elementary school students, regardless of gender, race, performance, disability, etc.
I	Intervention	Teaching with educational digital games. Games must be individual. Games using a screen. Cannot be used in combination with another intervention.
C	Control	Traditional teaching, defined as the usual teaching method at the school being evaluated, without the use of digital educational technologies. A control/comparison group is required.
O	Outcome	Algebra learning assessed by exam or test, with alphanumeric results.
N/A	Other criteria	Regardless of publication date or language. Private or public schools.

		Published studies only.

Source: The authors (2025).

5.3 Step three: define your reference search strategy

After designing the research question and defining the inclusion and exclusion criteria, it is necessary to establish a strategy for searching for references. This step usually includes databases and gray literature sources but may also include ongoing studies and unpublished research. In this section of the protocol, it is recommended that the researcher describe in detail the complete search strategy used in at least one of the databases consulted. (Moher *et al*., 2015).

5.3.1 Electronic search in databases

To define the bibliographic search, researchers must select which bibliographic databases will be consulted, and which descriptors and Boolean operators will compose the search strategies. Considering that the systematic review needs to be comprehensive, it is essential that the search be broad enough to identify as many relevant references on the topic in question as possible. In addition, it is essential to prioritize the use of the main databases recognized in the field of knowledge, ensuring the inclusion of significant and high-quality scientific productions.

In the field of Mathematics Education, some of the most widely used databases, according to the Institute of Mathematics and Statistics at the University of São Paulo (IME-USP), include: Google Scholar, European Mathematical Society (EMS), JSTOR, MathSciNet, CAPES Periodicals Portal, Scientific Electronic Library Online (SciELO), Science Direct, SpringerLink, Web of Science and ZbMATH. In addition, other databases relevant to research in Mathematics Education, such as Scopus, Educational Resources Information Center (ERIC) and PsycINFO, broaden access to interdisciplinary studies, especially in the areas of education, learning psychology, and social sciences.

It should be noted that Google Scholar has limitations in terms of advanced search structuring when compared to traditional databases and often returns an unworkable volume of references. Therefore, it is common practice to limit the analysis to the first 100 or 200 records retrieved, since the mechanism organizes the results in order of relevance, facilitating the identification of studies that are potentially relevant to the review.

When searching databases, researchers must define which terms will be used. Search strategies are usually constructed based on some of the elements of the PICO acronym. In systematic reviews aimed at evaluating the effectiveness of interventions, the most frequently used components are (P) Population, (I) Intervention, and, occasionally, (O) Outcome. We emphasize that the greater the number of PICO elements incorporated into the search strategy, the more restricted the set of studies retrieved tends to be.

Taking as an example the research question "What is the impact of using educational digital games on elementary school students' learning of algebra, compared to traditional teaching without the use of technology?", the components of the PICO acronym can be operationalized as shown in Table 2.

Table 2: Search criteria guided by the acronym PICO

Element	Definition	Example
P	Population	Elementary school students
I	Intervention	Teaching with educational digital games

C	Control	Traditional teaching
O	Outcome	Learning of algebra

Source: The authors (2025).

Based on the criteria presented in Table 2, the search strategy may include terms such as “elementary school students” or simply “elementary school”, as well as “teaching with educational digital games” or simply “educational digital games”. The terms corresponding to each element of PICO should be connected to each other using the Boolean operator OR, since the aim is to retrieve studies that use any of the expressions considered synonymous or conceptual variations of the same element. Subsequently, the set of terms used as synonymous for element (P) should be connected to the set of terms used as synonymous for element (I), using Boolean operator AND. This procedure allows studies containing both terms to be identified. In general, searches are conducted in English, as this is the language most commonly used in scientific database indexing systems, regardless of the language in which the studies are published. Table 3 presents an example of a structured search strategy:

Table 3: Example of electronic search

Element	Example	Example
P	Elementary school students	("elementary education" OR "primary education" OR "basic education" OR "fundamental education" OR "elementary school" OR "primary school" OR "compulsory education" OR "basic schooling" OR "mandatory schooling" OR "early school years" OR "initial years of elementary education" OR "final years of elementary education" OR "first stage of basic education" OR "second stage of basic education" OR "early childhood education" OR "lower school" OR "grade school")
I	Teaching with educational digital games	("digital educational games" OR "educational digital games" OR "electronic educational games" OR "educational electronic games" OR "pedagogical digital games" OR "educational digital learning games" OR "electronic games for education" OR "digital games for learning" OR "computational educational games" OR "computational learning games" OR "interactive digital games for education" OR "educational games" OR "educational gaming" OR "digital educational gaming" OR "pedagogical digital gaming" OR "serious games" OR "educational serious games" OR "digital games for teaching")
P AND I	(Elementary school) AND (educational digital games)	("elementary education" OR "primary education" OR "basic education" OR "fundamental education" OR "elementary school" OR "primary school" OR "compulsory education" OR "basic schooling" OR "mandatory schooling" OR "early school years" OR "initial years of elementary education" OR "final years of elementary education" OR "first stage of basic education" OR "second stage of basic education" OR "early childhood education" OR "lower school" OR "grade school") AND ("digital educational games" OR "educational digital games" OR "electronic educational games" OR "educational electronic games" OR "pedagogical digital games" OR "educational digital learning games" OR "electronic games for education" OR "digital games for learning" OR "computational educational

		games" OR "computational learning games" OR "interactive digital games for education" OR "educational games" OR "educational gaming" OR "digital educational gaming" OR "pedagogical digital gaming" OR "serious games" OR "educational serious games" OR "digital games for teaching")

Source: The authors (2025).

Besides combining descriptors using Boolean operators, you can apply filters and limits to your search, which vary depending on the database. In the Scopus database, for example, you can apply filters related to the year of publication, document type, and language. The limits – related to the descriptors used – applied in Scopus are the search for terms in the abstract, title, author, and keyword fields. Considering that the application of filters and limits tend to reduce the number of references found, we recommend that their use be restricted to specific situations. For example, if a particular approach emerged from a clearly identifiable time frame, such as the year 2000, it may be appropriate to restrict searches to references published from that period onwards. Nevertheless, considering potential errors in filtering and the consequent risk of eliminating references of interest, it is essential that researchers use such resources with caution and discretion.

5.3.2 Gray literature

Gray literature is usually defined as the set of documents and research materials that are not published through conventional commercial or editorial channels, making them difficult to locate and retrieve. This type of literature includes theses, dissertations, technical reports, conference proceedings, government documents, and other types of institutional publications (Farace & Frantzen, 1998). Its importance for systematic reviews lies in the fact that studies with negative results are often found in this type of source. The exclusion of gray literature can therefore contribute to meta-bias through publication bias, compromising the validity of the review findings.

Some databases, such as Scopus, partially incorporate references from gray literature. Another relevant source of Brazilian publications is the Digital Library of Theses and Dissertations (BDTD).

5.3.3 Manual search

In addition to electronic searches, the systematic review protocol may establish manual searches. It is recommended that the researcher explicitly state that manual searches will be conducted in the reference lists of the included studies and, ideally, in specialized journals or other systematic reviews already published on the topic of interest. (Lefebvre *et al*., 2025).

5.3.4 Other ways to build electronic searches (search hedges)

Search hedges consist of standardized search strategies designed to be used in conjunction with a search by subject to retrieve eligible studies. In other words, it is a search that has been constructed, tested, and refined to retrieve references based on a specific characteristic. An example of a search hedge is the search used to identify studies with the term "child in the Medline database: Infant[MeSH] OR Infant* OR infancy OR Newborn* OR Baby* OR Babies OR Neonat* OR Preterm* OR Prematur* OR Postmatur* OR Child[MeSH] OR Child* OR Schoolchild* OR School age* OR Preschool* OR Kid or kids OR Toddler* OR Adolescent[MeSH] OR Adoles* OR Teen* OR Boy* OR Girl* OR Minors[MeSH] OR Minors* OR Puberty[MeSH] OR Pubert* OR Pubescen* OR Prepubescen* OR

Pediatrics[MeSH] OR Paediatric* OR Paediatric* OR Peadiatric* OR Schools[MeSH] OR Nursery school* OR Kindergar* OR Primary school* OR Secondary school* OR Elementary school* OR High school* OR Highschool* (Boluyt, Tjosvold, Lefebvre, Klassen & Offringa, 2008).

Furthermore, there are tools that use text mining techniques to support the selection of search terms. These tools analyze texts indicated by the author of the systematic review and suggest combinations of descriptors capable of retrieving references similar to the source text. (Lefebvre *et al*., 2025).

5.4 Step four: describe the study selection process

The selection of studies corresponds to the stage in which the references located in the search stage will be evaluated for their relevance to answering the research question of the systematic review, based on the eligibility criteria established in section 5.2.

Essentially, this procedure involves two subsequent steps: (1) reading titles and abstracts and (2) reading the full text of studies that were not excluded in the initial step. In this section, the researcher should describe in detail how the selection process will be conducted in each of these phases, indicating, for example, whether the screening will be performed by two independent reviewers, both in the title and abstract stage and in the eligibility assessment based on the full texts. Ideally, it should be specified that the selection of studies will be carried out by two independent reviewers, since this procedure reduces the risk of undue exclusions of potentially relevant studies. (Higgins *et al*., 2024).

Systematic reviews aim to compile results from studies, not publications. A single study may have more than one related publication, either due to duplication of different journals, dissemination of partial results, or publication in different media (e.g., abstract at an event and full article in a journal). It is therefore essential that researchers avoid including multiple publications from the same study, as this compromises the validity of the review results. The duplicate inclusion of articles from the same study, for example, would give excessive weight to a single set of data, distorting the final synthesis by duplicating observations with identical characteristics.

The researcher must also define whether the selection of studies will be operationalized using a reference manager, a specific program, or even a spreadsheet, depending on the tools available. However, for the eligibility verification stage based on the full texts, we suggest using a preformatted spreadsheet containing guiding questions to complete the eligibility analysis of the studies. In the example presented, questions such as “Was the intervention group subjected to educational digital games?” and “Was the intervention applied to the teaching in algebra?” help the evaluator determine, in a systematic and transparent manner, whether the study will be “included” or “excluded” from the review.

5.5 Step five: Define how the methodological quality of the included studies will be assessed

The methodological quality of the included studies is analyzed by assessing the risk of bias in each study. This step allows for the examination of the reliability, internal validity, and relevance of the synthesized evidence by identifying potential sources of bias, methodological limitations, and the robustness of the methods used in individual studies, contributing to the interpretation and application of the results found. The researcher must specify whether the risk of bias assessment will be conducted by independent reviewers.

There are several instruments available to assess the risk of bias in a study, and the choice of tool should be appropriate to the type of design of the studies included. In this sense, it is essential that the selected instrument is compatible with the methodological characteristics of the studies analyzed. The Cochrane Collaboration[5] and the Joanna Briggs Institute (JBI)[6] provide different instruments to assist in this choice.

Experimental studies, for example, can be assessed using the *Risk of Bias* 2.0, developed by Cochrane (RoB 2.0), or the revised JBI instrument for assessing risk of bias in randomized clinical trials, among others. The methodological quality of quasi-experimental studies can be analyzed using the "*Risk of Bias of Non-randomized Studies – of Interventions*" (ROBINS-I) tool, the *Newcastle-Ottawa Scale*, or the JBI instrument for analyzing risk of bias in quasi-experimental studies, among others. In the case of reviews that include a combination of experimental and quasi-experimental studies, Cochrane recommends using RoB 2.0 to assess all studies.

The results of this step allow us to verify, for example, whether findings from studies with low methodological quality (high risk of bias) differ from those observed in studies with methodological quality (low risk of bias). This analysis can be performed using procedures such as subgroup analysis during qualitative or quantitative synthesis. Naturally, when the available evidence is based mainly on studies of low methodological quality, the results of the review should be interpreted with greater caution.

It is recommended that the methodological quality assessment be conducted by at least two independent reviewers, with disagreements resolved by consensus or through the participation of a third reviewer (Higgins *et al*., 2024).

5.6 Step six: define the data to be extracted from each study

At this stage of the protocol, the author should specify all the data that will be collected from the studies included in the review. This definition should be carefully structured to ensure that all the information necessary to answer the research question is adequately collected.

Although there are specific programs to support data extraction, their use depends on prior familiarization and, eventually, restricted access. However, it is possible to extract data using a preformatted spreadsheet in programs such as Microsoft Excel(R) or LibreCalc. These preformatted spreadsheets should be structured to facilitate standardized information collection. We suggest that the spreadsheet contain general data about the study, such as authors, year of publication, research title, country of origin, and methodological design, in addition to specific information of interest for the systematic review. This information may include population characteristics (sample size, age group, educational level, among others), intervention group data (description, duration, and frequency), control group data (description of the intervention or indication of the absence of intervention, for example), and evaluated outcomes (pre-test and/or post-test results, evaluation method, among others).

In operational terms, depending on the complexity and amount of data to be collected, it is possible to use a single preformatted spreadsheet to apply eligibility criteria, extract data, and assess risk of bias.

5.7 Step seven: write and publish the research protocol

Upon the completion of steps 5.1 to 5.6, the researcher will have gathered all the

[5] https://www.riskofbias.info/.
[6] https://jbi.global/critical-appraisal-tools.

information necessary to prepare the research protocol, which must be written and published. PRISMA provides a specific extension for systematic review protocols, PRISMA-P 2015, which contains a checklist of 17 essential items to be included. This tool is an important reference for ensuring the transparency, standardization, and methodological quality of the protocol, and is available on the official PRISMA website.

Several platforms and journals allow the registration or publication of systematic review protocols. In the field of education, two free platforms stand out: Campbell Collaboration and Open Science Framework Registries (OSF Registries). Campbell Collaboration requires that the publication of the protocol be accompanied by a commitment to conduct and publish the complete systematic review on the platform itself, aiming to ensure that the registered protocols are brought to completion, promoting transparency and scientific rigor. On the other hand, OSF Registries offers greater flexibility, allowing protocols to be published independently, without the obligation to conduct or publish the complete review on the same platform, which can be advantageous for researchers who wish to document the study design and subsequently publish the systematic review in another journal or platform.

It should be noted that the protocol for a systematic review allows for the planning and documentation of the methods that will be used in the review. It is essential because it allows researchers to carefully plan all stages of the review and anticipate potential problems. The protocol serves as a guide against possible arbitrary decisions made throughout the research, since the procedures are defined and recorded in advance. Thus, all steps described in the protocol must be strictly followed, and any subsequent modifications must be properly documented and justified. (Shamseer *et al*., 2015).

5.8 Step eight: Search and selection of studies

After publishing the research protocol, the next step is to search for references in electronic databases, using the previously defined descriptors and Boolean operators. It is essential to record the date on which the search was performed, as well as the number of references retrieved from each source, since this information must be included in the final review report. In the example presented, when executing the search strategy described in Table 3 in the Scopus database, 830 documents were identified on April 3, 2025.

The references obtained must be imported into a bibliographic reference manager, observing the file formats compatible with the chosen manager. Mendeley, for example, accepts files in BibTeX (.bib), RIS (.ris), EndNote XML (.xml), PubMed (.nbib), RefWorks (.ref) formats. In general, databases allow references to be exported in files with one of these formats. Therefore, after choosing a reference manager, you must check the formats it accepts before starting the import.

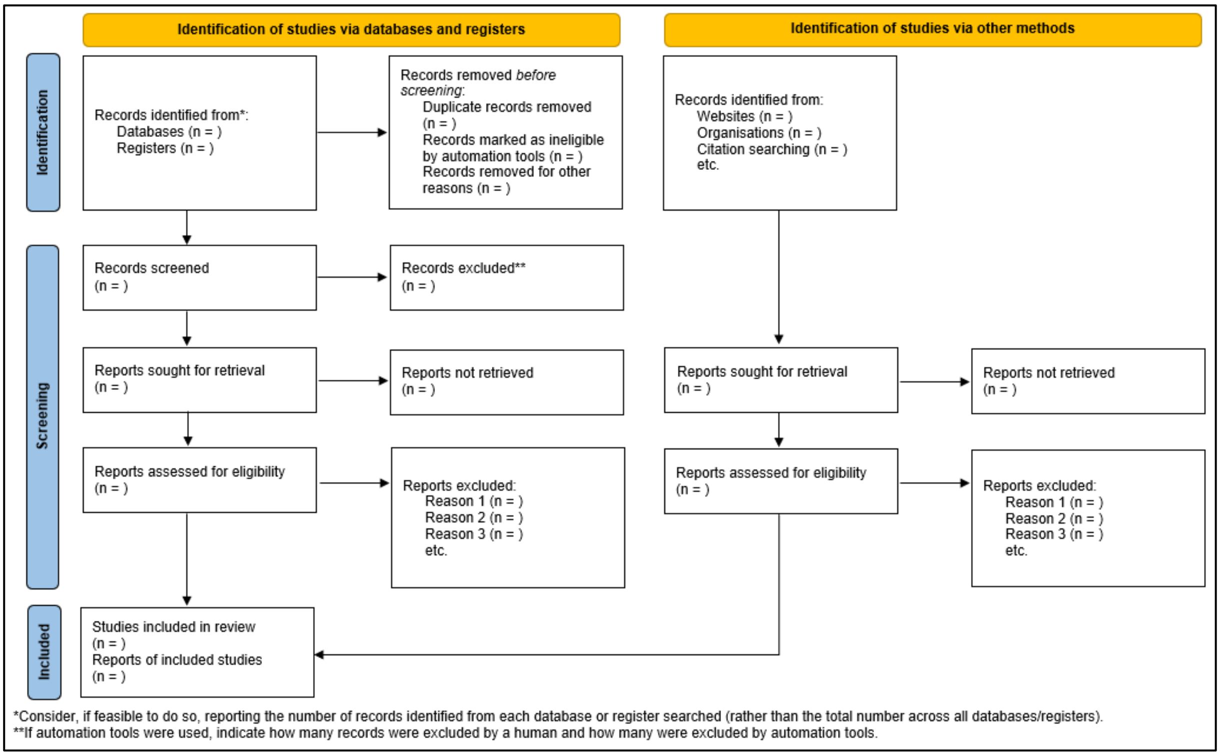
After importing, it is necessary to identify and exclude duplicate references using the managers' own features. Zotero, for example, has a tab called "Duplicate Items", in which references with a high degree of similarity are automatically grouped together. It is up to the researcher to decide whether the records should be merged into a single consolidated entry or kept as separate references.

Once the duplication removal stage is complete, the references should be exported to a spreadsheet containing information such as title, authors, place of publication, year of publication, and other relevant data. This information should then be transferred to a preformatted spreadsheet in order to begin selecting studies based on eligibility criteria.

The first phase of selection consists of reading the titles and/or abstracts, excluding those that are clearly ineligible. For this stage, a cautious approach should be adopted to avoid premature exclusion of relevant studies; in case of doubt, the study should be retained for the second phase, which corresponds to reading the full text. When evaluating the full texts, the reviewer should systematically verify whether the study meets all the eligibility criteria defined in the protocol.

The recommendation is that the selection process be conducted by at least two independent reviewers in order to ensure greater methodological rigor and reliability. In addition, it is essential to record the reasons for excluding studies during the full-text reading phase, as this information supports the preparation of the PRISMA Flow Diagram, a flowchart that summarizes all the stages of searching and selecting studies included in the systematic review, as shown in Figure 1:

Figure 1: PRISMA 2020 – Flowchart for new systematic reviews that include searches in databases, protocols, and other sources



Source: Page *et al.* (2021).

The flowchart should indicate the number of references identified in each data source, the references excluded at each stage of the selection process, and the respective reasons for exclusion at the eligibility assessment stage, based on analysis of the full text. In addition, it is necessary to report whether studies were identified by other means, such as manual searching of the reference lists of included studies. The editable version of PRISMA Flow, as shown in Figure 1, can be accessed on the official PRISMA website.

5.9 Step nine: data extraction and analysis of the methodological quality of the included studies

Data extraction and methodological quality analysis of studies consist of systematically collecting information contained in the studies included in the review. At this stage, the reviewer reads the texts in their entirety in search of the necessary information, following the parameters defined in the preformatted spreadsheet, as described in section 5.6. The reviewer must be attentive to the context and methodological design of each study to ensure that the extracted data are accurate.

It is recommended that this stage be conducted by at least two independent reviewers, with any disagreements decided by consensus or a third reviewer. The participation of more than one independent reviewer reduces the risk of errors in data extraction or omission of relevant information, thus ensuring the validity and reliability of the collection process.

5.10 Step ten: summary and interpretation of results and writing of the review

The final stage of the systematic review consists of summarizing and interpreting the results, as well as writing the final manuscript. Data synthesis can be performed qualitatively and quantitatively (meta-analysis), depending on the characteristics of the included studies.

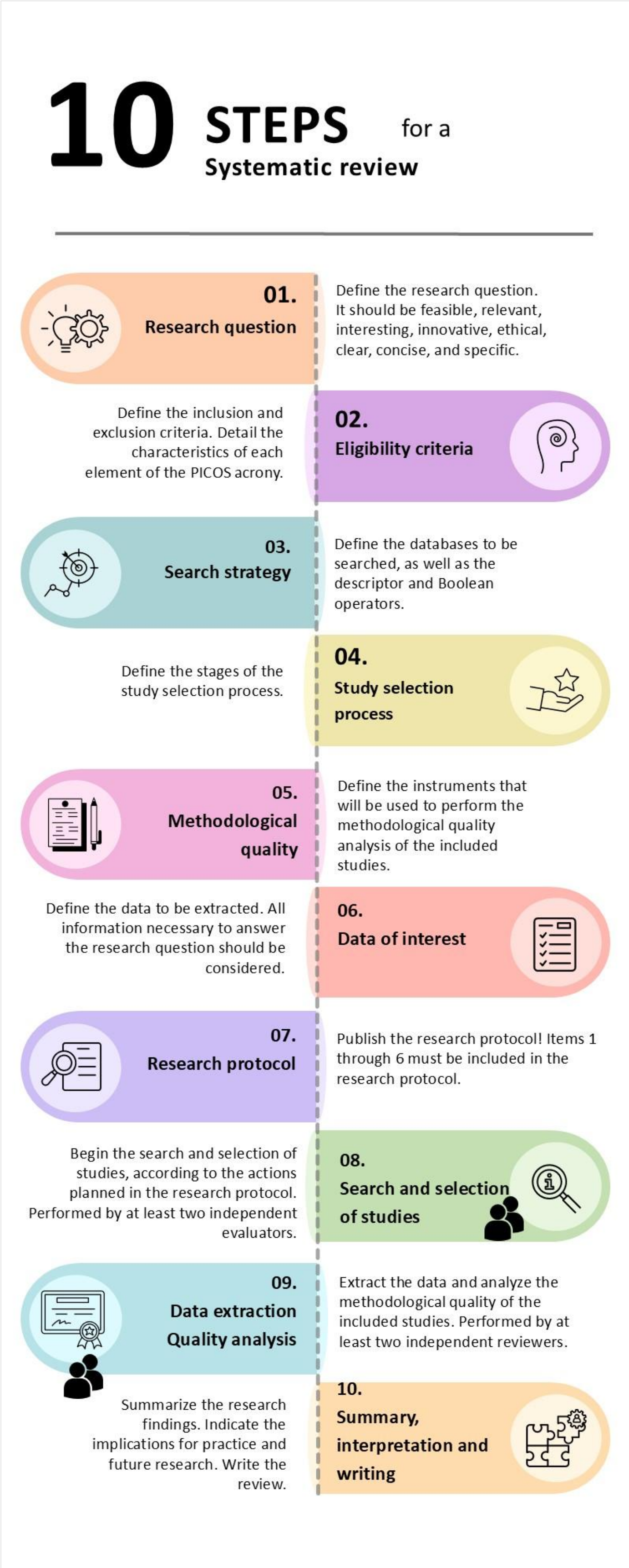
Qualitative synthesis is limited to cases where the data are highly heterogeneous or cannot be grouped numerically. In such situations, the results are organized and described narratively, highlighting similarities, differences, and patterns observed between studies. On the other hand, quantitative synthesis occurs when the studies are sufficiently similar to each other in terms of methodology, interventions, and outcomes. In this case, the results are grouped, and statistical methods are used to provide an overall estimate of the effect.

When interpreting the results, it is essential to provide an analysis of the practical relevance of the findings, highlighting their possible implications for the contexts of application. The limitations of the included studies should also be recognized and analyzed, explaining how these limitations may influence the results and the robustness of the conclusions of the systematic review.

The conclusion of the review should present the implications of the findings for both practice and the advancement of scientific knowledge, indicating the applicability of the results in different contexts and signaling any gaps that require further investigation. This conclusive analysis contributes to consolidating the scientific value of the review and guiding new paths for knowledge production.

Finally, we reinforce the recommendation to use PRISMA as a guide for writing systematic reviews. As already mentioned in section 4.1, PRISMA helps researchers report their research findings transparently, ensuring the robustness and reproducibility of the results. Figure 2 summarizes the 10 steps required to conduct a systematic review.

Figure 2: Steps for conducting a systematic review.



Source: The authors (2025).

6 Implications and challenges of systematic reviews in mathematics education

Mathematical education can be understood as a science, art, and social practice that dialogues with areas such as philosophy, mathematics, psychology, sociology, history, anthropology, semiotics, economics, and epistemology. Although situated at the intersection of multiple scientific fields, it has its own questions and problems and cannot be reduced to the simple specific application of these other areas (Fiorentini & Lorenzato, 2006).

We can affirm, based on Dario Fiorentini's interview with Maia (2023), that Brazil has a consolidated scientific community of researchers in Mathematics education, recognized nationally and internationally. This community is organized into 15 Working Groups (WGs) focused on research in this field, in addition to its activities in SBEM, whose scope goes beyond the limits of these groups. Each WG has contributed to the deepening and delimitation of specific lines of investigation, producing relevant results from different epistemological and methodological perspectives.

It is precisely in this context of theoretical diversity and expansion of scientific production that the need arises for methods capable of organizing, synthesizing, and critically evaluating the knowledge accumulated in the field. Systematic reviews meet this demand by offering a rigorous analysis of the existing literature on pedagogical practices, theories, and teaching methods, helping to identify educational interventions that positively influence mathematics learning.

As discussed throughout this paper, conducting a systematic review requires a meticulous process that demands effort and coordinated engagement from an entire team. Therefore, the success of a systematic review depends not only on the methodological rigor adopted, but also on the dedication, collaboration, and constant updating of the tools and approaches used throughout the different stages of the process.

Systematic reviews are also essential in guiding future research in the field, highlighting gaps in knowledge and pointing to new investigative directions. Based on the analysis of existing studies, they help direct research efforts toward under-explored areas or methodologies that need further investigation. Thus, systematic reviews not only synthesize available knowledge, but also serve as a starting point for the advancement of Mathematics education.

However, conducting systematic reviews in this area presents specific challenges. Silva, Curi and Schimiguel (2017) point to the difficulty of identifying the methodological approaches adopted in Mathematics education studies. According to the authors, "research in this field, published in the mapped journal, uses a variety of methods that are sometimes not easily identified, even after a more in-depth reading of the entire work" (Silva *et al*., 2017, p. 692). Considering that the quality of a systematic review is directly related to the quality of the studies included, it is essential that researchers in Mathematics education conduct primary studies with clear, rigorous, and transparent methodologies, thus ensuring the reliability and robustness of the evidence produced.

Moreover, in the field of education, there is a diversity of educational contexts. It is common for studies to involve different cultural, socioeconomic, and pedagogical contexts, which can directly influence the results obtained. In this sense, interventions and strategies that prove effective in a given context may not have the same relevance or applicability in other scenarios. Thus, the applicability of the evidence produced may be limited by contextual specificities, such as the diversity of resources, pedagogical practices, and student characteristics. This imposes the need for conclusions from systematic reviews to be adapted to the specific contexts in which they are implemented.

Furthermore, research in Mathematics education often adopts qualitative designs, which makes it equally pertinent to conduct systematic reviews of a qualitative nature. From this perspective, the procedures presented in this article can be extended to this type of review, with differences mainly in the mnemonic adopted for formulating the research question, the instruments used to assess methodological quality, and the strategies employed for analyzing and synthesizing qualitative data.

7 Conclusions

This article aimed to clarify the main step necessary for conducting systematic reviews of effectiveness, providing methodological support to researchers interested in adopting this approach. Throughout the text, we present conceptual and procedural elements that guide the reader in understanding what characterizes a systematic review, as well as guidelines for conducting one, with an emphasis on essential methodological aspects and the challenges inherent in this practice, illustrated by examples and specific applications for research on Mathematics education.

In the specific context of Mathematics education, the application of this methodology requires special attention to specific nuances, such as the precise definition of inclusion and exclusion criteria, the careful selection of databases, and the critical interpretation of the results obtained. These factors are essential to ensure that the systematic review produces a reliable and rigorous mapping of existing knowledge on a given topic.

Thus, we hope that this work will contribute to the dissemination of good practices in conducting systematic reviews in the field of Mathematics education, encouraging the development of more structured and evidence-based research, favoring more grounded professional practice supported by evidence accessible to educators and education managers.

Acknowledgements

This article was carried out with the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) / PROAP.

Conflicts of interest

The authors declare that there are no conflicts of interest that could influence the results of the research presented in this article.

Data Availability Declaration

The article was based on a theoretical study, so there is no data to be made available.

References

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (2024). *JBI Manual for Evidence Synthesis*. JBI.

Bhatt, A. (2010). Evolution of clinical research: A history before and beyond james lind. *Perspectives in Clinical Research*, *1*(1), 6.

Blaizot, A., Veettil, S. K., Saidoun, P., Moreno-Garcia, C. F., Wiratunga, N., Aceves-Martins, M., & Laig, N. M. (2022). Using artificial intelligence methods for systematic review in health sciences: A systematic review. *Research Synthesis Methods*, *13*(3), 353–362. https://doi.org/10.1002/jrsm.1553

Boluyt, N., Tjosvold, L., Lefebvre, C., Klassen, T. P., & Offringa, M. (2008). Usefulness of systematic review search strategies in finding child health systematic reviews in MEDLINE. *Archives of Pediatrics & Adolescent Medicine*, *162*(2), 111–116.

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation and the Health Professions*, *25*(1), 12–37. https://doi.org/10.1177/0163278702025001003

Clarke, M., & Chalmers, I. (2018). Reflections on the history of systematic reviews. *BMJ Evidence-Based Medicine*, *23*(4), 121–122. https://doi.org/10.1186/s13643-018-0833-3

Cummings, S. R., Hulley, S. B., & Browner, W. S. (2013). Conceiving the Research Question and Developing the Study Plan. *Designing Clinical Research*, 14–22.

Cumpston, M., Flemyng, E., Thomas, J., Higgins, J. P. T., Deeks, J. J., & Clarke, M. . (2024). Chapter I: Introduction. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.5*.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Elsevier. (26 de março de 2025). Resultados da busca para (TITLE-ABS-KEY (education) AND TITLE-ABS-KEY (mathematics OR math OR mathematical)). Scopus. Disponível em https://www.scopus.com HYPERLINK "https://www.scopus.com"

Farace, D. J., & Frantzen, J. (1998). Perspectives on the Design and Transfer of Scientific and Technical Information. *Proceedings of the Third International Conference on Grey Literature*.

Fiorentini, D., & Lorenzato, S. (2006). *Investigação em Educação Matemática: percursos teóricos e metodológicos*. Campinas, SP: Autores Associados.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.3102/0013189x005010003

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. . (2024). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5*.

Lasserson, T. J., Thomas, J., & Higgins, J. P. T. (2022). Chapter 1: Starting a review | Cochrane Training. *Cochrane Database of Systematic Reviews*.

Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Metzen, M.-I., Noel-Storr, A., Paynter, R., Rader, T., Thomas, J., & Wielandorf, L. S. (2025). Chapter 4: Searching for and selecting studies. In V. A. Higgins, J. P. T. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & L. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.5.1*.

Maia, M. G. B. (2023). Dario Fiorentini: reflexões pessoais e profissionais que influenciam a comunidade de professores e futuros professores que ensinam matemática. *Educação Matemática em Revista*, *28*(78), 3–19.

McKenzie, J. E., Brennan, S. E., Ryan, R. E., Thomson, H. J., Johnston, R. V., & Thomas, J. (2019). Defining the criteria for including studies and how they will be grouped for the synthesis. *Cochrane Handbook for Systematic Reviews of Interventions*, 33–65. https://doi.org/10.1002/9781119536604.ch3

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet*, *354*, 1896–1900. https://doi.org/10.1016/s0140-6736(99)04149-5

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Goup. (2015). Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1). https://doi.org/10.1186/2046-4053-4-1

Mulrow, C. D. (1994). Rationale for systematic reviews. *British Medical Journal*, *309*(6954), 597–599. https://doi.org/10.1136/bmj.309.6954.597

O'Gorman, C. S., Macken, A. P., Cullen, W., Saunders, J., Dunne, C., & Higgins, M. F. (2013). What are the differences between a literature search, a literature review, a systematic review and a meta-analysis? And why is a systematic review considered to be so good? *Irish Medical Journal*, *106*(2 Suppl), 8–10

O'Rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, *100*(12), 579–582. https://doi.org/10.1258/jrsm.100.12.579

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan — a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(210). https://doi.org/10.1186/s13643-016-0384-4

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

Pearson, K. (1904). Report on Certain Enteric Fever Inoculation Statistics. *BMJ*, *2*(2288), 1243–1246. https://doi.org/10.1136/bmj.2.2288.1243

PICO Portal. (2025). *PICO Portal.* www.picoportal.org

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., Altman, D. G., Booth, A., Chan, A. W., Chang, S., Clifford, T., Dickersin, K., Egger, M., Gøtzsche, P. C., Grimshaw, J. M., Groves, T., Helfand, M., … Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ (Online)*, *349*(January), 1–25. https://doi.org/10.1136/bmj.g7647.

Silva, J. F., Curi, E., & Schimiguel, J. (2017). Um cenário sobre a pesquisa em educação estatística no boletim de educação matemática -BOLEMA, de 2006 até 2015. *Bolema - Mathematics Education Bulletin*, *31*(58), 679–698. https://doi.org/10.1590/1980-4415v31n58a08.

JBI SUMARI. (n.d.). JBI Resource Center. https://sumari.jbi.global/

The Campbell Collaboration. (2019). Campbell systematic reviews : policies and guidelines. In *Campbell Policies and Guidelines Series No. 1* (Issue 1).

The Cochrane Collaboration. (2024). *Review Manager (RevMan)*. revman.cochrane.org

Thomas, J., Kneale, D., McKenzie, J. E., Brennan, S. E., & Bhaumik, S. (2024). Determining the scope of the review and the questions it will address. *Cochrane Handbook for Systematic Reviews of Interventions*, 13–31. https://doi.org/10.1002/9781119536604.ch2

Veritas Health Innovation. (2023). *Covidence systematic review software*. www.covidence.org