MedienPädagogik
Zeitschrift für Theorie und Praxis der Medienbildung
www.medienpaed.com ISSN 1424-3636

Themenheft Nr. 54:
Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen
Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers und Michael Kerres

Methodological Approaches to Evidence Synthesis in Educational Technology

A Tertiary Systematic Mapping Review

Katja Buntins[1], Svenja Bedenlier[2], Victoria Marín[3], Marion Händel[4] und Melissa Bond[5]

[1] Universität Duisburg-Essen

[2] Friedrich-Alexander-Universität Erlangen-Nürnberg

[3] Universitat de Lleida

[4] Hochschule Ansbach

[5] University College London; University of Stavanger

Abstract
Evidence synthesis methods are becoming increasingly popular in the social sciences, particularly in the field of educational technology, where secondary research has grown exponentially in recent years. Although review studies provide insight into these methods, questions have been raised about their methodological rigor and transparency. This tertiary review analyzed transparency and reproducibility in the reporting of evidence synthesis methods in the field of educational technology across different types of reviews indexed in the Web of Science, ERIC, Scopus, Google Scholar, Dialnet, and FIS. Reviews were included if they were published in English, German, or Spanish; if they synthesized the use of educational technology within formal teaching and learning settings; and if they contained a methods section. A sample of 446 evidence syntheses were included for data extraction and synthesis in EPPI Reviewer, with systematic reviews, meta-analyses, and literature reviews selected for deeper analysis as the most widely used review types in the corpus. Indicators of replicability at critical stages of the review were identified and analyzed in the sample by review type (research question, search strategy, data extraction, and synthesis). The results show significant room for improvement of methodological transparency in data extraction and synthesis, with certain types of reviews showing lower scores than others on some indicators. The article concludes with recommendations for improving the methodological transparency and rigor of evidence synthesis in the field of educational technology.

Buntins, Katja, Svenja Bedenlier, Victoria Marín, Marion Händel, and Melissa Bond. 2023. «Methodological Approaches to Evidence Synthesis in Educational Technology. A Tertiary Systematic Mapping Review». *MedienPädagogik* 54 (Research Syntheses): 167–191. https://doi.org/10.21240/mpaed/54/2023.12.20.X.

Methodische Ansätze zu Evidenzsynthesen in der Bildungstechnologie. Eine tertiäre Übersichtsarbeit

Zusammenfassung
Methoden der Evidenzsynthese werden in den Sozialwissenschaften immer beliebter, auch im Bereich der Bildungstechnologie, wo die Menge der durchgeführten Sekundärforschung exponentiell zugenommen hat. Während Übersichten einen wertvollen Einblick in den Stand des Feldes geben können, wurden Fragen zur methodischen Strenge und Transparenz aufgeworfen. In dieser tertiären Übersichtsarbeit wird daher untersucht, wie transparent und reproduzierbar die Berichterstattung über die Methoden der Evidenzsynthese im Bereich der Bildungstechnologie in den verschiedenen Arten von Übersichtsarbeiten ist. Zu diesem Zweck wurde eine kritische Analyse einer Stichprobe von 446 Evidenzsynthesen in drei Sprachen (Englisch, Deutsch und Spanisch) durchgeführt. Systematic Review, Meta-Analyse und Literaturreview als die am häufigsten verwendeten Review-Typen wurden für die weitere Analyse. In der Stichprobe wurden Indikatoren für die Replizierbarkeit in kritischen Phasen der Überprüfung identifiziert und nach Überprüfungstyp (Forschungsfrage, Suchstrategie, Datenextraktion und Synthese) analysiert. Die Ergebnisse zeigen, dass die methodische Transparenz bei der Datenextraktion und -synthese erheblich verbessert werden kann, wobei bestimmte Arten von Übersichten bei den Indikatoren schlechter abschneiden als andere. Es werden Empfehlungen zur Verbesserung der methodischen Transparenz und Strenge der Evidenzsynthese im Bereich der Bildungstechnologie gegeben.

1. Introduction

Evidence syntheses were originally used in the fields of medicine and psychology in the form of systematic reviews and meta-analyses to study effects of interventions across multiple studies. In the 1990s, evidence synthesis methods began to be used increasingly in the social sciences with the aim of conducting rigorous and transparent research to better inform policy, practice, and decision-making (Gough et al. 2020). In the early 2000s, as part of the move towards more evidence-based educational policy and practice, particularly in the United Kingdom, a lively debate emerged about the nature, place, and methodological quality of systematic research synthesis (e.g., Borrego, Foster, and Froyd 2014; Hammersley 2001). Since then, systematic reviews, meta-analyses, and other forms of evidence synthesis have grown at a rapid pace (Polanin, Maynard, and Dell 2017), gaining prominence in educational technology research (Kimmons and Rosenberg 2022). Evidence syntheses have been conducted, for example, on the impact of technology on learning (e.g., Tamim et al.

2011), on the effects of technology on student engagement (e.g., Bond et al. 2020), and on the benefits and challenges of using artificial intelligence in education (e.g., Crompton, Jones, and Burke 2022).

Depending on the aim and scope of the evidence synthesis, as well as on the primary studies included, various review approaches can be used. Sutton et al. (2019) categorized 48 review types into seven broader review families: traditional reviews, systematic reviews, review of reviews, rapid reviews, qualitative reviews, mixed methods reviews, and purpose-specific reviews.

Two of the most popular yet quite distinct types of reviews are the traditional literature review and the systematic review. A literature review is a comprehensive and critical analysis of the existing literature on a particular topic or research question. It involves identifying, evaluating, and summarizing relevant published research to provide an overview of the state of knowledge in the field. It does not aim to be systematic and complete in all respects, but to give an impression of the relevant findings and literature (Sutton et al. 2019). In contrast, a systematic review is a rigorous and transparent approach to reviewing the literature on a particular topic. It involves identifying, assessing and summarizing all relevant studies using predefined methods to minimize bias and improve reproducibility (Higgins et al. 2022). One form of systematic review is a meta-analysis, which is a quantitative summary of the results of multiple studies on a particular topic. It combines the results of individual studies using statistical methods to obtain a summary estimate of the effect size (Borenstein 2009). The main difference between a literature review and a systematic review is the degree of stringency. However, aside from certain review types in the traditional review family that often use a purposive sampling approach and less explicit reporting of methods (Sutton et al. 2019), all review types must follow the methodological requirement of transparency in terms of explicit and comprehensive reporting of their search approach.

1.1 Transparency and replicability in systematic reviews

Transparency and replicability are fundamental aspects of evidence synthesis research. Evidence syntheses systematically review, analyze, and synthesize the relevant primary literature to answer predefined research questions. The transparency of methods is of enormous importance for the results and helps to ensure the accuracy, reliability, and quality of the conclusions (Ioannidis et al. 2014). Clear and well-documented methods and results are important for repeating or replicating the synthesis and confirming and building confidence in the results (Higgins et al. 2022). Transparency and replicability are also beneficial in creating and promoting an open research culture (e.g., open science framework; OSF) that actively encourages collaboration and data sharing and can lead to new insights and discoveries.

One way to increase transparency and replicability in syntheses is by following guidelines such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al. 2009) or the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2022). By providing a framework for conducting evidence synthesis research and choosing methods for study selection, data extraction, analysis, and reporting, these guidelines help to ensure that the methods and results of evidence synthesis are well documented, standardized, and replicable. They also provide readers with a standardized framework for understanding how studies are conducted.

The present study addresses the overarching question of how evidence syntheses in educational technology research have been approached methodologically in the research to date and how transparent and replicable these syntheses are. In the following section, we first describe the method of evidence synthesis to provide the background against which a corpus of 446 evidence syntheses were assessed in terms of methodological approach (Buntins, Mulders, and Schröder 2023). We focus on how systematic reviews, meta-analyses, and literature reviews have been conducted to illustrate the range of possible syntheses used in educational technology research as reflected in the 379 syntheses in the corpus. The remainder of the article describes the approach used in this tertiary mapping review, presents the findings, and discusses their implications.

2. Synthesis of Evidence: General Process Steps

To evaluate the transparency and replicability of evidence syntheses in the field of educational technology, it is important to understand how each step of a synthesis contributes to transparency and replicability. Although evidence syntheses can be described in various ways, and the exact procedure depends on the type of synthesis chosen, there are commonly agreed stages (e.g., Pollock and Berge 2018; Atlam et al. 2020): (1) development of a research question (2) search and screening (3) coding and extraction of results, (4) synthesis.

2.1 Developing a research question

The research question defines the problems that will be addressed in the subsequent analysis and conclusions. Researchers can approach the development of research questions from a variety of angles, including examining intervention effects, exploring causal relationships, addressing research gaps, and synthesizing existing knowledge (Shea et al. 2007). In evidence synthesis, the formulation of precise and focused research questions plays a crucial role in guiding the entire research process and ensuring meaningful results. Approaches such as the PICO schema (Population, Intervention, Comparator, Outcome) have been developed to formulate research questions precisely. In this approach, the research question operationalizes the steps that follow, such as search and screening, based on the four constituent elements in the acronym. It aims to standardize the formulation of research questions and subsequent steps to maximize the accuracy of the results (Methley et al. 2014).

2.2 Search and screening

The search strategy process involves several decisions, including the definition of the search query, the selection of databases, and the use of other search strategies such as hand searching. In addition, this stage includes the documentation of study selection, for example, in the form of a flowchart (e.g., PRISMA, see Page et al. 2021), and determines what studies can be found (A. Campbell et al. 2018). Defining a search strategy for systematic reviews involves the challenge of identifying as many relevant studies as possible, while keeping the number of irrelevant studies as low as possible. In the literature, this dilemma is referred to as the "recall precision problem." The term "recall" or "sensitivity" refers to the proportion of studies found using the search strategy that are relevant to the overall pool of theoretically available studies (C. Cooper et al. 2018b). In contrast, "precision" describes the proportion of studies found with the search strategy that are relevant relative to all studies found, and allows the search strategy to be assessed by identifying relevant studies. Higher precision means a higher likelihood that the studies identified are relevant. To find as many relevant studies as possible, the search strategy should have a certain openness. This means that the criteria for study selection should be broad enough to ensure that potentially relevant articles are not overlooked (Eysenbach, Tuische, and Diepgen 2001). A search strategy that is too open, however, will inevitably lead to a reduction in accuracy, as a larger number of irrelevant studies will also be included in the results. Furthermore, to achieve the greatest possible retrieval, the selection of the right database is crucial. In studies exploring different approaches, Bramer, Giustini, and Kramer (2016) and Bramer et al. (2013) found that the choice of databases is highly dependent on the specific research question and stressed the importance of addressing these questions in educational science. Therefore, it matters how the grey literature is searched and what databases are used.

This stage includes defining not only the research question but also the inclusion and exclusion criteria applied to identify relevant studies for analysis during the screening process (Abrami, Cohen, and d'Apollonia 1988). Inclusion criteria refer to characteristics that must occur for a study to be included in the analysis, thereby defining the relevant sample of studies. Exclusion criteria describe the conditions under which a study is excluded from the analysis. They are often contrary to the inclusion criteria but may also include other aspects (Meline 2006).

2.3 Coding and data extraction

The development of a well-designed coding scheme is only possible once the authors have clearly defined the research question and have established a meaningful search strategy (C. Cooper et al. 2018a; H. Cooper and L. V. Hedges 1994; Glass et al., 1981; Glass and Smith 1979). Several coding approaches attempt to pre-structure the aspects that should be considered when coding for synthesis. Durlak and Lipsey (1991), for example, suggest coding by study context, methods, subjects, intervention, and effect, which echoes familiar approaches to PICO. Stock (2008) recommends considering the following factors when selecting a coding scheme: (a) the aims of the synthesis, (b) the quality of the information in the primary studies, (c) the ability of coders to reliably extract the required information, and (d) the associated coding burden. It is important to note that these approaches are primarily based on meta-analyses and that there is limited research on extraction and coding methods specifically for other synthesis approaches, particularly in the context of qualitative studies in education.

One indicator of clarity and accuracy in coding is inter-rater reliability (IRR), which is often overlooked in the planning phase. Various statistical methods can be used to measure IRR, including the calculation of simple agreement (Viera and Garrett 2005), Scott's π, Cohen's κ or Krippendorff's α (Lombard, Snyder-Duch, and Bracken 2002). IRR can be assessed at different stages, for instance, during the screening process or in relation to the coding process (Belur et al. 2018). Krippendorff (2018) highlights several crucial quality aspects, including stability over time, accuracy of the coding scheme, and reproducibility between different individuals. A low IRR may indicate not only low research quality but also weaknesses in the coding scheme or ambiguity of the construct (Kolbe and Burnett 1991).

Another important aspect is the quality of the primary studies. In systematic reviews and meta-analyses, it is generally assumed that the quality of the research synthesis decreases if the source studies are of low quality (Sotola 2022). Therefore, a quality assessment is often carried out before coding. There are various approaches to quality assessment, such as the Cochrane Risk-of-Bias (RoB 2) Tool (Higgins et al. 2011), Critical Appraisal Skills Program (CASP) (Singh 2013), Critical Appraisal Tools (Katrak et al. 2004) or the Newcastle-Ottawa Scale (NOS) (Hartling et al. 2013).

2.4 Synthesis

Researchers use various statistical and qualitative methods to analyze and integrate the collected data, depending on the type of review undertaken. This enables a comprehensive examination of the results and the derivation of meaningful conclusions. In systematic reviews, various methods of data analysis are used that are also common in primary studies. However, they must be modified and developed further in some cases. Researchers attempt to identify recurring themes, patterns, and concepts in the literature in various ways to gain a comprehensive understanding of the topic (Baumeister and Leary 1997). Qualitative techniques such as thematic analysis are often used to categorize and interpret qualitative data to gain insights into the underlying meanings and perspectives in the studies (Braun and Clarke 2006). In contrast, meta-analyses quantify the results of multiple studies (Higgins et al. 2022), combining specific effect sizes or outcome measures to produce a summary estimate of the magnitude and direction of the observed effect (Borenstein et al. 2010).

3. Research Questions

In order to understand the methodological approaches currently used in educational technology evidence syntheses, and given the increasing prevalence of syntheses in the field as well as the lack of methodological guidance on how to conduct them (e.g., Zawacki-Richter et al. 2022), this article presents a critical analysis of a sample of evidence syntheses and seeks to answer the following questions:

RQ1: How transparent and comprehensible is the reporting of evidence synthesis methods in reviews in the field of educational technology?

RQ2: How many systematic reviews, meta-analyses, and literature reviews are fully replicable?

RQ3: Are there differences in replicability depending on the type of evidence synthesis (i.e., between systematic reviews, meta-analyses, or literature reviews?

4. Methodology

This tertiary mapping review (Kitchenham et al. 2009; Lai and Bower 2020) was conducted following explicit, pre-defined methods (Gough et al. 2012; Zawacki-Richter et al. 2020), and the reporting here is guided by the PRISMA guidelines (Page et al. 2021) for increased transparency.

4.1 Search strategy

4.1.1 Search strings

Three different search strings were developed based on the preliminary work of Bond (2020), Bond et al. (2020), and Marín et al. (2023). Marín et al. (2023) focused on forms of evidence synthesis and educational technology, using * for truncations where appropriate (see Table 1).

Evidence synthesis	(“systematic review” OR “scoping review” OR “narrative review” OR “meta-analysis” OR “evidence synthesis” OR “meta-review” OR “evidence map” OR “rapid review” OR “umbrella review” OR “qualitative synthesis” OR “configurative review” OR “aggregative review” OR “thematic synthesis” OR “framework synthesis” OR “mapping review” OR “meta-synthesis” OR “qualitative evidence synthesis” OR “critical review” OR “integrative review” OR “integrative synthesis” OR “narrative summary” OR “state of the art review” OR “rapid evidence assessment” OR “qualitative research synthesis” OR “qualitative meta-summary” OR “meta-ethnography” OR “meta-narrative review” OR “mixed methods synthesis” OR “scoping study” OR “systematic map”)
AND	
Ed Tech	“education* technolog*” OR “digital technolog*” OR “ICT” OR computer* OR “information and communication*” OR “digital media” OR “online learning” OR “blended learning” OR “remote teaching” OR “remote learning” OR “remote education” OR “mobile learning” OR “online education” OR “social media” OR “eLearning” OR “learning analytics” OR “Facebook” OR “technology” OR “e-Learning” OR “multimedia learning” OR “media in education” OR “interactive learning environments” OR “computer mediated communication” OR “virtual reality” OR “distance education” OR “human-computer interface” OR gamification OR “game-based learning” OR “distance learning” OR “learning environments” OR “technology integration” OR “multimedia/hypermedia systems” OR “intelligent tutoring system*” OR “flipped classroom” OR “flipped learning” OR multimedia OR “evaluation of CAL systems” OR MOOC* OR “computer-supported collaborative learning” OR “distance education and telelearning” OR “serious game*” OR “learning management system*” OR “LMS” OR “CSCL” OR “m-learning” OR “human-computer interaction” OR “computer science education” OR “architectures for educational technology system” OR “distributed learning environment*” OR Moodle OR “online teaching” OR “technology-enhanced learning” OR “adaptive learning” OR “open educational resources” OR “OER” OR “technology enhanced learning” OR “digital technolog*” OR “virtual environments” OR “web-based learning” OR “video games” OR “augmented reality” OR “educational games” OR “massive open online course*” OR “computer-assisted instruction” OR “information and communication technolog*” OR “open education” OR “virtual learning environment*” OR “distributed learning” OR “learning technologies” OR “educational robotics” OR “computer-supported learning” OR “online educational videogames” OR “educational videos”
NOT	
Outside scope	Smoking OR clinic* OR pathology OR telemedicine OR telehealth OR inflammation OR patient* OR neurolog* OR disease* OR “mobile health”

Tab. 1: Search term in English.

The Spanish search term was developed based on the English search terms (see Table 2). However, the Spanish database used (Dialnet) has a search limit of 50 words, so the full search term used in English was not usable for this database and was reduced to more general terms for the three elements (evidence synthesis, EdTech, and education). The German search term was co-developed by an information scientist from the University of Erlangen–Nuremberg, as it required considerable adaptation to be used in the corresponding database (see Table 3).

Evidence synthesis	(“meta-análisis” OR “metanálisis” OR “metaanálisis” OR “metarevisión” OR “meta-revisión” OR revisión OR “síntesis cualitativa” OR “meta-síntesis” OR “metasíntesis”)
AND	
Ed Tech	(tecnologi* OR ordenador* OR computador* OR “TIC” OR “digital*”)
AND	
Education	(educa* OR aprend* OR enseña* OR docen*)

Tab. 2: Search term in Spanish.

Evidence synthesis	Review* ODER Synthes* ODER Meta-Analyse ODER Metaanalyse ODER Metanalysen ODER „narrative summary“ ODER Meta-Ethnographie ODER „scoping study“ ODER „systematische Übersichtsarbeit“ ODER Literaturstudie ODER Übersichtsarbeit ODER Meta-Synthese ODER „systematisches Literaturreview“ ODER Literaturüberblick ODER „systematische Übersicht“ ODER „Second-Order-Review“
AND	
Ed Tech	Bildungstechnolog* ODER Technolog* ODER IKT ODER ICT ODER computer* ODER Lerntechnolog* ODER “Informations- und Kommunikationstechnolog*“ ODER “augmentierte Realität*” ODER AR ODER “virtuelle Realität” ODER VR ODER Bildungsroboter ODER Bildungsrobotik ODER “Mensch computer Schnittstelle” ODER “Mensch computer Schnittstellen” ODER “intelligente Tutorensysteme” ODER „intelligentes Tutorensystem“ ODER “Architektur* für Bildungstechnolog*” ODER Moodle ODER Lernmanagementsystem ODER Lernmanagementsysteme ODER “Mensch-Computer-Interaktion” ODER “learning analytics” ODER LMS ODER Lernumgebung ODER Lernumgebungen ODER “multimedia System”

Tab. 3: Search term in German.

4.1.2 Search strategy and study selection

The searches were conducted in February and March 2022. As the author team is multilingual, studies in English, Spanish, and German were considered for possible inclusion. The English language databases and platforms searched were ERIC, Scopus, Web of Science, and Google Scholar. For Spanish studies, Dialnet was used, and for German-speaking countries, the FIS database was searched. These databases were selected because they provide broad coverage of the state of research suitable for evidence synthesis (Gusenbauer and Haddaway 2020), with the Spanish and German databases specifically selected because they are more comprehensive for the Spanish- and German-speaking contexts.

4.1.3 Screening and sampling procedures

The initial search yielded 9,050 English-language, 898 Spanish-language and 534 German-language references (see Fig. 1), which were imported into evidence synthesis software EPPI Reviewer (Thomas et al. 2023). After automatic removal of 3,207 duplicates, 7,275 references were identified for screening on title and abstract. The first 100 references were screened by four of the authors, achieving a moderate Fleiss kappa of k=0.60 (Landis and Koch 1977). Therefore, after reconciling differences, it was decided that pairs of authors would screen the remaining studies together to ensure substantial agreement. Articles were included if they were a form of evidence synthesis with a methods section; were focused on educational technology within a formal teaching and learning setting; and were not a workshop paper, poster, or editorial (see Table 4).

Given the large number of references to be screened on full text, a sample of English and Spanish articles was drawn from this corpus for further analysis. With the intention of drawing a sample that estimates the parameters of the population with a certain margin of error, we used methods for estimating sample size in the social sciences (Döring and Bortz 2017), stratifying by decade of publication and then drawing a random sample. Within the sampling, a margin of error of 5%, a percentage of 50%, and an alpha of 5% were assumed. A margin of error gives the range (in percent) by which the responses of your population may differ from those of the sample. The probability of committing an error to describe the attitude of the sample is equivalent to the alpha error in inferential statistics. Percentage refers to the size or probability of an effect. If there are no prior assumptions, this is set to 50%. Since we have numerous questions, we also proceeded in this way. If one expects a very large or small effect, then this number can be regulated higher or lower (Levy and Lemeshow 2011). The German studies were screened in their entirety, as it was assumed that many studies would not be suitable due to the change in the search term and that the total sample would otherwise be very small. After sampling, 734 studies remained that were screened on full text, with 446 studies considered for the synthesis (372 English-language, 61 Spanish-language, and 13 German-language studies) (see Figure 1).

This article specifically focuses on the three types of review previously described; literature reviews ("traditional review family"), systematic reviews and meta-analyses ("systematic review family"; Sutton et al. 2019), with these synthesis approaches being located on opposing ends of the spectrum of transparency and replicability, and at the same time being the most frequently conducted syntheses in the corpus. As such, 379 evidence syntheses were included in the final corpus for this article.

INCLUSION	EXCLUSION
Evidence synthesis	Primary research
Focuses on educational technology	No focus on educational technology
Education related (e.g., eAssessment, meta-analysis of experimental studies related to teaching and learning)	Not education related (e.g., student focus but about something in their private life – no connection to teaching and learning)
Journal articles, book chapters, reports, and conference papers (both full and short papers)	Workshop papers, poster contributions, editorials
Has a methods section	No methods section

Tab. 4: Inclusion and exclusion criteria.

4.2 Data extraction

In addition to operationalizing the research process research process (see Table 5), a range of data were extracted (see OSF[1] for the complete coding scheme), including publication and authorship information (e.g., publication type and name), review type (self-declared by the authors and informed by the typology of Sutton et al. 2019), specific educational and participant context (e.g., undergraduates, Health and Welfare), and EdTech thematic focus (e.g., blended learning). In regard to methodological characteristics, which are particularly pertinent to the present analysis, data were extracted based on an adaptation of both the Database of Abstracts and Reviews of Effects (DARE) tool (Centre for Reviews and Dissemination 1995), which has been used in previous tertiary reviews (e.g., Lai and Bower 2020), and the AMSTAR 2 tool (Shea et al. 2017). When interpreting these categories, it should be noted that replication is only possible if reporting was complete (yes). If only examples are given, this helps the reader to better understand the scope of the study but does not help to verify or update the results of the study.

1 https://osf.io/83vp7/

Stage	Operationalization	Categories
Method Section	How many studies did not report the methodology?	(Yes/No)
Research Question	Was a research question asked?	(Yes/No)
Search strategy	Were the search strings reported? Were inclusion and exclusion criteria specified? Were the databases or other search strategies specified? Were the years of publication indicated? Was a flow chart or PRISMA chart shown?	(Yes/Examples/No) (Yes/Examples/No) (Yes/No) (Yes/No) (Yes/No)
Data extraction	Was the inter-rater reliability reported? Was a coding scheme reported? Was a quality assessment reported? Was there a report or check for publication bias?	(Yes/No) (Yes/Examples/No) (Yes/No) (Yes/No)
Synthesis	Did they report how the data were synthesized in the methods section?	(Yes/Examples/No)

Tab. 5: Extracted methodological data.

4.3 Synthesis

The results were analyzed univariately and bivariately. Only descriptive results are given. Percentages are rounded to a full number and should be interpreted in an ordinal comparison as the exact percentage is not reliable due to sampling, selection, and researcher bias.

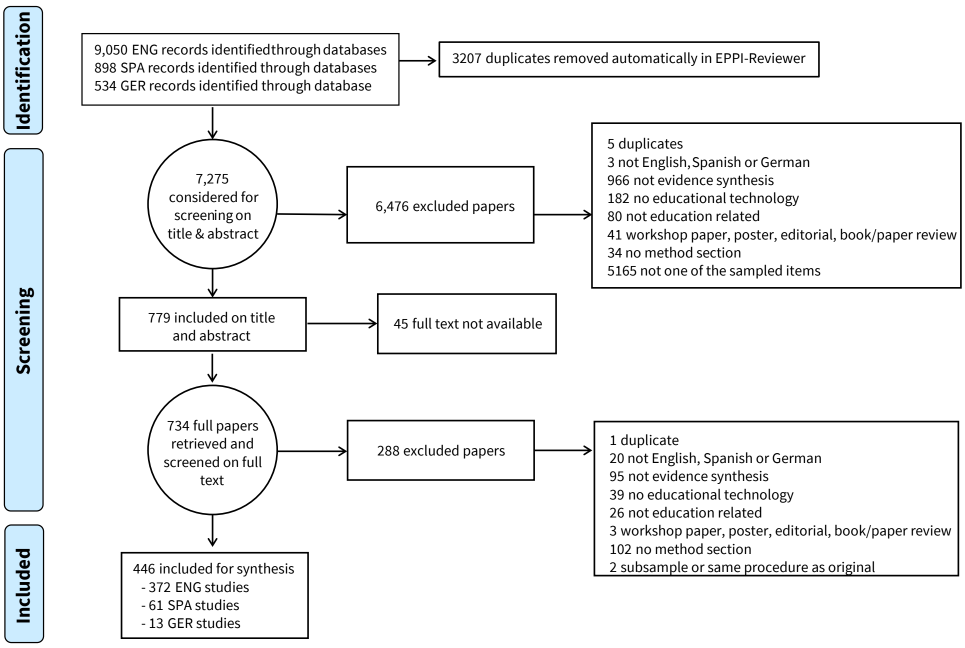

Fig. 1: PRISMA flow diagram.

4.4 Methodological limitations

Limited resources required a sampling procedure for the English and Spanish corpus. Although the sample was as representative as possible, the limitation that not all publications were included in the sample must be acknowledged. Furthermore, the search terms used for the German and Spanish databases differed from the English ones due to language issues and limitations in the databases' metadata and search functions, which may have led to relevant records being overlooked.

The review was conducted by five reviewers covering three languages, and although attempts were made to reduce inconsistency and bias, the human element remains. Data were only extracted and coded if the information could be found in the article itself, that is, no further research was done to determine, for example, the author's discipline or country of origin if this information could not be found in the manuscript itself. Furthermore, while a protocol was not pre-registered with an official systematic review registry such as Prospero, the full search details can be found on the OSF.[2] In the future, the authors will be able to use the newly created International Database of Education Systematic Reviews[3], which is now accepting protocols for all education disciplines.

4.5 Features of the study

Based on the authors' chosen term for the synthesis conducted, 28 different types of reviews were identified. According to Sutton et al.'s (2019) classification of review families, all seven review family types were represented to varying degrees, with the systematic review family being by far the most common in our sample (n=331) (see Table 6). It should be noted that some studies declared the use of more than one review type.

Review type	absolute frequency	Percentage
Systematic review	210	47%
Meta-analysis	131	29%
Literature review	38	9%
Scoping review	18	4%
Review	17	4%
Critical review	13	3%
Integrative review	10	2%
Mapping review / systematic map	10	2%
Narrative review/synthesis	9	2%
Meta-synthesis	6	1%

2 https://osf.io/83vp7/.
3 https://idesr.org.

Review type	absolute frequency	Percentage
Not specified/others	5	1%
Bibliographic review	4	1%
State-of-the-art review	4	1%
Qualitative meta-analysis	2	0%
Qualitative review	2	0%
Qualitative systematic review/qualitative evidence synthesis	2	0%
Systematized review	2	0%
Comparative review	1	0%
Descriptive review	1	0%
Framework synthesis	1	0%
Library method	1	0%
Meta review	1	0%
Mixed studies review/mixed methods review	1	0%
Quasi-systematic review	1	0%
Rapid review	1	0%
Research synthesis	1	0%
Second-order meta-analysis	1	0%
Systematic umbrella review	1	0%
Thematic review	1	0%

Tab. 6: Self-reported types of verification.

5. Results

To answer the question of how transparent and replicable evidence syntheses are in educational technology research (RQ1), the steps of the evidence syntheses were operationalized (see Table 5). As shown in the flow chart (see Fig. 1), 102 studies did not include any methods section. In the present study, we also investigated differences between review types, but this could only be done for the coded reviews; not for the question of whether a methods section was present. Although this was an official exclusion criterion, it can be considered an indicator of particularly poor replicability. In the following, therefore, the analytical focus is on the three most often used approaches across the corpus: systematic reviews, meta-analyses, and literature reviews.

5.1 Replicability of the research question

For the replicability of the research question, we identified two indicators that operationalize the extent to which the research question and the definition of the study population are designed to be replicable. In total, 86% of reviews included research questions, goals, or objectives (n=384). The degree of precision varied considerably. This means that 14% of the reviews did not state any research questions, aims, or objectives in the published text, meaning that the procedure of analysis must be interpreted by the reader from the theory section. Inclusion and exclusion criteria define the framework of the population, meaning that they indicate which studies belong to the object of study and which do not. Sixty-two percent of the studies stated the inclusion and exclusion criteria in full (n=275); 29% gave extracts or examples (n=131); and 9% of the studies did not provide any such information (n=40).
In relation to the different types of review (see Table 7), major differences emerge. For example, the values for literature reviews are significantly lower than for the other two review types when inclusion and exclusion criteria are specified. There are no major differences between meta-analyses and systematic reviews. Two-hundred forty-eight studies stated both the research question or objective and the inclusion and exclusion criteria. This is 46% of the studies that can be replicated up to this point. For the three types of reviews, the results show differences between the types. For literature reviews, 32% (n=10) reported both the research question and the inclusion and exclusion criteria; for meta-analyses, 60% reported both (n=78); and for systematic reviews, 61% reported both (n=128).

	Literature review	Meta-analysis	Systematic review
Research question			
Yes	84%	86%	90%
Inclusion and exclusion criteria			
Yes, fully, and completely reported	29%	68%	66%
Only a few examples were given	32%	25%	28%
Total [%]	61%	93%	94%
Total	38	131	209

Tab. 7: Indicators for the replicability of the research question.

5.2 Transparency of the search strategy

As our study is concerned with the transparency and reproducibility of reviews, we only consider criteria that are important for this: namely, specifying the search string, the search period, the databases searched, and the presence of a PRISMA diagram (see Page et al. 2021). In 87% of reviews, an example of the search string or the complete search string was given, with 44% reporting the full search string (n=192) and 43% (n=196) providing examples. In 13% of the research syntheses, the search string was not provided. The proportion of reported search terms is highest for systematic reviews, followed by literature reviews and meta-analyses. This order remains the same even when incomplete data are included (see Table 8). This shows that even if, in many cases, no direct replication of the search is possible, the reader at least gets an idea of what is meant by the constructs searched for. Six literature reviews (16%), 24 meta-analyses (18%) and 72 systematic reviews (34%) are still replicable at this point.

	Literature review	Meta-analysis	Systematic review
Yes, completely reproducible	34%	30%	55%
Only examples were given	47%	50%	39%
Total [%]	82%	80%	93%
Total	38	130	209

Tab. 8: Indicators for the replicability of the search strategy (I).

At this stage of a review, there are three further replication criteria: specifying the databases searched, specifying the years of publication, and including a flowchart or PRISMA diagram. Most studies indicated the years on which the research synthesis was based (80%, n=355), and only 4% of studies did not indicate where the studies were found (n=19). Far fewer studies included a flowchart or PRISMA diagram (n=192, 53%). Due to the high values, the variance in publication years and search strategies is comparatively low. However, the year of publication was most frequently reported in systematic reviews, while was is reported much less frequently in the other two (see Table 9).

Ninety-nine of the research syntheses (15%) are still replicable when the review period is included. Looking at the three specific review types, this means that five of the literature reviews (13%), 19 meta-analyses (15%), and 62 systematic reviews (30%) are still replicable up to this point. This continues the trend that the highest transparency is found in systematic reviews. However, even at this stage, only a small proportion of the studies can be replicated. When the databases are specified, the possibility of replicating the reviews does not decrease further. The situation is different when flowcharts or PRISMA diagrams are used. Only one of the literature reviews (3%), only eight meta-analyses (6%), and 45 systematic reviews (21%) are still replicable at this point. So, after two of the commonly agreed review steps, only a fraction of the reviews are still transparent and replicable.

	Literature review	Meta-analysis	Systematic review
Publication years			
Yes	76%	75%	83%
Database			
Presented	84%	96%	99%
Flowchart or PRISMA			
Presented	16%	33%	52%
Total	38	130	209

Tab. 9: Indicators for the replicability of the search strategy (II).

5.3 Transparency of screening and data extraction

The main replication indicator in this section is reporting the coding scheme. The other three indicators are reporting the performance of a quality assessment; reporting, checking, or correcting publication bias; and reporting inter-rater reliability.

In 37% of studies, the full coding scheme for data extraction was presented (n=163), and in another 36%, at least some examples were presented (n=160). In total, 28% of studies did not state how the data were extracted (n=124). There are major differences between the three types of reviews here; in contrast to the first two steps, the possibility of replication is substantially higher in meta-analyses than in the other two. Only in the partial representation of the coding scheme is there a notable difference between these two (see Table 10). Only 27 studies can be replicated up to this step of the coding (6%). If this is transferred to the three types of review considered in more detail, there are no literature reviews (n=0), only six meta-analyses (5%), and 15 systematic reviews (7%). The trend reversal described can be seen very clearly here.

	Literature review	Meta-analysis	Systematic review
Yes, the complete scheme is provided	26%	56%	29%
Examples are given, but not the complete list	24%	31%	41%
Total [%]	50%	87%	70%
Total	38	130	209

Tab. 10: Indicators for the reproducibility of data extraction (I).

Twenty-one percent of studies reported inter-rater reliability in screening or coding of primary studies (n=92). Publication bias was reported, corrected, or estimated in 19% of the studies (n=86), although it should be noted that this is only mathematically possible for quantitative reviews. The aim of a quality assessment is to exclude studies with inferior methods. This is not equally necessary for all types of review (e.g., scoping reviews), which partly explains why a quality or critical appraisal was only carried out in 24% (n=105) of cases.

When looking at the three review types, there are clear differences. The values for meta-analyses are notably higher than for the other two categories, although in terms of quality assessment, there is a clear difference between systematic reviews and literature reviews. However, as described above, these differences can be seen in the conceptualization of the studies (see Table 11). Since these three criteria are not necessary for replication, but merely illustrate the transparency of the process, they are not considered in our path analysis for the proportion of full replicability.

	Literature review	Meta-analysis	Systematic review
Inter-rater reliability			
Yes	11%	37%	15%
Quality assessment			
Yes	5%	32%	26%
Publication bias			
Yes	8%	55%	9%
Total	38	130	209

Tab. 11: Indicators for the reproducibility of data extraction (II).

5.4 Replicability of the synthesis

The final step in the process of evidence synthesis as it pertains to systematic reviews or meta-analyses is the synthesis of the primary studies. Forty-two percent of the studies in this sample explicitly stated how the results of the primary studies were synthesized, and in another 23%, there was some mention of what was done but no explicit description of the procedure. In 35% of the studies, this aspect was not mentioned at all. It is precisely in this last step where the described reverse trend becomes clear. Here, meta-analyses have a very high proportion of replicability, while systematic reviews and literature reviews have rather low values. Although the values for systematic reviews are higher, the differences are not so great that one could speak of a strong difference here (see Table 12).

	Literature review	Meta-analysis	Systematic review
Yes, with exact specification of the me-thod	24%	82%	31%
It is mentioned, but not explicitly	21%	8%	27%
Total [%]	45%	90%	58%
Total	38	130	210

Tab. 12: Indicators for the replicability of the synthesis.

As can be derived from the findings presented above, complete replication is possible in a total of 16 studies (see Table 13). The trend has thus completely reversed, although the values are very low overall.

	Number of Studies
All studies (only systematic review, meta-analysis, and literature review)	**361**
Was a research question asked?	**319**
Were the search terms specified?	**138**
Were inclusion and exclusion criteria specified?	**120**
Were the years of publication indicated?	**99**
Were the databases or other search strategies specified?	**99**
Was a flow chart or a PRISMA diagram shown?	**66**
Was a coding scheme reported?	**27**
Did they report how the data were synthesized in the methods section?	**16**

Tab. 13: Flow Table of Replication.

6. Discussion

This study examined methodological approaches to evidence syntheses in the field of educational technology to answer the question of how transparent and reproducible the reporting of evidence synthesis methods is in reviews in the field of educational technology (RQ1) and whether transparency and reproducibility differ between systematic reviews, meta-analyses, and literature reviews (RQ2, RQ3). The aim was not to assess the quality of reviews in detail, but to capture what data were used and how they were reported based on the information given. By systematically studying the current state of evidence synthesis methods in the context of educational technology, we can more clearly determine whether evidence syntheses in this area need further methodological development.

Looking at the different types of reviews and addressing the first research question, clear differences can be identified. Meta-syntheses and narrative reviews or syntheses and critical reviews tend to have lower transparency rates in most areas.

While the latter are not a problem, as these types of reviews (from the family of traditional reviews) usually do not contain research questions and do not use inclusion and exclusion criteria, this is not the case for other review types from the family of qualitative systematic reviews. Meta-analyses (from the systematic review family) are generally average in terms of stating the research question and search strategy, but above average in terms of replicability in data extraction and synthesis. Nevertheless, the analysis showed that of 361 literature reviews, systematic reviews, and meta-analyses, only 16 were fully transparent and replicable.

It is noticeable that a completely missing methods section is a frequent exclusion criterion. For the secondary studies with a methods section, it can be said that more care needs to be taken in the reporting of the methods, especially regarding inclusion and exclusion criteria and the way the data were extracted and synthesized. In contrast to the analysis of 73 EdTech systematic reviews and meta-analyses by Lai and Bower (2020), where 81% of studies included the full inclusion and exclusion criteria, only 62% of this much larger sample did so. Specifically, this means that not only should a table of inclusion and exclusion criteria be included, but also that the exact coding scheme should be provided, either as an appendix or in an openly accessible repository such as ResearchGate, OSF, or Mendeley Data. It should also be apparent to the reader how this extraction can be replicated, which may mean adding additional descriptions to the codes to enable full understanding.

Our analysis makes it clear that there is a need for better methodological representation of the approach to evidence syntheses in educational technology. However, the lack of replication also suggests a lack of methodological guidance. The combination of these findings suggests that there needs to be a greater focus on the documentation and selection of research methods and on the research protocols themselves, as well as enhanced guidelines and professional development for researchers at all career levels. There are already good examples of meta-analysis and systematic review guidance, such as the PRISMA guidelines (Page et al. 2021), quality assessment frameworks (e.g., CASP 2022), and methodological guides for education (e.g., Alexander 2020; Pigott and Polanin 2020). The lack of methodological quality is consistent with other analyses of primary research on methodological quality in the field of educational technology research. For example, Bulfin et al. (2014) analyzed methodological approaches in educational technology research and found that most studies used simple procedures for uni- or bivariate relationships. Hew et al. (2019) came to similar conclusions about the use of theory in educational technology research, and Buntins, Kerres, and Heinemann (2022) found a significant need for improvement in the analysis of measurement instruments in primary studies using an evidence synthesis. This raises the question of whether a deeper understanding is needed in the methodological training of researchers and peer reviewers and editors.

6.1 Further research

We posit that evidence syntheses in education should develop need their own research methodology rather than continuing to rely on approaches used in health and welfare, or even in the social sciences, which may or may not be appropriate. Not only should such syntheses include a theoretical background, they should also be iteratively developed and empirically evaluated based on empirical reviews. This development should be evaluated in line with various quality parameters (e.g., inter-rater reliability). In our opinion, this requires discipline-specific bespoke evaluation criteria that consider the interdisciplinarity and fuzziness of concepts. Further investigation is also needed into the extent to which a lack of methodological standards in reporting leads to a limitation of research results.

6.2 Concluding remarks

Given the exponential growth in the use of evidence synthesis in educational technology, this study offers a valuable contribution to the field by drawing attention to applications of evidence synthesis methodology and by suggesting ways to increase methodological transparency and replicability. In doing so, it aims to improve the quality of evidence synthesis in the field of educational technology, although the recommendations may also be applicable and helpful in other disciplines. Future work by this research team will explore further important and relevant aspects of evidence synthesis methodology, such as the issue of language bias.

References

Abrami, Philip C., Peter A. Cohen, and Sylvia d'Apollonia. 1988. «Implementation Problems in Meta-Analysis». *Review of Educational Research* 58 (2): 151–79. https://doi.org/10.3102/00346543058002151.

Alexander, Patricia A. 2020. «Methodological Guidance Paper: The Art and Science of Quality Systematic Reviews». *Review of Educational Research* 90 (1): 6–23. https://doi.org/10.3102/0034654319854352.

Atlam, Hany F., Muhammad Ajmal Azad, Madini O. Alassafi, Abdulrahman A. Alshdadi, and Ahmed Alenezi. 2020. «Risk-Based Access Control Model: A Systematic Literature Review». *Future Internet* 12 (6). https://doi.org/10.3390/fi12060103.

Baumeister, Roy F., and Mark R. Leary. 1997. «Writing Narrative Literature Reviews». *Review of General Psychology* 1 (3): 311–20. https://doi.org/10.1037/1089-2680.1.3.311.

Belur, Jyoti, Lisa Tompson, Amy Thornton, and Miranda Simon. 2018. «Interrater Reliability in Systematic Review Methodology: Exploring Variation in Coder Decision-Making». *Sociological Methods & Research* 50 (2): 837–65. https://doi.org/10.1177/0049124118799372.

Bond, Melissa. 2020. «Schools and Emergency Remote Education During the COVID-19 Pandemic: A Living Rapid Systematic Review». *Asian Journal of Distance Education* 15 (2): 191–247. https://doi.org/10.5281/zenodo.4425683.

Bond, Melissa, Katja Buntins, Svenja Bedenlier, Olaf Zawacki-Richter, and Michael Kerres. 2020. «Mapping Research in Student Engagement and Educational Technology in Higher Education: A Systematic Evidence Map». *International Journal of Educational Technology in Higher Education* 17 (1): 1–30. https://doi.org/10.1186/s41239-019-0176-8.

Borenstein, Michael, ed. 2009. *Introduction to Meta-Analysis:* John Wiley & Sons, Ltd.

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2010. «A Basic Introduction to Fixed-Effect and Random-Effects Models for Meta-Analysis». *Research Synthesis Methods* 1 (2): 97–111. https://doi.org/10.1002/jrsm.12.

Borrego, Maura, Margaret J. Foster, and Jeffrey E. Froyd. 2014. «Systematic Literature Reviews in Engineering Education and Other Developing Interdisciplinary Fields». *Journal of Engineering Education.* 103 (1): 45–76. https://doi.org/10.1002/jee.20038.

Bramer, Wichor M., Dean Giustini, and Bianca Kramer. 2016. «Comparing the Coverage, Recall, and Precision of Searches for 120 Systematic Reviews in Embase, MEDLINE, and Google Scholar: A Prospective Study». *Systematic Reviews* 5 (1): 39. https://doi.org/10.1186/s13643-016-0215-7.

Bramer, Wichor M., Dean Giustini, Bianca Kramer, and P. F. Anderson. 2013. «The Comparative Recall of Google Scholar Versus PubMed in Identical Searches for Biomedical Systematic Reviews: A Review of Searches Used in Systematic Reviews». *Systematic Reviews* 2 (1): 115. https://doi.org/10.1186/2046-4053-2-115.

Braun, Virginia, and Victoria Clarke. 2006. «Using Thematic Analysis in Psychology». *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Bulfin, Scott, Michael Henderson, Nicola F. Johnson, and Neil Selwyn. 2014. «Methodological Capacity Within the Field of «Educational Technology» Research: An Initial Investigation». *British Journal of Educational Technology* 45 (3): 403–14. https://doi.org/10.1111/bjet.12145.

Buntins, Katja, Michael Kerres, and Anna Heinemann. 2022. «A Scoping Review of Research Instruments for Measuring Student Engagement: In Need for Convergence». *International Journal of Educational Research Open* 2. https://doi.org/10.1016/j.ijedro.2021.100099.

Buntins, Katja, Miriam Mulders, and Nadine Schröder. 2023. «Cognitive Biases in Screening Processes – Search Strategies in Educational Technology Research a Systematic Review on Learning with Virtual Reality». *MedienPädagogik* 54: 103–124. https://doi.org/10.21240/mpaed/54/2023.11.19.X.

Campbell, Allison, Brian Taylor, Jessica Bates, and Una O'Connor-Bones. 2018. «Developing and Applying a Protocol for a Systematic Review in the Social Sciences». *New Review of Academic Librarianship* 24 (1): 1–22. https://doi.org/10.1080/13614533.2017.1281827.

CASP. 2022. «CASP Checklists. Critical Appraisal Skills Programme». https://casp-uk.net/casp-tools-checklists/.

Cooper, Chris, Rebecca Lovell, Kerryn Husk, Andrew Booth, and Ruth Garside. 2018a. «Supplementary Search Methods Were More Effective and Offered Better Value Than Bibliographic Database Searching: A Case Study from Public Health and Environmental Enhancement». *Research Synthesis Methods* 9 (2): 195–223. https://doi.org/10.1002/jrsm.1286.

Cooper, Chris, Joanna Varley-Campbell, Andrew Booth, Nicky Britten, and Ruth Garside. 2018b. «Systematic Review Identifies Six Metrics and One Method for Assessing Literature Search Effectiveness but No Consensus on Appropriate Use». *Journal of Clinical Epidemiology* 99:53–63. https://doi.org/10.1016/j.jclinepi.2018.02.025.

Cooper, Harris, and L. V. Hedges, eds. 1994. *Handbook of Research Synthesis:* Russell Sage Foundation.

Crompton, Helen, Mildred V. Jones, and Diane Burke. 2022. «Affordances and Challenges of Artificial Intelligence in K-12 Education: A Systematic Review». *Journal of Research on Technology in Education*, 1–21. https://doi.org/10.1080/15391523.2022.2121344.

Durlak, J. A., and M. W. Lipsey. 1991. «A Practitioner's Guide to Meta-Analysis». *American Journal of Community Psychology* 19 (3): 291–332. https://doi.org/10.1007/BF00938026.

Eysenbach, G., J. Tuische, and T. L. Diepgen. 2001. «Evaluation of the Usefulness of Internet Searches to Identify Unpublished Clinical Trials for Systematic Reviews». *Medical Informatics and the Internet in Medicine* 26 (3): 203–18. https://doi.org/10.1080/14639230110075459.

Glass, Gene V., Barry McGaw, and Mary Lee Smith. 1981. «Meta-analysis in social research». SAGE Publications.

Glass, Gene V., and M. L. Smith. 1979. «Meta-Analysis of Research on Class Size and Achievement». *Educational Evaluation and Policy Analysis* 1 (1): 2–16. https://doi.org/10.3102/01623737001001002.

Gough, David, Phil Davies, Gro Jamtvedt, Etienne Langlois, Julia Littell, Tamara Lotfi, Edoardo Masset et al. 2020. «Evidence Synthesis International (ESI): Position Statement». *Systematic Reviews* 9 (1): 155. https://doi.org/10.1186/s13643-020-01415-5.

Hammersley, Martyn. 2001. «On 'Systematic' Reviews of Research Literatures: A 'Narrative' Response to Evans & Benefield». *British Educational Research Journal* 27 (5): 543–54. https://doi.org/10.1080/01411920120095726.

Hartling, Lisa, Andrea Milne, Michele P. Hamm, Ben Vandermeer, Mohammed Ansari, Alexander Tsertsvadze, and Donna M. Dryden. 2013. «Testing the Newcastle Ottawa Scale Showed Low Reliability Between Individual Reviewers». *Journal of Clinical Epidemiology* 66 (9): 982–93. https://doi.org/10.1016/j.jclinepi.2013.03.003.

Hew, Khe Foon, Min Lan, Ying Tang, Chengyuan Jia, and Chung Kwan Lo. 2019. «Where Is the 'Theory' Within the Field of Educational Technology Research?» *British Journal of Educational Technology* 50 (3): 956–71. https://doi.org/10.1111/bjet.12770.

Higgins, Julian P. T., Douglas G. Altman, Peter C. Gøtzsche, Peter Jüni, David Moher, Andrew D. Oxman, Jelena Savović, Kenneth F. Schulz, Laura Weeks, and Jonathan A. C. Sterne. 2011. «The Cochrane Collaboration Tool for Assessing Risk of Bias in Randomised Trials». *BMJ* 343. https://doi.org/10.1136/bmj.d5928.

Higgins, Julian P. T., J. Thomas, J. Chandler, M. Cumpston, T. Li, Matthew J. Page, and Vivian A. Welch. 2022. *Cochrane Handbook for Systematic Reviews of Interventions,* Version 6.3*:* Cochrane.

Ioannidis, John P. A., Marcus R. Munafò, Paolo Fusar-Poli, Brian A. Nosek, and Sean P. David. 2014. «Publication and Other Reporting Biases in Cognitive Sciences: Detection, Prevalence, and Prevention». *Trends in Cognitive Sciences* 18 (5): 235–41. https://doi.org/10.1016/j.tics.2014.02.010.

Katrak, Persis, Andrea E. Bialocerkowski, Nicola Massy-Westropp, V. S. Kumar, and Karen A. Grimmer. 2004. «A Systematic Review of the Content of Critical Appraisal Tools». *BMC Medical Research Methodology* 4 (1): 1–11. https://doi.org/10.1186/1471-2288-4-22.

Kimmons, Royce, and Joshua M. Rosenberg. 2022. «Trends and Topics in Educational Technology, 2022 Edition». *TechTrends* 66 (2): 134–40. https://doi.org/10.1007/s11528-022-00713-0.

Kolbe, Richard H., and Melissa S. Burnett. 1991. «Content-Analysis Research: An Examination of Applications with Directives for Improving Research Reliability and Objectivity». *Journal of Consumer Research* 18 (2): 243–50. https://doi.org/10.1086/209256.

Krippendorff, Klaus. 2018. *Content Analysis: An Introduction to Its Methodology:* SAGE Publications.

Lai, Jennifer W. M., and Matt Bower. 2020. «Evaluation of Technology Use in Education: Findings from a Critical Analysis of Systematic Literature Reviews». *Journal of Computer Assisted Learning* 36 (3): 241–59. https://doi.org/10.1111/jcal.12412.

Landis, J. Richard, and Gary G. Koch. 1977. «The Measurement of Observer Agreement for Categorical Data». *Biometrics* 33 (1): 159–74. https://doi.org/10.2307/2529310.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. «Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability». *Human Communication Research* 28 (4): 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x.

Marín, Victoria I., Katja Buntins, Svenja Bedenlier, and Melissa Bond. 2023. «Invisible Borders in Educational Technology Research? A Comparative Analysis». *Educational Technology Research and Development: ETR & D* 71: 1349–1370. https://doi.org/10.1007/s11423-023-10195-3.

Meline, Timothy. 2006. «Selecting Studies for Systemic Review: Inclusion and Exclusion Criteria». *Contemporary Issues in Communication Science and Disorders* 33 (Spring): 21–27. https://doi.org/10.1044/cicsd_33_S_21.

Methley, Abigail M., Stephen Campbell, Carolyn Chew-Graham, Rosalind McNally, and Sudeh Cheraghi-Sohi. 2014. «PICO, PICOS and SPIDER: A Comparison Study of Specificity and Sensitivity in Three Search Tools for Qualitative Systematic Reviews». *BMC Health Services Research* 14:579. https://doi.org/10.1186/s12913-014-0579-0.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. «Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement». *PLoS Medicine* 6 (7): e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer et al. 2021. «The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews». *BMJ* 372. https://doi.org/10.1136/bmj.n71.

Pigott, Therese D., and Joshua R. Polanin. 2020. «Methodological Guidance Paper: High-Quality Meta-Analysis in a Systematic Review». *Review of Educational Research* 90 (1): 24–46. https://doi.org/10.3102/0034654319877153.

Polanin, Joshua R., Brandy R. Maynard, and Nathaniel A. Dell. 2017. «Overviews in Education Research: A Systematic Review and Analysis». *Review of Educational Research* 87 (1): 172–203. https://doi.org/10.3102/0034654316631117.

Pollock, Alex, and Eivind Berge. 2018. «How to Do a Systematic Review». *International Journal of Stroke* 13 (2): 138–56. https://doi.org/10.1177/1747493017743796.

Shea, Beverley J., Jeremy M. Grimshaw, George A. Wells, Maarten Boers, Neil Andersson, Candyce Hamel, Ashley C. Porter, Peter Tugwell, David Moher, and Lex M. Bouter. 2007. «Development of AMSTAR: A Measurement Tool to Assess the Methodological Quality of Systematic Reviews». *BMC Medical Research Methodology* 7 (1): 10. https://doi.org/10.1186/1471-2288-7-10.

Singh, Jatinder. 2013. «Critical Appraisal Skills Programme». *Journal of pharmacology and Pharmacotherapeutics* 4 (1): 76. https://doi.org/10.4103/0976-500X.107697.

Sotola, Lukas K. 2022. «Garbage In, Garbage Out? Evaluating the Evidentiary Value of Published Meta-Analyses Using Z-Curve Analysis». *Collabra: Psychology* 8 (1): 32571. https://doi.org/10.1525/collabra.32571.

Stock, William A. 2008. «Systematic Coding for Research Synthesis». *The Handbook of Research Synthesis* 236: 125–38. Russell Sage Foundation.

Tamim, Rana M., Robert M. Bernard, Eugene Borokhovski, Philip C. Abrami, and Richard F. Schmid. 2011. «What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study». *Review of Educational Research* 81 (1): 4–28. https://doi.org/10.3102/0034654310393361.

Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O›Driscoll, P., Bond, M., and Koryakina, A. 2023. «EPPI Reviewer: advanced software for systematic reviews, maps and evidence synthesis». EPPI Centre Software. UCL Social Research Institute. London. https://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4.

Viera, Anthony J., and Joanne M. Garrett. 2005. «Understanding Interobserver Agreement: The Kappa Statistic». *Family Medicine* 37 (5): 360–63. https://pubmed.ncbi.nlm.nih.gov/15883903/.

Zawacki-Richter, Olaf, Michael Kerres, Svenja Bedenlier, Melissa Bond, and Katja Buntins, eds. 2022. *Systematic Reviews in Educational Research.* Springer Fachmedien: Wiesbaden. https://doi.org/10.1007/978-3-658-27602-7.

Acknowledgments

Victoria I. Marín acknowledges the support of the Grant RYC2019-028398-I funded by MCIN/AEI/ https://doi.org/10.13039/501100011033 and FSE “El FSE invierte en tu futuro”.