Educational Review

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cedr20

Metasummary: examining the potential of a methodologically inclusive approach for conducting systematic reviews of educational research

Jason Anderson

To cite this article: Jason Anderson (20 Sep 2024): Metasummary: examining the potential of a methodologically inclusive approach for conducting systematic reviews of educational research, Educational Review, DOI: 10.1080/00131911.2024.2401079

To link to this article: https://doi.org/10.1080/00131911.2024.2401079

EDUCATIONAL REVIEW
https://doi.org/10.1080/00131911.2024.2401079

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Metasummary: examining the potential of a methodologically inclusive approach for conducting systematic reviews of educational research

Jason Anderson

Department of Applied Linguistics, School of Education, Learning and Communication Sciences, University of Warwick, Coventry, UK

ABSTRACT

This article critically examines metasummary as a methodology for systematic review that has, to date, been underused in the field of educational research. Because of its ability to combine and report on qualitative, quantitative and mixed methods studies in an integrated fashion, metasummary is potentially able to bridge the often-perceived paradigm divide between these different schools of data collection and analysis. The article begins by identifying key features of systematic review and identifying two broad schools of systematic review procedure in the social science literature. It then discusses more recent developments in mixed methods systematic reviews, focusing in on metasummary as the approach of interest and its potential for use in education. It reviews critically the limited number of metasummaries conducted in educational research to date relative to their adherence to recommended metasummary procedure. The final section of the article offers epistemological justification for more widespread use of metasummary before discussing the potential benefits and shortcomings of the approach. It concludes with methodological guidelines for researchers interested in using the approach in education.

ARTICLE HISTORY
Received 10 April 2024
Accepted 1 September 2024

KEYWORDS
Metasummary; systematic review; research methodology; guidelines

Introduction

Systematic review, as an approach to research synthesis, has expanded rapidly in recent years (Cohen et al., 2011). Since Glass (1976) first described the procedure for meta-analysis, a wide range of approaches have appeared, including a suite of qualitative approaches often grouped under the umbrella term "meta-synthesis" (Thorne et al., 2004), and, more recently, those proposing means for conducting mixed methods research synthesis (see, e.g. Heyvaert et al., 2017; Sandelowski et al., 2012). This article investigates one of these, called metasummary, which is widely used in healthcare research (Pollock et al., 2016; Sandelowski et al., 2007; Stern et al., 2021) yet is little known in education. The article

CONTACT Jason Anderson jason.anderson@warwick.ac.uk Department of Applied Linguistics, School of Education, Learning and Communication Sciences, University of Warwick, Coventry, CV4 7AL, UK

argues that, if used with critical awareness of its limitations, metasummary has the potential to be useful as a means to bridge the often-perceived paradigm divide (e.g. Andrews, 2005) between qualitative and quantitative methodologies in systematic reviews.

The article begins by identifying defining features of systematic reviews as discussed in the literature. Following Gough et al. (2012), it identifies two broad approach schools within the wider research synthesis literature: *aggregative* and *configurative* systematic reviews. It then introduces more recently developed mixed method designs for research synthesis, focusing in on metasummary, an approach introduced by Margarete Sandelowski and colleagues as a means to integrate findings systematically from qualitative, quantitative and mixed methods studies (e.g. Sandelowski et al., 2007). The article then examines the potential transferability of the approach to educational research and reviews the small number of metasummaries conducted in education to date. It offers two epistemological justifications for metasummary from the perspectives of mixed methods pragmatism and critical realism before highlighting some of the potential benefits and weaknesses associated with the approach. The article concludes by proposing initial guidelines for conducting metasummary in education.

A brief overview of systematic review methodologies

Understandings of the term "systematic review" as a form of research synthesis vary somewhat within different bodies of literature in social and interdisciplinary science, although areas of common ground can nonetheless be identified. Discussions most often contrast it with "narrative review" (e.g. Bearman et al., 2012; Cohen et al., 2011; Davies, 2000; Evans & Benefield, 2001; Waterfield, 2018), which is typically seen to involve a less systematic literature review of the type included in most journal articles, monographs and research student theses. In contrast to these traditional approaches to reviewing literature, the features that make research syntheses "systematic" tend to include several, most, or all of the following five features in definitions of the approach:

(1) The methodology used is transparent/explicit (Andrews, 2005; Cohen et al., 2011; Denner et al., 2017; EPPI Centre, 2019; Evans & Benefield, 2001; Gough et al., 2012; Maeda et al., 2022; Page et al., 2021; Thorne et al., 2004);
(2) Objectives and/or research questions are explicit and clearly formulated (Andrews, 2005; Bearman et al., 2012; Chapman, 2021; Evans & Benefield, 2001; Page et al., 2021; Thorne et al., 2004; Waterfield, 2018);
(3) The literature search is exhaustive and inclusion criteria are made explicit (Evans & Benefield, 2001; Gough et al., 2012; Thorne et al., 2004);
(4) There is an attempt to minimise bias or misinterpretation through use of the recommended procedure or protocol (Cohen et al., 2011; Cook et al., 1995; Evans & Benefield, 2001; Page et al., 2021);
(5) The procedure is reproducible or replicable (Bearman et al., 2012; EPPI Centre, 2019; Maeda et al., 2022; Waterfield, 2018).

In their discussion of different systematic review designs, Gough et al. (2012) draw upon earlier theorisation by Voils et al. (2008) to identify two broad approach schools,

yet also suggest that there are continua between them with regard to how different aspects of the review process are applied systematically. These schools are the *aggregative* and *configurative* systematic review types.

Aggregative reviews are typically located within realist or post-positivist philosophical camps, draw most often upon quantitative data and seek to make use of *a priori* declared objectives and methods to "test" aspects of theory by examining the weight of evidence available, while avoiding bias through specific practices and protocols (Gough et al., 2012). They are intended for broadly "instrumental" uses (e.g. to inform educational policy and practice). While this may include a number of systematic review approaches (e.g. vote-counting reviews, realist synthesis and reviews employing bibliometric analysis), by far the best known and most widely used of these is meta-analysis. There is general consensus in the literature on meta-analysis that it involves the statistical synthesis of results from independent but comparable studies that facilitates a quantitative summary of the pooled results, through, for example, an "effect size" statistic (e.g. Higgins, 2018; O'Rourke, 2007; Porta, 2014). Since Glass (1976) coined the term and basic procedure, tens of thousands of meta-analyses have been conducted, particularly in healthcare (see Davey et al., 2011), but also education (see Hattie, 2009). Web of Science currently lists 105,000 review articles with the term "meta-analysis" in the title, indicating that it is an established and very popular methodology. Numerous guides for conducting meta-analysis exist (e.g. Field & Gillett, 2010; Plonsky & Oswald, 2012), and while there is disagreement on finer points of analysis, procedures are generally standardised and the validity of the methodology is widely recognised.

Configurative reviews typically originate in interpretivist, constructivist or relativist philosophies and are more likely to draw upon qualitative data sources (Voils et al., 2008). They often make use of iterative methods and typically "generate" theory through transparent means "to provide enlightenment through new ways of understanding" (Gough et al., 2012, p. 3), rather than simply aggregating findings. They have evolved rapidly since the 1990s, when the term *meta-synthesis* (e.g. Jensen & Allen, 1996; Sandelowski et al., 1997) became established as an umbrella term to refer to many such systematic, qualitative reviews (Walsh & Downe, 2005). As Thorne et al. (2004) caution, the approaches that make use of the term meta-synthesis should be seen as a "family" rather than a single method, and include meta-ethnography, thematic synthesis, grounded formal theory, metastudy, critical interpretive synthesis and meta-aggregation (see Paterson, 2013), each with specific processes that aim to go beyond simply summarising findings to "deepen the understanding of a studied topic" (Maeda et al., 2022, p. 1). They often seek to "[move] the synthesis from the level of data to the level of interpretation" (Thorne et al., 2004, p. 1347), "reveal[ing] patterns or relationships between concepts and structures that remained hidden before" (Heyvaert et al., 2017, p. 9). Like meta-analysis, meta-synthesis is now well established, with nearly 2000 review articles listed on Web of Science including "meta(-)synthesis" in the title.

These two schools have been historically separated by the so-called paradigm divide in the social sciences (Andrews, 2005). However, as mixed methods research has evolved in its own right, supported by somewhat independent paradigmatic positions (e.g. pragmatist and critical realist; discussed below), researchers seeking to bring together findings from studies with diverse methodologies have more recently experimented with mixed methods synthesis approaches.

Mixed methods approaches to synthesising qualitative and quantitative data

Researchers have more recently begun investigating means for combining the findings of qualitative and quantitative studies, often through the use of methodologies originating in mixed methods research but adapted for use in secondary syntheses (e.g. Harden & Thomas, 2005; Heyvaert et al., 2013, 2017; Pluye et al., 2009; Sandelowski et al., 2006; Voils et al., 2008); these are typically referred to as mixed methods research synthesis (MMRS), which Heyvaert et al. (2013, p. 662) define as "a systematic review applying the principles of mixed methods research". They argue it should have "two well-developed distinct strands, one qualitative and one quantitative … [and] it must also integrate, link, or connect these strands in some way" (p. 662).

Both Heyvaert et al. (2017) and Stern et al. (2021) recommend careful selection of research design depending on aims and the nature of the data available, similar to primary mixed methods research. This means that a wide variety of methodological frameworks are seen as potentially valid. Following Sandelowski et al. (2006), Heyvaert et al. (2017) make a distinction between *segregated, integrated* and *contingent* literature review syntheses. Segregated designs treat qualitative and quantitative data as distinct sources of evidence to be synthesised only after reporting findings from the two datasets separately (e.g. in the Discussion section of a research report). Integrated designs involve combining qualitative and quantitative sources during analysis, for example by quantitizing qualitative data or qualitizing quantitative data; this includes metasummary, discussed below. Contingent designs involve sequential, adaptive phases of review to achieve the research objectives (e.g. a quantitative review to answer one research question, leading to a second review of qualitative data to address outstanding objectives). While they use slightly different terms, Stern et al. (2021) also recognise the key distinction between integration ("convergent integrated") and segregation ("convergent segregated") of findings in MMRS approaches.

Metasummary

As the first widely disseminated approach to mixed methods synthesis, "metasummary" (also "meta-summary") was developed in the late 1990s in the field of healthcare research by Margarete Sandelowski and colleagues, initially for summarising the findings of qualitative research. The approach was later refined and discussed in greater detail, often using the term (qualitative) "metasummary" (Sandelowski et al., 2006, 2007; Sandelowski & Barroso, 2003, 2007; Voils et al., 2008). While their early theorisation focused on it as a means for systematic review of qualitative studies, subsequent publications (e.g. Sandelowski et al., 2007; Voils et al., 2008) also discuss its potential for including quantitative research findings in reviews (also see Stern et al., 2021). According to Sandelowski and colleagues (Sandelowski et al., 2006, 2007; Sandelowski & Barroso, 2007), metasummary is a quantitatively-oriented means to extract and aggregate the findings of diverse qualitative and quantitative studies in a way that is both transparent and replicable and, if desired, also enables researchers to report both "frequency" and "intensity" effect sizes for the findings (Onwuegbuzie, 2003). It is the most widely used example of integrated MMRS (see above), with the findings of primary studies being integrated during data analysis.

Metasummary procedure

As often recommended for other types of systematic review (e.g. Andrews, 2005; Heyvaert et al., 2017), Sandelowski and Barroso (2007) recommend metasummary is carried out by a team of researchers (rather than an individual), including members with both methodological and topical expertise. Also similar to other systematic review approaches, the first stage of metasummary involves an exhaustive literature search across databases for studies of potential relevance, including both "high precision searches" using key terms and "high-recall searches" using a wider range of means (Sandelowski & Barroso, 2007, p. 35). Inclusion criteria should be pre-specified, but as inclusive as possible to avoid excluding potentially relevant studies. In contrast to some advice for systematic reviews (e.g. Greco et al., 2013), Sandelowski et al. (2007) recommend *not* filtering studies at this stage for indicators of quality (e.g. whether studies have been peer reviewed). Instead, all studies meeting basic inclusion criteria should be assumed into the dataset for the metasummary.

The second stage involves the identification and extraction of findings or results (hereafter "findings") of relevance to the aims of the metasummary. These are extracted directly from the research reports themselves and logged systematically by members of the research team through a manual process of reading and coding all findings presented in the reports regardless of whether they are qualitative, quantitative or mixed methods. Sandelowski and Barroso (2007, p. 152) recommend that a working definition of "target findings" is used to identify material to be extracted; this can be adapted if required. They stress the importance of ensuring that only findings that are presented as original to the empirical study in question are extracted, as opposed to the discussion of findings from previous studies (e.g. in literature review and discussion sections of reports). They also recommend that all research reports are given equal weighting regardless of sample size (Sandelowski & Barroso, 2007), and if two publications report findings from the same study, these findings should be counted only once. In case of disagreements concerning extraction or codification of findings, Sandelowski and Barroso (2007, p. 230) favour "negotiated consensus", in which team members try to reach agreement through reasoned discussion rather than inter-rater reliability tests.

Sandelowski and Barroso (2007) then recommend that findings are grouped together and abstracted. The first of these two stages requires researchers to group together all findings on the same topic regardless of whether they "say the *same thing* about the same topic" (p. 158). Each grouped set of findings is then reviewed critically to enable abstraction involving "more parsimonious renderings" of the findings that nonetheless "preserve [any] contradictions and ambiguities" present (p. 159) (i.e. complexity is retained as much as is possible within a research synthesis).

These abstracted findings are then presented in the research report according to the research objectives and/or questions, employing two elements that are the central presentation features of metasummary:

(1) a *summary table* in which the findings are presented according to frequency (i.e. how often they are found within the research report dataset), with the most frequently reported findings at the top of the table;

(2) a *descriptive summary* of these same findings, which allows authors to "preserve the complexity of their content" (Sandelowski et al., 2007, p. 6) by describing the findings qualitatively, including discussion of complexity, variation and examples in ways that are similar to a narrative review.

Finally, Sandelowski and Barosso (2007) recommend calculating and reporting two effect sizes as a means "to unite the empirical precision of quantitative research with the descriptive precision of qualitative research" (p. 160). A frequency effect size is calculated for each finding by dividing the number of reports in the dataset that contain that finding by the total number of reports in the dataset (e.g. if a finding is identified in six of a total of ten reports, the frequency effect size is 60%). An intensity effect size is calculated for each report by dividing the number of findings it contributes to the dataset by the total number of findings in the dataset to establish the apparent utility, reliability and representability of the report in question (Sandelowski et al., 2007; Sandelowski & Barroso, 2007).

As presented here, the procedure for metasummary includes all five elements common in systematic reviews, as discussed above. It involves a transparent methodology, explicit objectives, an exhaustive literature search with pre-defined inclusion criteria, a clear protocol that is replicable and helps to reduce bias (e.g. through inappropriate pre-filtering of studies into the dataset) and misinterpretation (e.g. through the use of negotiated consensus). Importantly, it is also able to integrate findings from qualitative, quantitative and mixed methods studies, making it of potentially great interest to researchers working in education.

Metasummary in educational research

As two of the largest fields of research involving social science, healthcare and education share a number of similarities to support the claim that those research methodologies that work in one are also likely to be similarly useful in the other (e.g. Evans & Benefield, 2001). Both involve a closely integrated combination of social, psychological and physiological phenomena, indicating the need for a range of data collection approaches and methods applied critically to enable researchers to capture the full complexity of any topic of investigation (Stern et al., 2021). Correspondingly, both qualitative and quantitative data are common and valued in both fields. Both also involve public services involving huge numbers of employees, a large range of professional fields and often extensive centralised (especially government) funding. As such, there are shared issues of accountability, monitoring and evaluation that are likely to require not only primary research but also secondary research summaries that are accessible to key stakeholders (e.g. policy-makers, doctors, school leaders, teachers, etc.) (Oakley, 2000). Since the move towards "evidence-based" practice around the turn of the twenty-first century (Evans & Benefield, 2001), aggregative and configurative systematic reviews (discussed above) have become common in both fields. As such it is not surprising that recent years have also seen an expansion in the use of metasummary in research in education; integrative MMRS of qualitative and quantitative research has an additional benefit over other systematic review types in its potential to capture the full picture of a phenomenon or topic in one set of findings.

However, there are also important differences between the two fields that should be taken into account when attempting to transfer a specific research approach from one to the other (see Biesta, 2007; Hammersley, 2001, 2004). These include the fact that healthcare is an interdisciplinary field, albeit with a strong social science element, whereas education sits primarily within social science. As a result, the latter involves more obviously cultural, symbolically mediated practices and traditions, in which values play a much greater role in determining appropriate practices and outcomes (Biesta, 2007). As such, it is likely that there is greater cultural variability between systems, such that research conducted in one context may not be relevant, appropriate or useful in another (Hammersley, 2001). This suggests that any systematic review needs to include discussion of such differences, including the role of how different factors/variables influence outcomes of interest, which themselves may be debated (see Biesta, 2015). Further, because of the complexity of the interactions involved in education (learners and teachers build up long-term, meaningful relationships), investigations of issues of effectiveness, quality or impact are likely to involve a wide range of cognitive, practical, professional and even personal factors, which may make the extraction of findings from research reports highly complex, as Anderson and Taner (2023) found in their meta-summary of teacher expertise research.

A potential example of a procedural element in Sandelowski and Barroso's (2007) guidelines that may be inappropriate in education is their recommendation to calculate what they call "frequency" and "intensity effect sizes" in the final stages of the metasummary procedure (see above). Both of these differ from normal understandings of effect size in educational research, where it typically denotes measurement of the likely impact of a specific practice or intervention on appropriate outcome measures (e.g. student learning or engagement), not simply how frequently a finding is recorded (Paterson, 2013). The latter (intensity effect size), as a measure of utility of an individual study, may not be reliable within the diverse subfields of educational research; in some of these fields, a higher intensity effect size may simply reflect a wider scope, a more exploratory research question, or more detailed reporting of findings (e.g. through "thick description").

Bearing this need for critical appraisal in mind, the remainder of this article reviews prior metasummaries conducted in education, then discusses the theoretical arguments for its use, as well as its potential benefits and shortcomings. Based on this critical appraisal, it suggests tentative guidelines for future use.

Prior metasummaries in education

In order to develop an initial understanding of how frequently and how well metasummary was being used in educational research, the author conducted a literature search of prior metasummaries in education. Three databases were consulted (ERIC, Google Scholar, Web of Science) using appropriate search terms (see Table 1). The titles and abstracts of studies returned in search results were analysed for the following inclusion criteria:

(1) The primary language of the study is English;
(2) The primary domain of the study in question is educational research;

Table 1. Literature searches conducted.

Database	Search syntax used	Date of search	No. of search results returned
Web of Science	AB = (metasummary OR meta-summary) OR TI = (metasummary OR meta-summary) OR TS = (metasummary OR meta-summary) AND ALL = (education OR studying OR teaching OR learning OR school* OR university OR student OR learner OR study OR classroom OR pedagogy)	24 July 2024	182
ERIC	title: "metasummary" OR title: "meta-summary" OR abstract: "metasummary" OR abstract: "meta-summary"	24 July 2024	26
Google Scholar	(metasummary OR meta-summary) AND (education OR studying OR teaching OR learning OR school* OR university OR student OR learner OR study OR classroom OR pedagogy)	24 July 2024	c. 5250[a]

[a]Only the first 1000 studies returned by Google Scholar were assessed for inclusion criteria.

(3) The study describes metasummary as the approach used (i.e. referencing the Sandelowski procedure, with justification for any differences adopted);
(4) There is sufficient detail in the reporting of the methodology to demonstrate that the authors attempted to identify and summarise the findings of all research reports included in the dataset;
(5) Findings are reported using both summary table(s) and descriptive summary.

The findings of this literature search confirmed that metasummary is not widely used in the field, returning only nine reviews that met all five inclusion criteria. A further eleven reviews in the field of education that used the term "metasummary/meta-summary" were considered and rejected (Ajlani et al., 2024; Dominguez & Hager, 2013; Habib & Pius, 2023; Nkoana & Ebersöhn, 2023; Novita et al., 2021; Pileggi, 2024; Saito et al., 2022; Shorey et al., 2023; Tibingana-Ahimbisibwe et al., 2022; Vieira et al., 2018; Vieira & Finardi, 2018). The two reasons for rejection were either that the study involved a non-systematic (e.g. narrative) review (i.e. not following Sandelowski and colleagues' recommended procedure) or that they failed to include either one or both of the key reporting vehicles of metasummary in the findings (e.g. the summary table or descriptive summary). The nine studies that met inclusion criteria are summarised in Table 2.

A number of observations can be made concerning the variety and quality of the studies described in Table 2. Firstly, the use of metasummary in education, while still rare, is a recent and apparently increasing trend, with the first study dating from 2013, and six of the nine published since 2020. Secondly, these nine studies investigate a wide range of topics, including areas of focus that are both general (e.g. the impact and contributions of STEM education) and more specific (e.g. challenges with curriculum reform in Indonesia). They include both peer-reviewed published articles and PhD studies. Despite the small number of studies, this evidence of a wide range of topics and purposes indicates the potential versatility of metasummary as an approach for reviewing educational research. Thirdly, the number of research reports in the different studies' datasets varies considerably from 7 to 106, yet all were able to present key findings using the methodology. Two studies that included fewer research reports (Demir, 2018; Kanadlı, 2019) were nonetheless able to report on their datasets convincingly, likely due, in part, to the comparatively large number of total participants (>600 in both cases) across the reports in question. Fourthly, significant variation in quality is apparent among the studies, including variation in transparency when reporting procedure, variation in

Table 2. Metasummaries in education.

Study	No. of coders	No. of RRs in dataset	QL, QN or MM?	Peer-reviewed?	Followed MS recs?	Topic	Notes, incl. good practice (+) and critique (–)
Al-Salman (2013)	1	49	QL, QN, MM	N (PhD study)	****	The role of faculty staff in making online tertiary programs effective	(+) Careful use of MS: Extensive literature review. Summary tables used (3 tables, for different RQs with % and "effect size"). Detailed and informative descriptive summaries with appropriate citation. (–) Only 1 coder. No transparent discussion on how themes were extracted.
Demir (2018)	1	8	QL	Y	**	Metaphors used by teachers, trainee teachers and students to describe EFL teacher roles in Turkey	No QN due to topic. (+) Transparent method. Extensive literature search. Large number of total participants ($n = 626$). Used thematic analysis to extract themes inductively. Clear definitions for key themes are offered transparently. Used summary table (incl. %) and diagram. (–) 8 RRs only. Initial quality filtering not consistent with MS (e.g. rejected non-peer reviewed studies). Only 1 coder. Descriptive summary of findings is brief, restricted to Discussion only.
Kanadlı (2019)	1 or 2	22	QL	Y	***	Impact and contributions of STEM education	390 themes in total. (+) Good transparency on procedure. Large n. of total participants ($n = 699$). Second coder involved. Summary table included (incl. % as FES). (–) Initial quality filtering not consistent with MS. QN studies rejected from analysis. Only informal IRR was conducted. Descriptive summary is weak, focused mainly on unpacking the statistics. FES calculations unclear, but may be for subsets of data rather than complete dataset.
Taskaya and Gül (2020)	2	50	N/A	Y	**	Teaching [pedagogic] principles in teaching guides	Dataset is published books for teachers, rather than research reports. (+) Detailed summary table. FES and "density" [i.e. intensity] effect size both calculated. (–) English used in article is at times ambiguous. Poor reporting of procedure (literature search, coding). Descriptive summary of findings is very brief with little expansion on terms used.
Eberhardt (2021)	1	7	QL	N (PhD study)	**	Perspectives of members of school personnel on LGBTQ bullying in schools	Unusually brief methodology for a PhD study. Combines MS with "conventional qualitative analysis' (p. 25). (+) Inclusion criteria clear. Extensive discussion of specific findings. (–) 7 RRs only. Poor reporting of procedure (e.g. literature search, data analysis). Only QL studies accepted. Only 1 coder.
Kamil (2021)	1	81	?	Y	**	Problems encountered during implementation of Indonesian national curriculum reform of 2013	(+) Uses summary table (incl. %) to report findings. (–) Literature search not very extensive. Only 1 coder. V. little detail/

(*Continued*)

Table 2. Continued.

Study	No. of coders	No. of RRs in dataset	QL, QN or MM?	Peer-reviewed?	Followed MS recs?	Topic	Notes, incl. good practice (+) and critique (−)
							transparency on issues of extraction/coding. Studies included are not referenced, little information on type.
Anderson and Taner (2023)	2	106	QL, QN, MM	Y	****	Teacher expertise in K12 education	Very broad topic. (+) Transparency on procedure. Extensive literature review with flow chart. Large number of total participants ($n = 1124$). 2 coders. Detailed discussion of coding procedure. Code saturation achieved. Used independent agreement criterion, IRR and threshold criterion to reduce coding bias/error. Summary table and descriptive summary with example studies both presented. (−) Large dataset means that potential findings may have been overlooked. Fairly low IRR result (moderate agreement).
Şanlı and Karakuş (2023)	?	14	QL, MM		**	Views of students and teachers on gender equality education in Turkey	(+) Cited and followed recommended procedure fairly well. Five summary tables are informative. Sufficient descriptive summary. (−) Replicates aspects of erroneous procedure from Kanadlı (2019), incl. exclusion of QN studies and initial quality filtering that is not consistent with MS. No detail on how coders worked together.
Haidiri and Koçoğlu (2024)	2	31	QL, MM	Y	***	Good teacher characteristics in Turkey: perceptions of students, teachers and school administrators	Broad topic. (+) Cited and followed recommended procedure generally well. Literature review extensive and well documented. 2 coders worked independently. Procedure reported in sufficient detail. 3 summary tables presented. (−) Initial quality filtering not consistent with MS. QN studies not considered for inclusion. Descriptive summary could be more extensive.

Note: Literature search conducted in July 2024. RR: research reports; QL: qualitative; QN: quantitative; MM: mixed methods; MS recs: metasummary recommendations of Sandelowski and colleagues, scored from 1 asterisk (low compliance) to 5 asterisks (high compliance); RQs: research questions; FES: frequency effect size; IRR: inter-rater reliability.

systematicity of literature search, variation in procedure for quality evaluation of research reports, transparency and rigour with regard to coding procedures (e.g. use of multiple raters) and variation concerning reporting procedures (e.g. several include only minimal descriptive summaries of the findings). Finally, it is evident that some studies, particularly those conducted recently at Turkish institutions (e.g. Haidiri & Koçoğlu, 2024; Şanlı & Karakuş, 2023; Taskaya & Gül, 2020), may be replicating research procedures adopted in Kanadlı's well-cited study (2019), but not necessarily those recommended in Sandelowski and colleagues' original procedure, discussed above. This includes the choice to pre-filter studies for quality before analysis (*contra* Sandelowski et al., 2007), the choice to exclude quantitative studies from the dataset (*contra* Sandelowski et al., 2007) and a primary focus on reporting findings through the summary table, sometimes to the detriment of the descriptive summary (*contra* Sandelowski et al., 2007).

The variation in quality and procedure identified through this review indicates strongly that guidelines for conducting metasummary in education may be useful to future researchers interested in using the approach. Such guidelines are presented later in this article.

Theoretical arguments for using metasummary in education

The evidence presented above indicates that metasummary is feasible and useful as a research approach in education. However, critical readers may justifiably ask why another approach to conducting systematic reviews is required. In this section of the paper two overlapping justifications are offered, from pragmatist and critical realist perspectives.

The pragmatist argument for metasummary

While both meta-analysis and meta-synthesis, as the two most common approaches to systematic review, are valuable tools in themselves, it can be argued that each has shortcomings in areas where the other has strengths (Stern et al., 2021), particularly when researching aspects of the complex, dynamic systems (e.g. learners, teachers, classrooms, schools, curricula, policies, etc.) that make up the practice of education (Bruner, 1996). The advantages of each largely replicates the advantages of quantitative and qualitative methodologies respectively, and thus the advantages of combining them also correspond to the pragmatic advantages of conducting mixed-methods research (Stern et al., 2021), which is argued to offer a "breadth and depth of understanding and corroboration" (Johnson et al., 2007, p. 123; also Teddlie & Tashakkori, 2011). Meta-analysis, as a quantitative, nomothetic approach, is able to "condense" the findings of large numbers of studies into simple, comparable metrics that are potentially generalisable across larger populations. This allows "best guess" theoretical comparison between, for example, certain methodologies, policies or classroom practices (e.g. Hattie, 2009), typically through the use of effect sizes. However, because meta-analysis draws upon multiple studies, such summary metrics are always oversimplifications (Thorne et al., 2004). And while many meta-analyses today include empirical moderator analyses to allow for comparison across contexts, these are limited to variables available in the source data, which may not account for all relevant differences (see, e.g. Voils et al., 2008). Further, if the

influence of moderators is not carefully discussed, readers may assume that a similar effect size can be expected in any educational context.

In contrast to meta-analysis, meta-synthesis, as a school of qualitatively-oriented, idiographic approaches, is able to retain more of the "thick description" of specific procedures and practices that make qualitative research valuable (e.g. Tracy, 2010). It is able to offer important insights from inside what quantitatively-oriented researchers often call the "black box" of the classroom (see Alexander, 2015), potentially enabling us to understand where, how and even why there seems to be variation – the nuanced "contingent causality" that emerges from small-n studies as a useful, theory-generating alternative to the explanatory causality of large-n quantitative research (Anderson, 2023). Nonetheless, many types of meta-synthesis neglect to also paint the wider picture that can be useful for readers who would like to understand exactly how often, how much, or how widely a given finding seems to apply, and to what effect.

The inclusive nature of metasummary (insomuch as it treats all studies as equally valid and all findings as potentially insightful) reduces the danger of overlooking the importance either of context or human experience when investigating a phenomenon of interest (Sandelowski et al., 2006). Thanks to its integrated design, audiences are able to see both which findings seem to have the greatest *generalisability* or *applicability* (through the summary table) and how different contexts or participant experiences may influence the potential *transferability* of findings (through the descriptive summary), thereby reducing the likelihood of overgeneralisation, a key danger in education research (Hammersley, 2004). As such, it enables us to "zoom in" and "zoom out" on the metaphorical map of research findings – a pragmatic solution to the challenges posed by systematic review. Perhaps more importantly, given that most phenomena of interest to educational researchers have been studied both qualitatively and quantitatively, any systematic review that attempts to investigate a phenomenon without examining research adopting both approaches risks misunderstanding, misinterpreting, under- or overgeneralising findings in ways that are potentially deleterious (Stern et al., 2021). This point is generally acknowledge in the field of healthcare research, where a number of scholars have argued for the pragmatic importance of inclusivity when conducting systematic reviews (see Stern et al., 2021; Thorne et al., 2004). Similar arguments can – and arguably should – be made for educational research.

The critical realist argument for metasummary

Because metasummary is, essentially, a mixed methods approach to systematic review, the above discussed pragmatic argument may be seen to suffice. However, more recently, a number of researchers (e.g. Maxwell & Mittapalli, 2010; Mukumbang, 2023; Zachariadis et al., 2013) have argued that mixed-methods methodologies may also be supported by a critical realist (in its broadest sense) position. Zachariadis et al. (2013) argue that critical realism occupies a middle position between the paradigms of positivism and interpretivism, also evident in Maxwell's (2012, p. 5) definition of critical realism, which

> retain[s] an ontological realism (there is a real world that exists independently of our perceptions, theories, and constructions) while accepting a form of epistemological constructivism and relativism (our understanding of this world is inevitably a construction from our own perspectives and standpoint).

In order to understand as much as possible about this independent reality, critical realist researchers may choose to draw flexibly and pragmatically upon a range of data collection tools available, including in conjunction with one another, to triangulate their findings on the phenomenon in question (Vincent & O'Mahoney, 2019). As Maxwell and Mittapalli (2010, p. 147) note, "realism provides a philosophical stance that is compatible with the essential methodological characteristics of both qualitative and quantitative research, and it can facilitate communication and cooperation between the two". Metasummary facilitates such communication effectively, particularly thanks to its ability to draw upon both qualitative and quantitative data sources, but also in its approach to data analysis; the simultaneous quantification and detailed description of findings. Indeed, Sandelowski and colleagues have adopted a critical realist position in their discussion of validity in metasummary (e.g. Sandelowski & Barroso, 2007), cognisant of "a 'real world that exists independently of' researchers conceptions of it", while also "recogniz [ing] that all knowledge is inescapably partial and socially constructed" (Sandelowski et al., 2012, p. 321). Similarly, in his discussion of validity in critical realism, Maxwell anticipates the approach adopted in metasummary when he notes that "descriptive validity can also pertain to numerically descriptive aspects of accounts … "quasi-statistics' – simple counts of things to support claims that are implicitly quantitative" (2012, p. 137).

Critical discussion of potential benefits and shortcomings of metasummary

The increasing number of metasummaries being conducted in education in recent years indicates that the approach is here to stay as part of the ever-increasing interest in systematic reviews in social science research. As such, like all methodologies, it has both benefits and shortcomings that researchers and readers should be aware of when choosing to conduct or make use of metasummaries in education. These are discussed critically in this section of the paper.

Probably the primary advantage of metasummary, when compared with the problem-specific approach of mixed methods research synthesis (Heyvaert et al., 2017), is that it serves as a means for the synthesis of data during the early stages of analysis (data coding) in a relatively straightforward process that is transparent, rigorous and replicable. This results in clearly displayed and interpretable findings, particularly through the summary table, which is likely to be useful to key non-academic readers of systematic review (e.g. practitioners, policy-makers), for whom clarity and brevity can be useful. A potential pitfall related to the summary table is the danger of oversimplification of findings, evident in those studies reviewed in Table 2 that offered little descriptive summary to interpret the findings further (e.g. Taskaya & Gül, 2020). Nonetheless, providing it *is* detailed, the descriptive summary serves to temper this danger, offering useful discussion of the general findings alongside example studies and discussion of contextual factors that may be important, as Hammersley (2001) argues is necessary. As such, it is notable that Anderson and Taner's (2023) well-publicised teacher expertise metasummary was reported with reasonable accuracy both in national newspapers (Cordano, 2023; Pinkstone, 2022) and science and education blog posts (e.g. Murray, 2023), all of which drew upon information presented in both the summary table and descriptive summary appropriately.

Nonetheless, as the range of national and educational contexts in which metasummary is being conducted worldwide continues to expand, like all forms of systematic review, there will always be a danger of "neophyte researchers" (Thorne et al., 2004, p. 1343) attempting to use the tools of metasummary without the necessary experiential understanding either of the research processes or the field of study in question. As Hammersley (2001) observes, judgement is always required when conducting systematic review; mechanical rule-following should never be allowed to replace the interpretation of principles. An example of this need for researcher judgement relates to variation in metasummary procedure concerning how key decisions are made, particularly when multiple team members are involved (e.g. whether to code a research report for a specific finding or not). Some authors may choose to prioritise systematicity and transparency, as in Anderson and Taner's (2023) use of independent coding and inter-rater reliability reporting. While more easily replicable, this led to only "moderate agreement" between raters (p. 6) and the possibility that findings of importance were overlooked in their large database. Others may choose to follow Sandelowski and Barroso's (2007) original recommendation for the use of "negotiated consensus". While the latter is inevitably more subjective and opaque (i.e. arguably unsystematic), it can also enable researchers to heed Hammersley's (2001) call for principled judgements to be made. Such negotiation discussions are likely to be particularly useful when dataset sizes are small and it is important to identify all key findings.

Another potential advantage of metasummary results from its potential to be able to include findings or results from all types of research report; qualitative, quantitative or mixed methods. These findings are extracted through a systematic process of coding of the research reports, rather than the primary research data (unlike meta-analysis, for example). Nonetheless, because qualitative studies often present more extensive, "thick description" in their findings (see e.g. Tracy, 2010), they also offer the researcher more opportunities to identify potential findings than quantitative studies, some of which may present only one or two codable results. As such, there is a danger that metasummaries may exhibit a bias towards the qualitative studies in the dataset, particularly because of the assumption that all findings should be presented with equal weighting (Sandelowski & Barroso, 2007), despite the fact that quantitative studies may offer more systematically robust evidence to support the findings identified. Authors wishing to counter this bias may choose to indicate the balance of study types in the summary table presented and identify quantitative studies and their effect sizes in the descriptive summary.

Finally, metasummary has the advantage of being capable of theory generation and theory testing simultaneously, roles that Gough et al. (2012) perceive to be specific to either aggregative and configurative review types respectively. This may occur if the grouping, organisation and presentation of findings is based on inductive coding that leads to a framework, typology or hierarchy emerging as a result of the analysis. Examples of this in the studies reviewed include Kanadlı's (2019) identification of three areas and five dimensions in which STEM education contributes to the science curriculum, and Anderson and Taner's (2023) six domains of teacher expertise, which have already been drawn upon as a theoretical framework by other researchers (e.g. Ries et al., 2024).

Table 3. Guidelines for using metasummary in educational research.

Stage	Description, including minimal requirements	Optional recommendations and dependent elements
1	Formulate study objectives and research questions to clarify and delimit study focus.	• Metasummary is likely to be most useful and more reliable with larger datasets (>20 research reports), hence research questions should be broad and exploratory whenever possible. Larger datasets will also allow for stratified analysis of subsets of data.
2	Establish initial inclusion criteria, and report these transparently, covering "topical (what), population (who), temporal (when), and methodological (how) parameters" (Sandelowski & Barroso, 2007, p. 35).	• Wherever possible, metasummary should be inclusive of all methodologies (quantitative, qualitative and mixed methods). If the dataset is more restricted (e.g. only qualitative), a justification for rejecting other methodologies should be included.
3	Conduct the widest possible literature search for relevant studies, consistent with recommendations for systematic reviews (e.g. Page et al., 2021). Report on all databases searched, search terms and languages used.	• Use a flow diagram to summarise literature search procedure. • Conduct forward and backward citation checks. • If possible, multilingual research teams may search in several languages (e.g. Chinese, English, Spanish). Advances in machine translation may also soon make this possible for monolingual research teams. • Any studies falling in grey areas between inclusion and exclusion criteria should be assessed by multiple researchers for inclusion.
4	If two or more publications report on findings from a single study or dataset, lump these into a single "research report" to ensure that one dataset does not exert a disproportionate influence on findings.	• If necessary, contact research report authors to clarify whether different publications on the same project repeat any findings.
5	Avoid pre-filtering research reports for quality (e.g. non-peer-reviewed studies should be retained in the dataset). Any choice to pre-filter should be justified.	• Because the presentation of findings in metasummary gives greater weight to those found most often, any erroneous findings can be de-emphasised or excluded from the report (e.g. through use of a frequency count threshold for reporting; see below).
6	Coding of research reports should involve two or more coders. Coders should be sufficiently experienced (or trained) to enable them to separate the empirical findings of a research report from the report's discussion of findings from prior studies (e.g. in the literature review) or the authors' extrapolations beyond the findings (e.g. in the discussion). Ensure definitions for themes with higher levels of inference are agreed upon by the coding team. Organise related codes into topic areas or domains if required. Report coding procedure transparently.	• If the anticipated number of findings is likely to be high (e.g. over 20), consider developing an initial coding framework before coding itself begins (Anderson & Taner, 2023). This can be done inductively by reading through a sufficient sample of the dataset to ensure that the majority of potentially relevant findings are identified (code saturation). For added transparency, a coding framework may be included in the appendices. • Different coding procedures are likely to suit different projects, dependent on research focus, complexity of constructs involved, variety and extent of research reports in the dataset. • If more robust interpretive validity is prioritised, coders may opt to consult at the end of an independent coding phase, discuss disagreements, and come to an agreement (Sandelowski & Barroso's "negotiated consensus"; 2007). This option may also work well in the case that only a small number of research reports are available. • If greater rigour and replicability are prioritised, coders may opt to work independently. In such cases, a theme may be assigned to a specific research report only if two coders have independently assigned that code to the report (Anderson & Taner's "independent agreement criterion"; 2023). Inter-rater reliability can be used to assess coding reliability.

(*Continued*)

Table 3. Continued.

Stage	Description, including minimal requirements	Optional recommendations and dependent elements
		• Studies involving only one principle coder (e.g. single-authored PhD studies) should employ assistant coders to code an appropriate sample of the data after training (e.g. > 20%) to assess reliability of the coding procedure.
7	Consider carefully which findings to report.	• If studies have not been pre-filtered for quality, findings identified in only one research report should be reported on with caution. • If required, an agreement count threshold (see Anderson & Taner, 2023) can be set to establish which findings to report to avoid reporting false positives.
8	Ensure that both summary table and descriptive summary are included in the metasummary research report.	• The summary table should order findings according to frequency counts as a useful measure of reliability of findings. • The descriptive summary should offer a description and discussion of findings presented in the summary table, informed by the researchers' understanding of the dataset gained from conducting the review. Note any variations within the dataset and indicate potential reasons for these variations (e.g. contextual differences, influence of different research objectives, etc.). • Consider Sandelowski and Barroso's (2007) recommendation to report "frequency effect sizes" and "intensity effect sizes" with critical caution, given that these are unlikely to be appropriate in education, and may be misinterpreted by audiences. This is particularly important if standard effect sizes (e.g. using Cohen's D) from research reports included in the dataset are also discussed in the descriptive summary.
9	Present and discuss limitations as appropriate.	• In addition to common limitations of systematic reviews, this may also include challenges faced when extracting findings and when deciding upon what to report.

Guideline recommendations for conducting metasummary in education

Given evident variation in quality in the small but increasing number of metasummaries in education documented in Table 2, this article here offers initial, contingent guidelines for conducting metasummary in education. The guidelines draw extensively upon Sandelowski and colleagues' tried and tested recommendations (Sandelowski et al., 2006, 2007; Sandelowski & Barroso, 2007; Voils et al., 2008), but also discuss a range of alternative pathways and procedures specific to educational research, including personal recommendations based on lessons learnt from conducting one of the studies presented in Table 2 (Anderson & Taner, 2023) and from reviewing the other educational metasummaries in Table 2. These guidelines are presented in Table 3.

The guidelines include recommended obligatory steps (in column 2) which are useful for all metasummaries in education, as well as optional or dependent elements (column 3), which may vary depending on the focus, scope, complexity and available resources for a given metasummary project. The guidelines are intended to help research teams without being overly prescriptive – this is important and necessary given how rarely metasummary has been used in education to date and the diversity of uses already apparent.

Conclusion

This article has discussed the process, justification and merits of metasummary, a systematic review methodology originating in healthcare research but also likely to be useful in educational research, where it has been relatively neglected. It has identified and reviewed the small number of educational metasummaries conducted to date, which, given the variety of objectives, topic areas and sample sizes involved, strongly suggests the possibility for more widespread application in the field. It has discussed two potential underlying philosophies (pragmatism and critical realism), with which metasummary is argued to be compatible, and it has offered discussion of its potential benefits and weaknesses as a systematic review methodology. Based on the review involved, the article has also offered a number of guidelines as an initial, contingent framework intended to help other researchers interested in bringing together the findings of qualitative, quantitative and mixed methods studies on a given topic in education systematically and informatively. This methodologically inclusive feature of metasummary constitutes a potential means to bridge – albeit cautiously – the paradigm divide that has fragmented educational research communities to the detriment of our wider audiences beyond academia. In this sense, metasummary may be a useful addition to systematic review methodology in education.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID


Jason Anderson http://orcid.org/0000-0002-3400-0135

References

Ajlani, A., Laube, S., & Meyer, U. (2024). Sociotechnical experiences and strategies of people with reading/writing disorders in higher education: A systematic literature review in sociology. In K. Miesenberger, P. Peňáz, & M. Kobayashi (Eds.), *International conference on computers helping people with special needs* (pp. 275–284). Springer Nature.

Alexander, R. J. (2015). Teaching and learning for all? The quality imperative revisited. *International Journal of Educational Development*, *40*, 250–258. https://doi.org/10.1016/j.ijedudev.2014.11.012

Al-Salman, S. (2013). *The role of faculty in the effectiveness of fully online programs* [Doctoral dissertation]. https://core.ac.uk/download/pdf/51097824.pdf

Anderson, J. (2023). *Teacher expertise in the global south: Theory, research and evidence*. Cambridge University Press. https://doi.org/10.1017/9781009284837

Anderson, J., & Taner, G. (2023). Building the expert teacher prototype: A metasummary of teacher expertise studies in primary and secondary education. *Educational Research Review*, *38*, 100485. https://doi.org/10.1016/j.edurev.2022.100485

Andrews, R. (2005). The place of systematic reviews in education research. *British Journal of Educational Studies*, *53*(4), 399–416. https://doi.org/10.1111/j.1467-8527.2005.00303.x

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, *31*(5), 625–640. https://doi.org/10.1080/07294360.2012.702735

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*(1), 1–22. https://doi.org/10.1111/j.1741-5446.2006.00241.x

Biesta, G. (2015). Freeing teaching from learning: Opening up existential possibilities in educational relationships. *Studies in Philosophy and Education*, *34*, 229–243. https://doi.org/10.1007/s11217-014-9454-z

Bruner, J. (1996). *The culture of education*. Harvard University Press.

Chapman, K. (2021). Characteristics of systematic reviews in the social sciences. *The Journal of Academic Librarianship*, *47*(5), 102396. https://doi.org/10.1016/j.acalib.2021.102396

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Routledge.

Cook, D. J., Sackett, D. L., & Spitzer, W. O. (1995). Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *Journal of Clinical Epidemiology*, *48*(1), 167–171. https://doi.org/10.1016/0895-4356(94)00172-M

Cordano, M. (2023, March 20). *Los mejores profesores crean relaciones sólidas con sus alumnos y saben lidiar con lo inesperado*. El Mercurio.

Davey, J., Turner, R. M., Clarke, M. J., & Higgins, J. P. (2011). Characteristics of meta-analyses and their component studies in the cochrane database of systematic reviews: A cross-sectional, descriptive analysis. *BMC Medical Research Methodology*, *11*, 1–11. https://doi.org/10.1186/1471-2288-11-160

Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, *26*(3–4), 365–378. https://doi.org/10.1080/713688543

Demir, Y. (2018). A meta-summary of the research on metaphors regarding Turkish EFL teachers' roles. *Advances in Language and Literary Studies*, *9*(2), 65. https://doi.org/10.7575/aiac.alls.v.9n.2p.65

Denner, J., Marsh, E., & Campe, S. (2017). Approaches to reviewing research in education. In D. Wyse, N. Selwyn, E. Smith, & E. Larry (Eds.), *The BERA/Sage handbook of educational research* (pp. 143–164). Sage. https://doi.org/10.4135/9781473983953

Dominguez, N., & Hager, M. (2013). Mentoring frameworks: Synthesis and critique. *International Journal of Mentoring and Coaching in Education*, *2*(3), 171–188. https://doi.org/10.1108/IJMCE-03-2013-0014

Eberhardt, A. K. (2021). *School personnel perspectives on LGBTQ bullying: A qualitative meta-summary* [PhD thesis]. https://www.proquest.com/docview/2604826606

EPPI Centre. (2019). *What is a systematic review?* http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=67

Evans, J., & Benefield, P. (2001). Systematic reviews of educational research: Does the medical model fit? *British Educational Research Journal*, *27*(5), 527–541. https://doi.org/10.1080/01411920120095717

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 665–694. https://doi.org/10.1348/000711010X502733

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.2307/1174772

Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews*, *1*, 1–9. https://doi.org/10.1186/2046-4053-1-28

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart, Lung and Vessels*, *5*, 219–225.

Habib, A., & Pius, A. (2023). Flexible learning approach: A systematic review of pedagogical tools for higher education. *Social Science Research Network*. Advance online publication. https://doi.org/10.2139/ssrn.4574286

Haidiri, S. M., & Koçoğlu, A. (2024). Do the learners are inclining [*sic*] towards the student credit card scheme? The reality and its prospect. *International Journal of Educational Innovation and Research*, *3*(1), 1–16. https://doi.org/10.31949/ijeir.v3i1.6720

Hammersley, M. (2001). On 'systematic' reviews of research literatures: A 'narrative' response to Evans & Benefield. *British Educational Research Journal*, *27*(5), 543–554. https://doi.org/10.1080/01411920120095726

Hammersley, M. (2004). Some questions about evidence-based practice in education. In R. Pring & G. Thomas (Eds.), *Evidence-based practice in education* (pp. 133–149). McGraw-Hill Education.

Harden, A., & Thomas, J. (2005). Methodological issues in combining diverse study types in systematic reviews. *International Journal of Social Research Methodology*, *8*(3), 257–271. https://doi.org/10.1080/13645570500155078

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Heyvaert, M., Hannes, K., & Onghena, P. (2017). *Using mixed methods research synthesis for literature reviews*. SAGE Publications.

Heyvaert, M., Maes, B., & Onghena, P. (2013). Mixed methods research synthesis: Definition, framework, and potential. *Quality & Quantity*, *47*(2), 659–676. https://doi.org/10.1007/s11135-011-9538-6

Higgins, S. (2018). *Improving learning*. Cambridge University Press.

Jensen, L. A., & Allen, M. N. (1996). Meta-Synthesis of qualitative findings. *Qualitative Health Research*, *6*(4), 553–560. https://doi.org/10.1177/104973239600600407

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Kamil, D. (2021). Embracing the 2013 curriculum: A qualitative metasummary of problems. *Indonesian Research Journal in Education*, *5*, 260–278. https://doi.org/10.22437/irje.v5i1.12499

Kanadlı, S. (2019). A meta-summary of qualitative findings about STEM education. *International Journal of Instruction*, *12*(1), 959–976. https://doi.org/10.29333/iji.2019.12162a

Maeda, Y., Caskurlu, S., Kenney, R. H., Kozan, K., & Richardson, J. C. (2022). Moving qualitative synthesis research forward in education: A methodological systematic review. *Educational Research Review*, *35*, 100424. https://doi.org/10.1016/j.edurev.2021.100424

Maxwell, J. (2012). *A realist approach for qualitative research*. Sage Publications.

Maxwell, J., & Mittapalli, K. (2010). Realism as a stance for mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social & behavioral research* (pp. 145–168). SAGE Publications.

Mukumbang, F. C. (2023). Retroductive theorizing: A contribution of critical realism to mixed methods research. *Journal of Mixed Methods Research*, *17*(1), 93–114. https://doi.org/10.1177/15586898211049847

Murray, J. (2023, January 17). *Casting a wide net to help define teacher expertise*. The Thomas B. Fordham Institute. https://fordhaminstitute.org/ohio/commentary/casting-wide-net-help-define-teacher-expertise

Nkoana, L. N., & Ebersöhn, L. (2023). A meta-summary review of enabling school-based interventions in highly challenged schools in a postcolonial, global south space. In V. Scherman & L. Liebenberg (Eds.), *African schools as enabling spaces* (pp. 76–95). Routledge.

Novita, L. D., Sarkadi, & Maksum, A. (2021). Group investigation learning in developing 21st century skills of elementary school students. *International Journal of Multicultural and Multireligious Understanding*, *8*, 268–278. https://doi.org/10.18415/ijmmu.v8i6.2751

Oakley, A. (2000). *Experiments in knowing: Gender and method in the social sciences*. Polity Press.

Onwuegbuzie, A. J. (2003). Effect sizes in qualitative research: A prolegomenon. *Quality and Quantity*, *37*, 393–409. https://doi.org/10.1023/A:1027379223537

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/10.1177/0141076807100012020

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., & McGuinness, L. A. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, *372*(71), 1–9. https://doi.org/10.1136/bmj.n71

Paterson, B. (2013). Metasynthesis. In C. T. Beck (Ed.), *Routledge international handbook of qualitative nursing research* (pp. 331–346). Routledge.

Pileggi, F. (2024). The contribution of visual aesthetic experiences conducted through AItech to educate critical and creative thinking in the primary school: The field of visual arts. In

E. Guarcello & A. Longo (Eds.), *School children and the challenge of managing AI technologies* (pp. 102–118). Routledge.

Pinkstone, J. (2022, December 19). Why the best teachers "think on their feet." *Daily Telegraph*, 2.

Plonsky, L., & Oswald, F. (2012). How to do a meta-analysis. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 275–295). Blackwell. https://doi.org/10.1002/9781444347340

Pluye, P., Gagnon, M.-P., Griffiths, F., & Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. *International Journal of Nursing Studies*, *46*(4), 529–546. https://doi.org/10.1016/j.ijnurstu.2009.01.009

Pollock, M., Fernandes, R. M., Becker, L. A., Featherstone, R., & Hartling, L. (2016). What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. *Systematic Reviews*, *5*(1), 190. https://doi.org/10.1186/s13643-016-0367-5

Porta, M. S. (2014). *A dictionary of epidemiology* (6th ed.). Oxford University Press.

Ries, N., Wolf, K., Baier-Mosch, F., Roth, A., & Kunter, M. (2024). Cooperative learning before and during the COVID-19 pandemic: The predictive power of teacher expertise. *Learning Environments Research. Advance online Publication*, https://doi.org/10.1007/s10984-024-09507-y

Saito, H., Sawaki, Y., & Kasahara, K. (2022). Evaluating fairness and justice of external English language test score interpretation and Use for Japanese university admission. *Language Assessment Quarterly*, *19*(4), 422–448. https://doi.org/10.1080/15434303.2022.2083965

Sandelowski, M., & Barroso, J. (2003). Creating metasummaries of qualitative findings. *Nursing Research*, *52*(4), 226–233. https://doi.org/10.1097/00006199-200307000-00004

Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. Springer Pub. Co.

Sandelowski, M., Barroso, J., & Voils, C. I. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in Nursing & Health*, *30*(1), 99–111. https://doi.org/10.1002/nur.20176

Sandelowski, M., Docherty, S., & Emden, C. (1997). Qualitative metasynthesis: Issues and techniques. *Research in Nursing & Health*, *20*(4), 365–371. https://doi.org/10.1002/(SICI)1098-240X(199708)20:4<365::AID-NUR9>3.0.CO;2-E

Sandelowski, M., Voils, C., & Barroso, J. (2006). Defining and designing mixed research synthesis studies. *Research in the Schools*, *13*, 29.

Sandelowski, M., Voils, C. I., Leeman, J., & Crandell, J. L. (2012). Mapping the mixed methods–mixed research synthesis terrain. *Journal of Mixed Methods Research*, *6*(4), 317–331. https://doi.org/10.1177/1558689811427913

Şanlı, B., & Karakuş, F. (2023). A meta-summary of qualitative findings on gender equality education. *The Journal of Academic Social Science Studies*, *16*(94), 39–54. https://doi.org/10.29228/JASSS.68340

Shorey, S., Min, C., & Yap Seng, C. (2023). Turbulent academic journey of female academics: A meta-synthesis. *International Journal of Inclusive Education*. Advance online publication, 1–21. https://doi.org/10.1080/13603116.2023.2227862

Stern, C., Lizarondo, L., Carrier, J., Godfrey, C., Rieger, K., Salmond, S., Apostolo, J., Kirkpatrick, P., & Loveday, H. (2021). Methodological guidance for the conduct of mixed methods systematic reviews. *JBI Evidence Implementation*, *19*(2), 120–129. https://doi.org/10.1097/XEB.0000000000000282

Taskaya, S., & Gül, A. (2020). An analysis on teaching principles in "teaching principles and methods" books. *International Journal of Education Technology and Scientific Researches*, *5*(11), 550–579. https://doi.org/10.35826/ijetsar.83

Teddlie, C., & Tashakkori, A. (2011). Mixed methods research: Contemporary issues in an emerging field. In N. Denzin & Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 285–300). SAGE Publications.

Thorne, S., Jensen, L., Kearney, M. H., Noblit, G., & Sandelowski, M. (2004). Qualitative metasynthesis: Reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, *14*(10), 1342–1365. https://doi.org/10.1177/1049732304269888

Tibingana-Ahimbisibwe, B., Willis, S., Catherall, S., Butler, F., & Harrison, R. (2022). A systematic review of peer-assisted learning in fully online higher education distance learning programmes. *Open Learning: The Journal of Open, Distance and E-Learning*, *37*(3), 251–272. https://doi.org/10.1080/02680513.2020.1758651

Tracy, S. J. (2010). Qualitative quality: Eight "Big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, *16*(10), 837–851. https://doi.org/10.1177/1077800410383121

Vieira, G., & Finardi, K. (2018). The case of EFL: Does technology lead to learning? *Raído-Revista Do Programa de Pós-Graduação Em Letras Da UFGD*, *12*(30), 197–210.

Vieira, G. V., Finardi, K. R., & Piccin, G. F. O. (2018). Going international: The challenges for the Brazilian federal institutes of education, science and technology. *Revista Ibero-Americana de Estudos Em Educação*, *13*(1), 391–406. https://doi.org/10.21723/riaee.nesp1.v13.2018.11428

Vincent, S., & O'Mahoney, J. (2019). Critical realism and qualitative research: An introductory overview. In C. Cassell, A. L. Cunliffe & G. Grandy (Eds.), *The SAGE handbook of qualitative business and management research methods: History and traditions* (pp. 201–216). SAGE Publications. https://doi.org/10.4135/9781526430212

Voils, C. I., Sandelowski, M., Barroso, J., & Hasselblad, V. (2008). Making sense of qualitative and quantitative findings in mixed research synthesis studies. *Field Methods*, *20*(1), 3–25. https://doi.org/10.1177/1525822X07307463

Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing*, *50*(2), 204–211. https://doi.org/10.1111/j.1365-2648.2005.03380.x

Waterfield, J. (2018). Literature review. In B. B. Frey (Ed.), *The Sage encyclopedia of educational research, measurement, and evaluation* (pp. 983–987). Sage. https://doi.org/10.4135/9781506326139

Zachariadis, M., Scott, S., & Barrett, M. (2013). Methodological implications of critical realism for mixed-methods research. MIS Quarterly, 37(3), 855-879. https://doi.org/10.25300/MISQ/2013/37.3.09