Coverage and similarity of bibliographic databases to find most relevant literature for systematic reviews in education

Tamara Heck[1] · Carolin Keller[1] · Marc Rittberger[1,2]

Abstract

Systematic literature reviews in educational research have become a popular research method. A key point hereby is the choice of bibliographic databases to reach a maximum probability of finding all potentially relevant literature that deals with the research question analyzed in a systematic literature review. Guidelines and handbooks on review recommend proper databases and information sources for education, along with specific search strategies. However, in many disciplines, among them educational research, there is a lack of evidence on the relevance of databases that need to be considered to find relevant literature and lessen the risk of missing relevant publications. Educational research is an interdisciplinary field and has no core database. Instead, the field is covered by multiple disciplinary and multidisciplinary information sources that have either a national or international focus. In this article, we discuss the relevance of seven databases in systematic literature reviews in education, based on results of an empirical data analysis of three recently published reviews. To evaluate the relevance of a database, the relevant literature of those reviews served as the gold standard. Results indicate that discipline-specific databases outperform international multidisciplinary sources, and a combination of discipline-specific international and national sources is most efficient in finding a high proportion of relevant literature. The article discusses the relevance of the databases in relation to their coverage of relevant literature, while considering practical implications for researchers performing a systematic literature search. We, thus, present evidence for proper database choices for educational and discipline-related systematic literature reviews.

Keywords Systematic review · Research database · Coverage · Similarity · Educational research

1 Introduction

Systematic literature reviews (SLR) are a research method used to synthesize the current state of research for a specific question or topic and summarize research findings in a systematic way. Given the increasing number of scientific publications and studies, systematic reviews aim at an overview of relevant research findings [1, 2]. While aggregating relevant research outputs, such reviews are used to make evidence-based decisions in research, practice, and politics [3].

Methodological approaches for SLR should be transparent and guarantee comprehensiveness over the condensed outputs. For several disciplines standards and guidelines are provided, e.g., by Cochrane [4] specifically for medical systematic reviews and by the Evidence for Policy and Practice Information and Co-coordinating Centre (EPPI), for educational research [5]. Besides, guidelines to enhance the quality of searching literature for reviews exist, such as the guide to information retrieval and searching for studies by the Campbell Collaboration [6]. Those guidelines describe best-practice approaches to conduct searches for reviews, but evidence-based guidance on relevant sources to guarantee a high recall and precision rate of relevant literature is still lacking for specific disciplines like educational research.

Cochrane [4] recommends explicit databases to search health and medical literature for reviews. Those sources are said to be sufficient to cover most of the relevant literature in this discipline [7]. However, for the social sciences

✉ Tamara Heck
  heck@dipf.de

1 DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

2 Darmstadt University of Applied Sciences, Darmstadt, Germany

and humanities, there is a plurality of sources. This circumstance challenges the choice of "the proper" databases to conduct systematic reviews in those disciplines. For educational research, an interdisciplinary field, this becomes even more challenging as many relevant questions worth synthesizing in reviews, is researched in several disciplines such as pedagogics, psychology, and sociology. Thus, relevant literature is spread across numerous databases and other web sources. As systematic reviews have become more significant in educational research [8, 9], there is a lack of knowledge of optimal databases and database combinations for reviews.

This article discusses results from data analyses of three SLR datasets. It extends the results and discussion published in the paper by Keller et al. [10]. The goal of the studies was to investigate the effects of database choices on finding relevant literature for SLR in education and argue for more evidence on the relevance of different information sources for SLR.

Our research questions are as follows:

- Which databases index relevant literature for SLR in education?
- Which combination of databases most efficiently covers all relevant literature?

Section 2 introduces systematic reviews and search guidelines and discusses the differences in bibliographic databases and their coverage of disciplines. In Sect. 3, we describe our method and the open review datasets we applied, before we show and discuss the results in Sect. 4 and address practical implications in Sect. 5. Section 6 concludes the paper.

2 State-of-the-art for systematic reviews

SLR are meant to give an overview of the most relevant literature published on a question or topic and help make evidence-based decisions, e.g., identifying research gaps to be investigated, or informing practitioners to initiate changes in performance and implementations of processes. Several types of review approaches exist [11], and with it a diverse terminology, such as scoping review or critical review. As such, the review types are based on different methodological approaches relating to the needs and objectives of the researchers, and those are continuously being developed and improved [12]. The systematic reviews approach is the gold standard of a research review. Systematic reviews aim at systematically reviewing relevant research. The four key activities include "clarifying the question […], identifying and describing the relevant research […], critically appraising research reports in a systematic manner […], known as synthesis; and establishing what evidence claims can be made from the research" [13].

2.1 Search strategies in review guidelines

Guidelines and handbooks list detailed and structured steps for systematic reviews [8, 14–17]. Specifically, for all types of reviews, a systematic approach to the literature search is crucial in order to avoid bias and to ensure the replicability of the method [2, 13, 18]. Reviews can only provide reliable evidence based on a completely searched dataset. Consequently, whoever is accountable for the search for relevant literature, bears a high responsibility for the quality and validity of a review. Thus, many guidelines recommend consulting an experienced information professional or librarian, such as Cochrane Collaboration and Campbell Collaboration [6, 19], which "recognize the importance and value of consulting with an information specialist during the (un)systematic information retrieval stage of the review process" [20, p. 115]. Studies investigated the impact of information specialists on the quality of search methods in reviews, mainly with a focus on the transparency and documentation of the search process. For example, there are studies on how often librarians were mentioned or listed as co-authors in review papers. According to the studies, librarians have an influence on the review process, especially those reviews show a better reproducibility of the literature searches and more database sources are used, which is relevant with respect to completeness [21–24].

A further step recommended in all handbooks is the conduction of a test search to modify the search strategy and selection of databases at the beginning. Furthermore, investigators should consider advanced search options and syntax of selected databases. A further revision of search strategies might be necessary after reviewing results. Additionally, instructions for an extensive documentation of all steps guarantee the replication of the method in the conducted review [2, 6, 15]. Most of the guidelines, however, focus on reviews in the fields of health and medicine, like the PICO framework [25], whose structured inclusion criteria are not always adaptable for searches in other disciplines like educational research [3]. Only a very few publications recognize the specifics of the information infrastructure in educational research [6, 8]. In addition, many of the decisions for the literature search depend on the research question and methodology so a general specification is not possible. For example, the selection of the document types depends on the purpose and scope of the planned review. Investigators need to decide if they only include peer-reviewed journal articles or also gray literature.

For assessing the quality of the literature search itself, a peer review for electronic search strategies (PRESS) exists [26]. PRESS defines quality criteria for database searches, e.g., the services shall allow Boolean operators to formulate proper queries, the selection of search words should be considered and different functions of the databases. Besides, however, this strategy does not give any explicit criteria

for database selection, but only generally lists available databases.

2.2 Criteria for database selection

A crucial element for characterizing bibliographic databases is their coverage [27]. Rittberger and Rittberger specifically name scope and coverage as important subject-related criteria for the quality of databases [28]. For a complete search coverage, the reviewer needs to choose databases carefully with respect to these criteria: They need to cover all relevant literature, i.e., all types of documents (not only journal articles) with regard to the review's topic, question, or discipline investigated. In some disciplines, the geographic coverage of the database may play an important role, e.g., for educational science, when reviews focus on questions concerning the national educational system. Some review guidelines do give explicit database recommendations. Cochrane, who sets standards for health and medical reviews, recommends Central, Medline, and Embase [4]. Although study results slightly differ and researchers recommend [29, 30] or indicate the trend [31] to apply several databases in medical and health sciences, others conclude that the "majority of relevant studies [for medicine and life sciences] can be found in a limited number of databases" [32]. Moreover, with a proper search strategy applying Boolean search and ranking, searching Medline as only source might be sufficient [7]. Thus, the discipline seems to have a manageable and known set of databases that cover relevant literature properly.

Similarly for the social and educational sciences, various guidelines and textbooks have provided lists of selected sources, often named are ERIC, the Web of Science, and FIS Bildung [2, p. 111 ff, 6, p. 47 ff]. However, those guidelines strongly indicate that relevant literature is found in a variety of different sources, i.e., not only in the major bibliographic databases, but also in research registries, search engines, on websites of important institutions, and through hand searches [33, p. 107 ff]. In contrast with the quite large number of studies investigating databases for reviews in the medical and health sciences, currently, there is a lack of evidence-based research on the impact of sources on reviews in the social and educational sciences.

Besides the large bibliographic databases, other sources mentioned by the guidelines for reviews in the social and educational sciences do not only cover journal articles, but as well other document types potentially relevant for questions and topics in those disciplines. Educational research is highly interdisciplinary with heterogeneous study designs. Multiple disciplines address research on education and learning and teaching, such as pedagogics, psychology, and sociology. The publication culture varies within the field, reaching from journal article publications popular, e.g., in psychology, to essay collections, books, and reports from practice

[3]. Much of that literature cannot be found in bibliographic databases that often include journal articles only. Moreover, many international databases do not cover social sciences and humanities literature properly. A study on German university profiles showed that the coverage of the Web of Science with regard to social science literature, including educational research, is less than 50% [34]. Other studies found similar results and conclude that the Web of Science over represents English language publications for those disciplines [35]. A study that compared Google Scholar with the Web of Science, Microsoft Academic, Scopus, Dimensions, and the database COCI by Open Citations showed that all of the five international services have a limited coverage of the social sciences and humanities, not exceeding 50% of analyzed citations [36]. Google Scholar performed best, but it comes with some drawbacks for conducting reviews, as we will discuss later.

As the named larger international bibliographic databases cannot offer a satisfying coverage of the literature for all disciplines, educational research reviewers need to draw on discipline-specific databases and additional web searches. ERIC, provided by the Education Resources Information Centre, is often named as one of the main bibliographic and full-text databases for educational research publications [9]. The Centre collects journals and non-journal sources according to its selection policy (ERIC 2018). However, it only indexes English language articles, crucial fact educational researchers need to consider, specifically when they investigate questions with a national focus. For German language educational literature, the database of references FIS Bildung (German Education Index) is a relevant source. The database is hosted in Germany and subject to cooperation agreement, about 30 partners collect and index educational research literature.

Besides a high recall, databases need to provide the necessary functionalities for a systematic review search, like allowing Boolean operators, an option to enter complex search strings o or filtering via metadata fields, optimally based on controlled vocabulary. Here as well, medicine profits from the well-kept Medical Subject Headings (MeSH) thesaurus available in all larger medicine databases such as Medline or PubMed. Whereas a good database's coverage raises recall, search functionalities can boost precision. Both are relevant and not always provided, as is often shown for Google Scholar which generally has a high coverage of research literature, but precision is very low [18], and it is unclear which and how many publications Google Scholar explicitly includes [37, 38]. Boeker et al. applied "realistic search expressions" from published reviews to show the effects of Google Scholar's limited search options [18]. The authors conclude that the database "does not provide necessary elements for systematic scientific literature retrieval" [18], including optimizing queries and exporting references. In the following study, we included Google Scholar to show

the effect as well, but report on a review where the information professionals omitted the database in the second search phase due to the limited search functionalities.

3 Method

To investigate the relevance of databases, we choose relevant literature from three review studies as the gold standards and analyzed their coverage in seven databases often applied in SLR in education.

3.1 Datasets

The chosen datasets of 15 reviews are part of the cooperative project "digitizing in education," which aims at investigating central aspects of digital learning in five educational sectors. One of our authors was co-responsible for conducting the review searches. The project description says that it is meant to conduct critical reviews with narrative overviews that summarize essential findings for each specific research question within the project. Systematic literature searches were performed, and data were published at a research data center [39, 40]. The results were published in two proceedings [41, 42], while a third proceeding is being prepared. We chose those datasets because they are a good example of reviews in educational research: They refer to a current relevant topic, digitization in education, and investigate sub-reviews on three research questions in relation to five different educational sectors.

Splitting a broad research topic into several sub-reviews is common in the social sciences and called mixed or multicomponent reviews [43]. The first question asks about the role of pedagogical staff in implementing digital devices (SLR1). The second question asks about organizational development in educational institutions (SLR2). The third question concerns didactics and teaching (SLR3). For each question, the sectors are early childhood, general education, vocational education, adult and teacher education. Thus, for each research question, the researchers compiled five reviews in different search phases between February 2020 and March 2022 (Table 1).

The literature inclusion criteria were German and English resources and a publication date later than 2016. In contrast with other reviews that often only include peer-reviewed journal articles, the publication type was not restricted due to the publication culture in educational research, where other types of publications are often most relevant [44, p. 114]. For further details of the search, we refer to the original data documentation [39, 40] and proceedings [41, 42]. In the following, we refer to SLR1 in relation to the first review question and

Table 1 Topics and educational sectors of the 15 reviews

Topic	Educational sector
	School
SLR1: digital services applied by pedagogical staff	Early childhood
SLR2: organizational development regarding digitalization	Teacher
SLR3: didactics and teaching regarding digitalization	Adult
	Vocational

the five sub-reviews for each educational sector, and, respectively, SLR2 and SLR3 in relation to the second and third review questions and their five sub-reviews.

3.2 Databases in the original SLR studies

In SLR1, two discipline-specific information sources for educational research were chosen, i.e., FIS Bildung and ERIC, additionally, the German National Library (DNB) and Google Scholar. Based on their experiences made during the search and screening of the retrieved literature in SLR1, the investigators expanded and adapted the choice of databases for the SLR2 and SLR3 to cover more discipline-related and multidisciplinary research [41, 42]. They added the Web of Science and Education Research Complete, as well as LearnTechLib, which indexes research reports relevant for the investigated questions. Moreover, Google Scholar was excluded due to the low precision rate of searches and limited search functionalities [18].

For the vocational and adult education sector, the reviewers added sources in all three SLR. This was due to the experienced low number of search results in the primarily chosen databases. Additional sources for vocational education were the VET Repository and Library, offered by the Federal Institute for Vocational Education and Training (BIBB). In SLR3, the search included the Social Science Open-Access Repository (SSOAR) and the Sociology Information Service (SocioHub), however, only in relation to the vocational sector. By using the additional sources for vocational education, one relevant publication in SLR1 was found and two in SLR2 and SLR3, respectively. Due to the low number of search results for adult education, the Bielefeld Academic Search Engine (BASE) was searched in SLR3 to identify further relevant publications—this led to one study additionally identified in BASE.

Besides the database searches, advanced search strategies were applied like hand searching (search on websites from institutions and associations) and citation searching based on relevant authors. Sixty-four documents were originally added

Table 2 Number of relevant publications found via search strategies applied additionally to database searches

	SLR1	SLR2	SLR3
Added by authors	28	20	5
Hand search	4	–	–
Author search	–	7	–

by advanced search strategies such as hand, and author search done by the information specialist or the review authors themselves. These are 32 in SLR1, 27 in SLR2, and 5 in SLR3[1] (Table 2).

For the data analyses presented in this article, we investigated the main databases searched either one of the three SLR, except the VET Repository, BASE, SSOAR, and Socio-Hub, which were only applied for specialized educational sectors to find additional publications, and not as core sources to find relevant literature for the SLR. Table 3 shows the databases analyzed.

3.3 Analytic approach

For the following study, we searched 445 publications from SLR1, SLR2, and SLR3. Those publications were considered relevant by the expert researchers involved in the reviews. They mark the final datasets synthesized according to the research questions. Those relevant publications are our gold standard. We searched for each relevant publication in the seven databases searched in any of the original reviews. We conducted our searches in December 2021 for SLR1 and SLR2, the results are published in Keller, Heck, and Rittberger [10]. We did the search for SLR3 in September 2022. We used title and author details and DOIs to match and validate the search. The dataset of our search is available at OSF [45]. Please note that the figures concerning relevant publications per database in our dataset differ from the figures in the original data [39, 40]. The original data show the source of a publication after deduplication. Using these data leads to a bias in analyzing database coverage.

For measuring the relevance of databases, we applied the following indicators.

The *coverage* of a database indicates which relevant literature known to a user a database includes [46, p. 83]. We measured the coverage with:

$$\text{coverage CO} = \frac{|A|}{|U|}, \tag{1}$$

where $|A|$ is the number of retrieved relevant documents in database A, and $|U|$ is the number of relevant documents known, i.e., our gold standard.

For the measurement of *similarity* of coverage in databases, we used the cosine coefficient as a common similarity coefficient:

$$\text{Similarity SI} = \frac{|C|}{\sqrt{|A||B|}}, \tag{2}$$

where $|C|$ is the number of common relevant documents found in two databases, and $|A|$ and $|B|$ are the number of retrieved relevant documents in databases A and B, respectively.

To measure the effect of *combined database* search, we count relevant documents at least indexed in one of the two databases:

$$\text{Combination CB} = \frac{|A| + |B| - |C|}{|U|}. \tag{3}$$

4 Results and discussion

In the following, we will show our results and discuss them referring to the research questions, before we turn to practical implications of systematic reviews.

4.1 Coverage of databases

Overall, we found 445 relevant documents in the seven databases for all three reviews. However, nine documents of those—six from SLR1, two from SLR2, and one from SLR3—do not seem to be indexed in any of the seven databases (Table 4). Table 5 shows the number of relevant documents in the original SLR data. The number of final relevant papers seem to vary, specifically comparing SLR1 and SL2 with SLR3, which seems to focus on the first three sectors in Table 5.

As other studies show [47], the relevant literature retrieved differs for the seven databases. In comparable studies, 5.7% of the results were indexed in all investigated three databases (WoS, Scopus, and EBSCO) [47].

In our analysis, none of the relevant documents was retrieved in all seven databases for all three SLR. For SLR1, less than 1% was retrieved in six, 25.74% in five, and about 20% in four, three, or two databases, respectively. For SLR2, no document was found in six databases, and between 10 and 17% in either five, four, three, or two databases. For SLR3, about 3% was found in six databases, 17% in five databases, 28 1% in four databases, and about 32% in three databases (Table 6).

---

[1] In our previous publication [10], the three publications found via the VET Repository were counted as "found via hand search", which led to the number 62. This has been modified.

Table 3 Investigated databases and acronyms

DNB*	Catalog German National Library (dnb.de)
ERC[+!]	Education Research Complete (ebsco.com/products/research-databases/education-research-complete)
ERIC*[+!]	Education Resources Information Center (eric.ed.gov)
FIS*[+!]	FIS Bildung Literaturdatenbank/ German Education Index (available via fachportal-paedagogik.de)
GS*	Google Scholar (scholar.google.de)
LTL[+!]	LearnTechLib (learntechlib.org)
WoS[+!]	Web of Science Social Science Citation Index (webofknowledge.com, conducted via DIPF access)

*Applied in SLR1, [+]applied in SLR2, and [!]applied in SLR3

Table 4 Publications found in databases

	SLR1	SLR2	SLR3
Relevant publications	202	126	117
Publications found in any database	196	124	116
Publication found in any database	97%	98%	99%

Table 5 Relevant documents *U* per educational sector

	SLR1	SLR2	SLR3
School education	122	56	29
Early childhood ed	10	12	35
Teacher education	15	24	38
Adult education	41	23	8
Vocational education	14	11	7

Table 6 Percentage of single publications in different databases for the three SLR

Number of databases	SLR1 (%)	SLR2 (%)	SLR3 (%)
7	0	0	0
6	0.99	0	2.56
5	25.74	10.40	17.95
4	22.77	13.37	29.06
3	19.31	16.83	31.62
2	21.78	11.88	16.24
1	6.44	8.91	1.71
0	2.97	0.99	0.85

Table 7 reports on the analysis of the coverage of the databases. GS coverage is highest for all three SLR, followed by ERIC and ERC, both international and discipline-specific databases focusing on educational research.

The coverage is similar for all three SLR. The average deviation from the average coverage is 10% in one case and below in all other cases. On the level of each SLR, we can say that—regardless of GS as the exceptional web source—the international and discipline-specific databases seem highly relevant to find literature for SLR in education. International multidisciplinary databases (WoS) and more specialized ones like LTL have a lower coverage. The lowest coverage seems to have the national discipline-focused (FIS) or generic databases (DNB) databases.

We come to different results when comparing the coverage for sub-disciplines. Those different numbers of relevant papers might be why the coverage of databases per educational sector and per SLR highly shows high differences (Table 8).

For example, FIS has the highest coverage for early childhood education in SLR1 and with GS and DNB the only coverage for this field for SLR2. However, matters are different for SLR3, where the other sources are predominant. We know from the reviewers that the search focus differed in SLR3 for early childhood education, focusing as well on "teenager" and "young people," whereas in SLR1 and SLR2, the focus was "Kindergarten." Teacher and general education are internationally broadly investigated research fields. Here, WoS, ERIC, and ERC show higher coverage, whereas the databases do not well cover specific literature on vocational education. The 0% for FIS for teacher education might be explainable due to the database policy. FIS focuses on indexing literature in education that is not indexed elsewhere. As teacher education as well indexed in the international sources, FIS does not index the literature of this sector. FIS is accessible via the German Education Portal (Fachportal Pädagogik[2]), which allows meta-searches for the literature in ERIC, the Library of Congress, BASE, and other databases. The reviewers searched in the German Education Portal in the original SLR studies. In their data, they reported that they

---

[2] https://www.fachportal-paedagogik.de.

Table 7 Coverage *CO* of relevant publications (number and percentage) for each SLR, and average, median, and deviation over all three SLR for each database

Database	# SLR1	(%) SLR1	# SLR2	(%) SLR2	# SLR3	(%) SLR3	# all SLR	(%) all SLR	Average all SLR (%)	Median all SLR (%)	Average deviation all SLR (%)
FIS	45	22	30	24	10	9	85	19	18	22	6
ERIC	131	65	62	49	88	75	281	63	63	65	9
ERC	112	55	66	52	82	70	260	58	59	55	7
WoS	71	35	43	34	55	47	169	38	39	35	5
LTL	99	49	25	20	37	32	161	36	33	32	10
GS	190	94	116	92	114	97	420	94	95	94	2
DNB	26	13	39	31	24	21	89	20	21	21	6

Table 8 Coverage *CO* (%) of databases per sector and per SLR (1, 2, and 3)

	FIS			ERIC			ERC			WoS			LTL			GS			DNB		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
School	12	16	3	85	66	76	70	64	76	43	38	45	59	25	31	99	89	100	5	18	21
Early childhood	70	8	14	0	0	74	0	0	66	0	0	60	0	0	20	90	100	100	60	50	23
Teacher	0	0	0	67	79	92	93	100	84	73	67	50	93	46	53	100	100	100	7	17	11
Adult	37	61	13	37	22	50	24	22	50	15	22	13	24	0	0	88	100	88	20	52	25
Vocational	57	55	43	14	9	14	14	9	14	7	9	14	21	0	14	71	64	71	29	64	57

Table 9 Documents not found in WoS SSCI

Document type	Number
Article	199
Book (monography/edited)	32
Proceedings	26
Gray literature	16
Inexplicit document type	3

found 46% auf the relevant documents for SLR1 via the German Education Portal, respectively, 37% for SLR2 and 44% for SLR3 [39, 40]. One should as well notice that the search portal does not allow to search for all indexed literature in ERIC.

As the WoS is often applied for SLR in many disciplines due to its reputed international and interdisciplinary character, we had a closer look at its coverage. A total of 276 documents of SLR 1–3 were not found in the WoS SSCI, where we searched for the literature. A total of 199 of those 276 documents are marked as article in a supposed research journal (Table 9).

Forty-one of those 199 articles are published in journals indexed in WoS ESCI and one in WoS SCIE, respectively. Seven articles should have been found as they are published in a journal indexed in WoS SSCI. However, those journals are not fully indexed (yet). The volumes of our relevant articles in SLR 1–3 are missing. Overall, 148 articles marked as relevant in SLR 1–3 are published in journals that are not indexed in any of the main WoS indices (SSCI, SCIE, ESCI, and AHCI).

To conclude the coverage of databases, we can say that the overall performance shows a consistency of database relevance. However, on the level of single sub-reviews, we see varying numbers, which might be the result of different foci in the research questions and search strategies of the original studies. As mentioned above, GS is not efficient for systematic review searches [18] and was omitted by the reviewers after SLR1. The reviewers reported that they retrieved only a few documents via GS, i.e., the high coverage shown in our results does not reflect the efficiency of the database during the review search phase. If GS is not used, the importance of some databases for finding relevant documents is visible. WoS as an interdisciplinary database has a high coverage but lacks relevant peer-reviewed articles. As access to the WoS depends on different licenses, investigators should be aware of WoS databases (indices), they have access to and are able to apply in their searches.

4.2 Efficiency of database combinations

Regarding the second research question, the measurement of the coverage of relevant literature in combination of two databases (Fig. 1) again shows the predominance of GS. If ignoring GS, we see a combination of an international with a national database is efficient, such as ERIC or ERC with
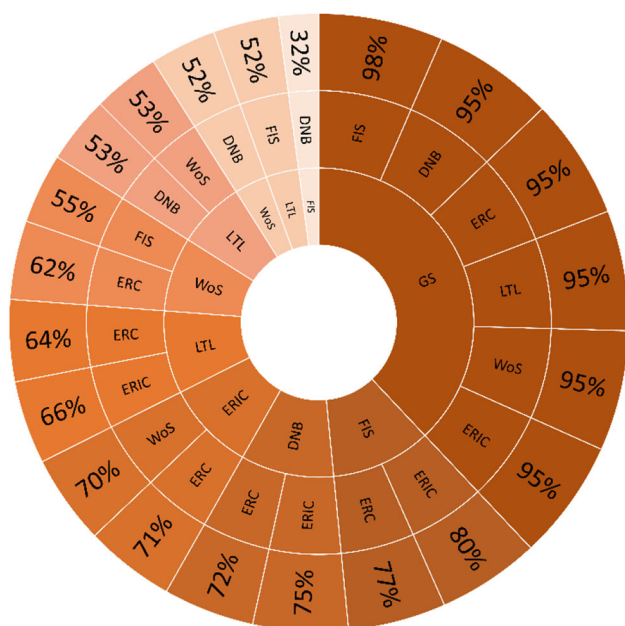
Fig. 1 Percentage of relevant documents found via database combination (*CB*). Numbers show average (%) of SLR 1–3

either FIS or DNB. The interdisciplinary WoS seems to play a less important role.

Figure 1 shows the average percentage of database combinations over all three reviews; however, the inner and outer circles do not indicate the proportion on the literature coverage of any of the two databases referring to. The average deviation is under 10% for all combinations, except for FIS and LTL with 13% and WoS and ERIC with 10%. Thus, the combination efficiency seems to be stable over SLR 1–3.

The most efficient combinations over more than two databases are shown in Table 10, which includes the five best combinations for each SLR, while leaving out databases that do not lead to a higher coverage. E.g., adding WoS to the first combination in SLR1 that covers 88% of relevant documents does not have any effect, and thus, WoS is left out in Table 10. In most cases for SLR1, WoS has no effect on the coverage, while it has for SLR2. LTL is less important for SLR2. In contrast, these results differ for SLR3, where we see WoS and LTL in the most efficient combinations. Thus, results cannot provide any general advice for applying WoS for discipline-specific review topics. Overall, Table 10 shows that for all SLR, a combination of FIS and either ERIC or ERC is fruitful, while other sources adding a few more relevant documents.

4.3 Similarity of database

The similarity of the databases provides insights into why the database combinations lead to different coverages of relevant documents. The numbers show a tendency toward coherence

Table 10 Multi-database combinations (*CB*), percentage shows number of relevant documents

SLR1	
FIS/ERIC/ERC/DNB/LTL	93%
FIS/ERIC/ERC/DNB	93%
FIS/ERIC/ERC/LTL	91%
FIS/ERIC/[ERC OR WoS]	90%
FIS/ERIC/LTL	90%
SLR2	
FIS/ERIC/ERC/WoS/DNB	90%
FIS/ERIC/ERC/DNB	88%
FIS/ERC/DNB/LTL/WoS	88%
FIS/ERC/DNB/[LTL OR WoS]	87%
FIS/ERC/DNB	84%
SLR3	
FIS/ERIC/ERC/WoS/LTL/DNB	98%
FIS/ERIC/ERC/WoS/[LTL OR DNB]	97%
FIS/ERIC/ERC/WoS	97%
FIS/ERIC/ERC/LTL/DNB	95%
ERIC/ERC/LTL/WoS/DNB	95%

Table 11 Average of similarity *SI* of databases (cosine coefficient) over SLR1–SLR3

	FIS	ERIC	ERC	WoS	LTL	GS	DNB
FIS		0.02	0.02	0.00	0.00	0.36	0.41
ERIC			0.85	0.65	0.66	0.81	0.25
ERC				0.77	0.63	0.79	0.24
WoS					0.51	0.65	0.22
LTL						0.58	0.07
GS							0.45

over all three SLR but are not that consistent as the average deviation indicates. Table 11 shows the average similarity over all three SLR.

ERIC and ERC are the most similar databases, besides GS and either both. The national databases FIS and DNB show little similarity to more internationally oriented and multidisciplinary databases. LTL shows an average deviation over 10%, i.e., for SLR1, it shows a similarity over 0.8 with ERIC and ERC. For SLR2 and SLR3, the similarity between LTL and either ERIC or ERC lies between 0.61 and 0.54. The more overlap between two databases can be seen, the smaller the benefit of their combination. To search the widest possible range of sources and therefore relevant publications, it is necessary to combine as heterogeneous databases as possible. E.g., in combination, ERIC and ERC cover 71% of relevant documents on average (Fig. 1), while ERIC itself already covers 63% (Table 7). Thus, adding ERC in a search strategy would not disclose many more relevant

Table 12 Number of unique documents per database

	SLR1	SLR1 (w/o GS)	SLR2	SLR2 (w/o GS)	SLR3	SLR3 (w/o GS)
FIS	5	28	5	13	1	4
ERIC	0	15	0	2	0	7
ERC	0	0	1	2	0	3
WoS	0	0	0	2	0	4
LLT	0	1	0	0	0	1
DNB	0	5	9	11	0	1
GS	8	–	1	–	1	–

Table 13 Document types in SLR 1–3

	SLR1	SLR2	SLR3
Journal article	168	91	109
Monography, incl. single chapters	23	29	8
Gray literature	11	6	0

documents. Similarly, WoS is quite similar to ERIC and ERC, and adding this source would not lead to significantly more relevant literature. Instead, more literature would be found while adding a national-oriented database such as FIS or DNB, of course always considering the concrete research topic of a SLR study.

4.4 Unique documents and document types

The relevance of databases becomes visible when looking at the outcomes for unique documents found in only one of the seven databases (Table 12).

Here, FIS shows its relevance, specifically when we do not consider GS. ERIC and DNB also have a few documents not covered by any other database. These results support the conclusions based on database combinations and their similarity. It indicates that more heterogenous databases contribute to a potential higher coverage of relevant literature.

We also took a closer look at the document types of the relevant literature chosen in SLR 1–3 (Table 13). Relevant literature for educational research findings are not exclusively published in journal articles. Gray literature as well as book chapters were added to the final dataset. Contributions to books and proceedings might not be indexed as a single document and thus not be retrievable via a database. This was the case in SLR3, where the document not found in any database is a book chapter. As such, a search in databases that include monographs is useful. In contrast, there seems to be a shift in the choice of literature for SLR3, which does only include eight monographs or chapters and no gray literature. Journal articles dominate specifically for teacher education. A reason here might be the high amount of literature for this sector and a more precise choice of relevant and high-quality literature.

For adult education, a sector where less literature was found, gray literature was considered. Moreover, for adult and vocational education, the reviewers considered relatively more relevant literature from books and book chapters. The example of adult education in SLR1 and SLR2 shows that the reviewers found a high number of relevant literature via different search strategies (Table 2).

Our analysis reveals specifically for SLR1 and SLR2 the importance of searching for gray literature. Many guidelines discuss the search for different document types and mention sources, such as OpenGrey, ProQuest for thesis, or catalogs of special libraries [2]. It is not trivial to search for gray literature in a systematic way. There is no gold standard for the methodological approach or suitable sources for educational research. Handbooks such as the Cochrane Handbook [4] suggest using GS as a source for gray literature. Unfortunately, a systematic search is nearly impossible in this source. Moreover, we lack information about what is included in this source. Resources include are not professionally indexed and might have missing or invalid metadata.

4.5 Limitations

The datasets chosen cover a broad and heterogenous research field and included three concrete research questions applied to five educational sectors. However, all SLR relate to just one research project and were done by the same group of researchers. This might bring a bias toward the choice of relevant literature, as SLR have different intentions and scopes and other research teams would have chosen different relevant literature. Moreover, because the SLR focused on educational structures in Germany, the search terms, sources, and finally studies were selected based on these criteria. A similar research question in the context of other geographical regions, of course, requires different databases—focusing also on databases of the geographical region—as well as different methods in the literature search and generate different results from the same database selection.

5 Practical implications

With regard to the coverage reflected in the databases, GS ranks first, but might not be useful for systematic review searches in practice as the precision rate is too low [18]. Discipline-specific databases like ERIC are more appropriate, whereas even more specific databases focusing on discipline-specific and national literature such as FIS and DNB add unique relevant resources. Reviewers should consider such databases regarding the topic and scope of their review. Meta-search portals facilitate searching over multiple databases. Coverage needs to be born in mind and possible selection policies of such services. Moreover, databases differ with respect to their frequency of updates. Our analysis shows that the reviewers of the original SLR found a lot of literature indexed in ERIC through the German Education Portal, but not all relevant documents in ERIC were retrieved via this source.

Regarding database combination, SLR should focus on databases that are not similar as this increases changes of finding more relevant literature, instead of duplicates indexed in multiple databases. If capacities are limited, one might consider choosing only one database out of two if coverage is similar.

In the original SLR, not all relevant documents were found via database searches. In total, 64 resources were found via other search strategies such as author or hand search, or reviewers added relevant literature they might have known already and did not name an explicit search strategy (Table 2). Nearly 4% of relevant literature is classified as gray literature. Yet, in our current analysis, we retrieved all but nine relevant documents in at least one database—however considering GS. When not considering GS, 20 documents are not found in any of the other databases, including a large proportion of gray literature and contributions in monographs.

Another reason for not finding documents in databases was the keywording. Some documents are low-keyworded, i.e., metadata is missing even in professional information databases. In addition, investigators seem to have used search terms too specific for the research topic or the educational sector. We took a closer look at the 64 documents originally not found in any database. About half of them were not found because their bibliographic metadata did not include the applied search terms. The search syntax seemed partly too complex, and documents were not retrieved via databases although we showed that these documents were indexed.

Some of the nine documents not found belong to German essay collections not indexed in any database. In the original reviews, they were identified via author searches, primarily done via search engines and institutional websites. Many government publications or final reports of institutional studies are not published in a traditional scientific format. Investigators, therefore, should consider extending database searches and ask if the literature relevant for a review might be published by ministries, institutions, or stakeholder groups that publish their reports on private websites or in repositories.

If investigators leave out gray literature and limit their relevant SLR documents to peer-reviewed articles—despite the criterion in many SLR guideline to reach a high recall of any relevant literature—they should be aware of the different WoS indices, their journal coverage, and their access to them, e.g., via their university library. Moreover, not all volumes of a journal might be indexed in those indices. Our data show that WoS, as all other databases, does not cover all research articles.

Despite the database coverage, the functions and usability of search systems is relevant and determines review search strategies. As such, investigators need to adapt search queries according to the databases they apply. Relevant criteria for appropriate review searches provide further evidence for reviewers for choosing the right sources [48].

6 Conclusion

We analyzed the coverage of seven databases based on 445 publications considered relevant in three larger SLR studies consisting of five sub-reviews from the educational field. We could retrieve most of the publications via databases, although a large amount of them was originally found via other search strategies such as hand and author search.

The overall database coverage regarding the first research question indicates a tendency toward a higher relevance of international discipline-specific databases for the educational field, compared to international multidisciplinary or national discipline-specific databases. This trend is measured over all analyzed SLR and supported by results on the similarity of databases. International discipline-specific databases are similar in coverage, while national discipline-specific databases hold unique publications that cannot be found elsewhere. However, results are more heterogeneous for sub-reviews focusing on more specialized educational sectors. Here, we see also a difference on the choice of relevant literature regarding document types.

GS outperformed all databases regarding the coverage, but due to poor precision, this database is considered inadequate for review purposes. The reviewers of the original studies applied GS only in SLR1, stating that it offers inadequate search options.

Analyses, regarding the second research question on database combinations, support the results on the coverage and indicate that a combination of discipline-specific international and national databases is most efficient. Two to three databases in combination add the most relevant documents,

while adding up to five different sources leads to a top coverage of 93% of relevant documents.

For practical implications, it should be noticed that a crucial methodological criterion for SLR is to get a recall rate as high as possible, i.e., to retrieve all relevant literature. As such, the more sources are applied, databases and other information sources, the higher the changes to reach this criterion. In research practices, SLR investigators need "to balance the thoroughness of the search with efficiency in use of time and funds" [6]. To better support them, further research is needed to compare more reviews in educational studies to confirm our results and give evidence-based information on databases that should be considered for reviews in this field, as well as measure possible biases resulting from these choices. In the future, semi-automated tools based on text mining and statistics might be useful to reach these goals.

At last, we want to emphasize that reviewers should carefully consider their choice of databases and give rationales on criteria for inclusion and exclusion of sources. We can only support the argument that researchers should not only name the databases included, but also details on the database access and concrete indices used, and as well give rationales on the inclusion and exclusion of databases [47].

References

1. Beelmann, A.: Möglichkeiten und Grenzen systematischer Evidenzkumulation durch Forschungssynthesen in der Bildungsforschung. Z. Erzieh. **17**, 55–78 (2014). https://doi.org/10.1007/s11618-014-0509-2

2. Booth, A., Sutton, A., Papaioannou, D.: Systematic Approaches to a Successful Literature Review, 2nd edn. SAGE, Los Angeles (2016)

3. Zawacki-Richter, O., Kerres, M., Bedenlier, S., et al. (eds.): Springer, Wiesbaden (2020)

4. Higgins, J., Thomas, J.: Cochrane Handbook for Systematic Reviews of Interventions: Version 6.2 (2021). https://training.cochrane.org/handbook/current

5. Tripney, J.: Introduction to systematic reviews: evidence 2016, Pretoria, South Africa, 20–22 September 2016 (2016). https://archive.org/details/tripney-2016-introduction-to-systematic-reviews

6. Kugley, S., Wade, A., Thomas, J., et al.: Searching for studies: A guide to information retrieval for Campbell systematic reviews. Campbell Syst. Rev. **13**, 1–73 (2017). https://doi.org/10.4073/cmg.2016.1

7. Sampson, M., de Bruijn, B., Urquhart, C., et al.: Complementary approaches to searching MEDLINE may be sufficient for updating systematic reviews. J. Clin. Epidemiol. **78**, 108–115 (2016). https://doi.org/10.1016/j.jclinepi.2016.03.004

8. Petticrew, M., Roberts, H.: Systematic Reviews in the Social Sciences: A Practical Guide. Blackwell, Malden (2006)

9. Newman, M., Gough, D.: Systematic reviews in educational research: Methodology, perspectives and application. In: Zawacki-Richter, O., Kerres, M., Bedenlier, S., et al. (eds.) Systematic Reviews in Educational Research: Methodology, Perspectives and Application, pp. 3–22. Springer Fachmedien Wiesbaden, Wiesbaden (2020)

10. Keller, C., Heck, T., Rittberger, M.: How many sources are needed? In: Aizawa, A., Mandl, T., Carevic, Z. et al. (eds). Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries. ACM, New York, NY, USA, pp. 1–11. https://doi.org/10.1145/3529372.3530933

11. Grant, M.J., Booth, A.: A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info. Libr. J. **26**, 91–108 (2009). https://doi.org/10.1111/j.1471-1842.2009.00848.x

12. Gough, D., Thomas, J., Oliver, S.: Clarifying differences between review designs and methods. Syst. Rev. **1**, 28 (2012). https://doi.org/10.1186/2046-4053-1-28

13. Gough, D., Oliver, S., Thomas, J.: Introducing systematic reviews. In: Gough, D., Oliver, S., Thomas, J. (eds). An Introduction to Systematic Reviews, 2nd edition. SAGE, Los Angeles, pp. 1–17

14. Centre for Reviews and Dissemination: CRD's guidance for undertaking reviews in healthcare, 3rd edn. Systematic reviews. York Publ, Services, York (2009)

15. Cooper, C., Booth, A., Varley-Campbell, J., et al.: Defining the process to literature searching in systematic reviews: A literature review of guidance and supporting studies. BMC Med. Res. Methodol. **18**, 85 (2018). https://doi.org/10.1186/s12874-018-0545-3

16. Fink, A.: Conducting Research Literature Reviews: From the Internet to Paper. Sage publications, Los Angeles, CA (2019)

17. Patole, S.: Principles and Practice of Systematic Reviews and Meta-Analysis. Springer. (2021). https://doi.org/10.1007/978-3-030-71921-0

18. Boeker, M., Vach, W., Motschall, E.: Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. BMC Med. Res. Methodol. **13**, 131 (2013). https://doi.org/10.1186/1471-2288-13-131

19. Higgins, J., Green, S.: Cochrane Handbook for Systematic Reviews of Interventions (2008). https://training.cochrane.org/handbook/current

20. Wade, A., Turner, H., Rothstein, H., et al.: Information retrieval and the role of the information specialist in producing high-quality systematic reviews in the social, behavioural and education sciences. Evid. Policy J. Res. Debate Pract. **2**, 89–108 (2006). https://doi.org/10.1332/174426406775249705

21. Schellinger, J., Sewell, K., Bloss, J.E., et al.: The effect of librarian involvement on the quality of systematic reviews in dental medicine. PLoS ONE **16**, e0256833 (2021). https://doi.org/10.1371/journal.pone.0256833

22. Aamodt, M., Huurdeman, H., Strømme, H.: Librarian co-authored systematic reviews are associated with lower risk of bias compared to systematic reviews with acknowledgement of librarians or no participation by librarians. EBLIP **14**, 103–127 (2019). https://doi.org/10.18438/eblip29601

23. Meert, D., Torabi, N., Costella, J.: Impact of librarians on reporting of the literature searching component of pediatric systematic reviews. J. Med. Lib. Assoc. JMLA **104**, 267 (2016)

24. Rethlefsen, M.L., Farrell, A.M., Trzasko, L.C.O., et al.: Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J. Clin. Epidemiol. **68**, 617–626 (2015). https://doi.org/10.1016/j.jclinepi.2014.11.025

25. Schardt, C., Adams, M.B., Owens, T., et al.: Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med. Inform. Decis. Mak. **7**, 16 (2007). https://doi.org/10.1186/1472-6947-7-16

26. McGowan, J., Sampson, M., Salzwedel, D.M. et al.: PRESS - Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration, Ottawa (2016)

27. Cleverdon, C.W.: The critical appraisal of information retrieval systems. In: The International Congress of the International Federation for Documentation, Moscow (1968)

28. Rittberger, M., Rittberger, W.: Measuring quality in the production of databases. J. Inf. Sci. **23**, 25–37 (1997)

29. Bramer, W.M., Giustini, D., Kramer, B.M.R.: Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Syst. Rev. **5**, 39 (2016). https://doi.org/10.1186/s13643-016-0215-7

30. Bramer, W.M., Rethlefsen, M.L., Kleijnen, J., et al.: Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst. Rev. **6**, 245 (2017). https://doi.org/10.1186/s13643-017-0644-y

31. Lam, M.T., de Longhi, C., Turnbull, J., et al.: Has Embase replaced MEDLINE since coverage expansion? J. Med. Lib. Assoc. JMLA **106**, 227 (2018). https://doi.org/10.5195/jmla.2018.281

32. Hartling, L., Featherstone, R., Nuspl, M., et al.: The contribution of databases to the results of systematic reviews: a cross-sectional study. BMC Med. Res. Methodol. **16**, 127 (2016). https://doi.org/10.1186/s12874-016-0232-1

33. Brunton, G., Stansfield, C., Caird, J., et al.: Finding relevant studies. In: Gough, D., Oliver, S., Thomas, J. (eds.) An Introduction to Systematic Reviews, 2nd edn., pp. 93–122. SAGE, Los Angeles (2017)

34. Heinze, T., Tunger, D., Fuchs, J.E. et al.: Fachliche Forschungs- und Lehrprofile staatlicher Universitäten in Deutschland. Eine Kartierung ausgewählter Fächer (2019). https://www.researchgate.net/publication/337324104_Heinze_T_Tunger_D_Fuchs_JE_Jappe_A_Eberhardt_P_2019_Fachliche_Forschungs-_und_Lehrprofile_staatlicher_Universitaten_in_Deutschland_Eine_Kartierung_ausgewahlter_Facher_Wuppertal_BUW. https://doi.org/10.25926/9242-ws58

35. Archambault, É., Vignola-Gagné, É., Côté, G., et al.: Benchmarking scientific output in the social sciences and humanities: the limits of existing databases. Scientometrics **68**, 329–342 (2006). https://doi.org/10.1007/s11192-006-0115-z

36. Martín-Martín, A., Orduna-Malea, E., Thelwall, M., et al.: Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. J. Informet. **12**, 1160–1177 (2018). https://doi.org/10.1016/j.joi.2018.09.002

37. Kousha, K., Thelwall, M.: Google Scholar citations and Google Web/URL citations: a multi-discipline exploratory analysis. J. Am. Soc. Inform. Sci. Technol. **58**, 1055–1065 (2007). https://doi.org/10.1002/asi.20584

38. Khabsa, M., Giles, C.L.: The number of scholarly documents on the public web. PLoS ONE **9**, e93949 (2014). https://doi.org/10.1371/journal.pone.0093949

39. Rittberger, M., Keller, C.: Literaturrecherche für Dossier Digi EBF 1 (2020). https://doi.org/10.7477/414:1:0

40. Rittberger, M., Keller, C.: Literaturrecherche für Dossier Digi EBF 2 (2021). https://doi.org/10.7477/414:2:0

41. Wilmers, A., Anda, C., Keller, C. et al.: Bildung im digitalen Wandel: Die Bedeutung für das pädagogische Personal und für die Aus-und Fortbildung. Waxmann (2020). https://doi.org/10.31244/9783830991991

42. Wilmers, A., Achenbach, M., Keller, C.: Bildung im digitalen Wandel: Organisationsentwicklung in Bildungseinrichtungen. Waxmann (2021). https://doi.org/10.31244/9783830994558

43. Gough, D., Thomas, J.: Systematic reviews of research in education: Aims, myths and multiple methods. Rev. Educ. **4**, 84–102 (2016). https://doi.org/10.1002/rev3.3068

44. Dees, W.: Bildungsforschung, Humboldt-Universität zu Berlin, Philosophische Fakultät III. https://doi.org/10.18452/17101 (2015)

45. Keller, C., Heck, T., Rittberger, M.: Systematic Reviews in Education (2022). https://doi.org/10.17605/OSF.IO/NJTWC

46. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, New York, USA (1999)

47. Wanyama, S.B., McQuaid, R.W., Kittler, M.: Where you search determines what you find: the effects of bibliographic databases on systematic reviews. Int. J. Soc. Res. Methodol. (2021). https://doi.org/10.1080/13645579.2021.1892378

48. Gusenbauer, M., Haddaway, N.R.: Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res. Synth. Methods **11**, 181–217 (2020). https://doi.org/10.1002/jrsm.1378