Keller, Carolin; Heck, Tamara; Rittberger, Marc

How many sources are neded? The effects of bibliographic databases on systematic review outcomes. [Preprint]

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
Aizawa, Akiko [Hrsg.]; Mandl, Thomas [Hrsg.]; Carevic, Zeljko [Hrsg.]; Hinze, Annika [Hrsg.]; Mayr, Philipp [Hrsg.]; Schaer, Philipp [Hrsg.]: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2022 (JCDL '22), hybrid conference, Cologne, Germany and online, June 20 - 24, 2022. New York : Association for Computing Machinery 2022, 10.1145/3529372.3530933

Preprint of Carolin Keller, Tamara Heck, and Marc Rittberger. 2022. How many sources are needed? The effects of bibliographic databases on systematic review outcomes. In *The ACM/IEEE Joint Conference on Digital Libraries in 2022 (JDCL'22). June 20-24, 2022, Cologne, Germany.* https://doi.org/10.1145/1234567890

How many sources are needed?

The effects of bibliographic databases on systematic review outcomes

Carolin Keller
Information Center for Education
Leibniz Institute for Research and Information in Education
Frankfurt am Main Germany
keller@dipf.de

Tamara Heck
Information Center for Education
Leibniz Institute for Research and Information in Education
Frankfurt am Main Germany
heck@dipf.de

Marc Rittberger
Information Center for Education
Leibniz Institute for Research and Information in Education
Frankfurt am Main Germany;
Darmstadt University of Applied Sciences
rittberger@dipf.de

ABSTRACT

Systematic reviews are an established method to synthesize the current state of research for a specific question to make evidence-based decisions in research, politics, and practice. A key activity of a review approach is a systematic and comprehensive search strategy to find all potentially relevant literature. Although guidelines and handbooks address relevant methodological aspects and recommend strategies, the right choice of databases and information sources is unclear. Specifically in educational research, an interdisciplinary field, with no core database at hand and multiple potentially relevant sources available, investigators lack guidance for choosing the most appropriate ones. The presented study investigates the coverage in terms of scope, similarity and combination efficiency of seven multidisciplinary, discipline-specific and nationally focused databases. The evaluation is based on relevant assessed literature of two extensive recently published reviews in German educational research that serve as gold standard to evaluate the databases. Results indicate distinct variations in the databases, while also detecting databases with equal coverage. The paper contributes to guidance in choosing databases for educational review studies, while stressing that this process depends on a review's topical and geographical focus. Moreover, general implications resulting from the study refer to the relevance of database choice for review outcomes, the careful consideration of diverse search strategies beyond database search and a rigorous documentation of database inclusion and exclusion criteria.

CCS CONCEPTS

Information Systems - Information retrieval diversity, Information Systems - similarity measures, Information Systems - Combination, fusion and federated search

KEYWORDS

Systematic review, research database, coverage, similarity, educational research

1 Introduction

Systematic reviews 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 systematic reviews 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 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 are researched in several disciplines like 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.

In the following, we present a study investigating the effects of database choices on finding relevant literature for systematic reviews in educational research. We thus analyze relevant literature from published systematic reviews in this field. The study concludes on the literature coverage of those databases and their combinations to inform recommendations for database choices in systematic reviews for educational research. We analyzed literature from ten reviews with regard to its findability in main database sources. Our research questions are:

- Which databases index relevant literature for reviews?
- Which combination of databases most efficiently covers all relevant literature?

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

2 State of Art for Systematic Reviews

Reviews 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 [10], and with it a diverse terminology, like scoping review or critical review. As such, the different types are based on different methodological approaches relating to the needs and objectives of the researchers and those are continuously being developed and improved [11]. 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" [12, p. 4].

2.1 Search Strategies in Review Guidelines

Guidelines and handbooks list detailed and structured steps for systematic reviews [13–16, 8]. 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, 12, 17]. Reviews can only provide reliable evidence based on a completely searched data set. Consequently, investigators, who are accountable of the search for relevant literature, bear 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 [18, 6], which "recognize the importance and value of consulting with an information specialist during the (un)systematic information retrieval stage of the review process" [19, 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 this, they examined how often librarians were mentioned or listed as co-authors in review papers. According to the studies, librarians have 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 [20–23].

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 the first results. Additionally, instructions for an extensive documentation of all steps guarantee the replication of the method in the conducted review [2, 6, 14]. Most of the guidelines, however, focus on reviews in the fields of health and medicine, like the PICO framework [24], 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 [8, 6] . 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: Should only peer-reviewed journal articles or as well reports, i.e. grey literature, be included?

For assessing the quality of the literature search itself, a peer review for electronic search strategies (PRESS) exists [25]. PRESS defines quality criteria for database searches, e.g. the services shall allow Boolean operators to formulate proper queries, investigators shall consider the selection of search words and different functions the databases offer. Besides, however, this strategy does not give any explicit criteria for database selection, but only provides a generic list of available databases.

2.2 Criteria for Database Selection

A crucial element for characterizing bibliographic databases is their coverage [26]. Rittberger and Rittberger specifically name scope and coverage as important subject-related criteria for the quality of databases [27]. For a complete search coverage, the

investigator of a review 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]. This explicit recommendation on databases derives from the fact that the health and medical science sector has core databases, which seem to cover most relevant research outputs in this field. With a proper search strategy applying Boolean search and ranking, researchers claim that searching Medline only might be sufficient [7]. Other studies conclude in a similar way by saying that the "majority of relevant studies can be found in a limited number of databases" [28]. The authors suggest applying Medline in combination with one other database like Embase or Biosis, choice depending on the topic. This limited number of sources would be sufficient for most cases and omitting other database sources has less or no significant effects on the outcomes of a review. Moreover, the geographical coverage of US-based Medline and Elsevier's Embase might differ, but Embase has been aiming to include all Medline references and using Embase only might suffice for medical reviews [29]. In contrast, Bramer et al. [30], who investigated optimal database combinations for review searches, suggest applying at least Embase, Medline, Web of Science, and Google Scholar for an efficient coverage. The trend to use more than one database in medical and health sciences is measurable [29]. Despite slightly different study outcomes, the discipline seems to have a manageable and known set of databases that cover relevant literature.

Similarly for the social and educational sciences, various guidelines and textbooks provide lists of selected sources, often named are ERIC, the Web of Science, and FIS Bildung [6, p. 47 ff, 2, p. 111 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 [31, p. 107 ff]. In contrast to the quite large number of studies investigating databases for reviews in the medical and health sciences, we currently lack 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 % [32]. Other studies found similar results and conclude that the Web of Science over represents English language publications for those disciplines [33]. 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 [34]. Google Scholar performed best, but this database 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 literature for all disciplines, investigators of educational research reviews 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, a crucial fact educational researchers need to consider, specifically when they investigate questions of national importance. 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 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 like 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 [17] and it is unclear which and how many publications Google Scholar explicitly includes [35, 36]. Boeker et al. applied "realistic search expressions" from published reviews to show the effects of Google Scholar's limited search options [17]. The authors conclude that the database "does not provide necessary elements for systematic scientific literature retrieval" [17], 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 analyze the coverage and overlapping of databases, we took a closer look at search outcomes and relevant literature from two review studies with ten sub-reviews from an extensive educational research project in the field of digital education. In the following, we describe those reviews and the conducted searches, before we introduce our analytic approach.

3.1 Datasets

The chosen datasets of the ten 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 it is meant to conduct critical reviews with narrative overviews that summarize essential findings for each specific research question within the project, the investigators did systematic literature searches and published their data at a research data center [37, 38]. The results are published in two proceedings [39, 40].

We chose those datasets because they are a good example of reviews in educational research, as they focus on two 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 multi-component reviews [41]. The first question asks about the role of pedagogical staff in implementing digital devices (review 1). The second question asks about organizational development in educational institutions (review 2). For each question, the sectors are early childhood, general education, vocational education, adult and teacher education. Thus, for each research question, the researchers complied five reviews in two different search phases. Literature inclusion criteria were German and English resources and a publication date later than 2016. In contrast to 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 [42, p. 114]. For further details of the search, we refer to the original data documentation [37, 38] and proceedings [39, 40]. We will mention *review 1* (*R1*) referring to the first review question and the five sub-reviews for the five educational sectors, and *review 2* (*R2*) referring to the second review question and the five sub-reviews, respectively.

In the search for review 1, the investigators chose two main discipline-specific sources for educational research plus the German national library and Google Scholar as main web source. Google Scholar is said to outperform traditional databases [34]. Based on the experiences made during this search and the screening of the retrieved literature, the investigators expanded and adapted the choice of databases for the search in review 2 to cover more discipline-related and multidisciplinary research. The Web of Science and Education Research Complete were included, as was LearnTechLib, which indexes research reports relevant for the investigated questions. Moreover, the researchers excluded Google Scholar due to the low precision rate of searches and limited search functionalities [17]. Besides the database searches, the investigators applied advanced search strategies like hand searching (search on websites from institutions and associations) and citation searching based on relevant authors. These hand searches revealed 62 publications. For the database analysis presented in this paper, we investigated all sources either applied in review study 1 and/or review study 2 (table 1).

Table 1. 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 review study 1 + applied in review study 2

3.2 Analytical Approach

For the following study, we searched 328 publications from *R1* and *R2*. Those publications were considered relevant by the expert researchers involved in the reviews. They mark the final datasets synthesized. They are our gold standard to measure the coverage of databases on relevant literature for review in educational research, adapting the measurement of database coverage in the life sciences (cp. [30]). We searched for each publication in the seven databases the investigators applied in any of the original studies. We conducted our search December of 2021. We used title and author details and DOIs. It has to be noted that the original review datasets have information on the source of a publication as well. However, the removal of duplicates biases this output. The dataset of our search is published at OSF [43].

The *coverage* of a database indicates which relevant literature known to a user a database includes [44, p. 83]. We measured

the coverage with:

$$\text{coverage} \quad \frac{|R|}{|U|} \tag{1}$$

where |R| is the number of retrieved relevant documents, and |U| the number of relevant documents known, i.e. our gold standard.

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

$$\text{similarity} \quad \frac{|C|}{\sqrt{|D_a| * |D_b|}} \tag{2}$$

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

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

$$\text{combination} \quad \frac{|D_a| + |D_b| - |D_{ab}|}{|U|} \tag{3}$$

where $|D_{ab}|$ is the number of relevant documents retrieved in databases D_a and D_b.

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 for investigators of systematic reviews.

4.1 Coverage of Databases

Tables 2 and 3 show the number of the retrieved publications, i.e. the database coverage referring to our first research question. Overall, we found 320 relevant documents in the seven databases for *R1* and *R2*, as well, 54 of the 62 documents originally added by hand and author search. However, eight documents of those, six from *R1* and two from *R2* seem not to be indexed in any of the seven databases. As expected, the GS coverage is highest for both *R1* and *R2*, followed by ERIC and ERC, both discipline-specific databases focusing on educational research (table 1). As other studies show [45], 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, EBSCO) [45]. In our analysis, none of the relevant documents was retrieved in all seven databases for *R1* and *R2*. For *R1*, less than 1 % was retrieved in six, 25.74% in five, and about 20 % in four, three or two databases, respectively. For *R2*, no document was found in six databases, and between 10% and 17 % in either five, four, three or two databases.

Moreover, outcomes vary highly for the two review datasets we investigated. For example, FIS Bildung has the highest coverage for early childhood education (R1) and with DNB the only coverage for this field for R2. Similar results apply for vocational education. In contrast, the best databases for literature on teacher education seem to be ERC, ERIC and WoS. These results are striking, as we are not aware of database policies to focus on specific educational sectors. The high coverage of the DNB for adult and vocational education (table

Table 2. Numbers of retrieved relevant literature and coverage per database and per educational sector for *R1*.

Educational sectors	FIS		ERIC		ERC		WoS		LTL		GS		DNB		# relevant documents
School education	15	12%	104	85%	86	70%	52	43%	72	59%	121	99%	6	5%	122
Early childhood ed.	7	70%	0	0%	0	0%	0	0%	0	0%	9	90%	6	60%	10
Teacher education	0	0%	10	67%	14	93%	11	73%	14	93%	15	100%	1	7%	15
Adult education	15	37%	15	37%	10	24%	5	12%	10	24%	36	88%	8	20%	41
Vocational education	8	57%	2	14%	2	14%	1	7%	2	14%	10	71%	4	29%	14
TOTAL	45		131		112		69		98		191		25		202
COVERAGE TOTAL	22%		65%		55%		34%		49%		95%		12%		100%

Table 3: Numbers of retrieved relevant literature and coverage per database and per educational sector for *R2*.

Educational sectors	FIS		ERIC		ERC		WoS		LTL		GS		DNB		# relevant documents
School education	9	16%	37	66%	36	64%	20	36%	14	25%	50	89%	10	18%	56
Early childhood ed.	1	8%	0	0%	0	0%	0	0%	0	0%	12	100%	6	50%	12
Teacher education	0	0%	19	79%	24	100%	16	67%	11	46%	24	100%	4	17%	24
Adult education	14	61%	5	22%	5	22%	5	22%	0	0%	23	100%	12	52%	23
Vocational education	6	55%	1	9%	1	9%	1	9%	0	0%	7	64%	7	64%	11
TOTAL	30		62		66		42		25		116		39		126
COVERAGE TOTAL	24%		49%		52%		33%		20%		92%		31%		100%

2) might be because R2 focused on literature about organizational development in education especially for German institutions. DNB includes many monographs on this topic.

Teacher and general education are internationally broadly investigated research fields. Here, WoS, which is biased towards English language literature, shows higher coverage, whereas the database does not cover well more specific literature on early childhood or vocational education.

Furthermore, the discipline-specific databases cover more relevant literature than the WoS. FIS, which has a focus on German language educational literature, might have less coverage due to the fact that it concentrates on indexing literature, which has not yet been indexed in other databases. Moreover, FIS is part of a meta-search portal, Fachportal Pädagogik (German Education Portal[1]), which has interfaces to ERIC, the Library of Congress, BASE and other databases, and allows a meta-search for educational literature. The investigators included this search portal in the original review studies, where it covered 88% of the total documents for *R1*

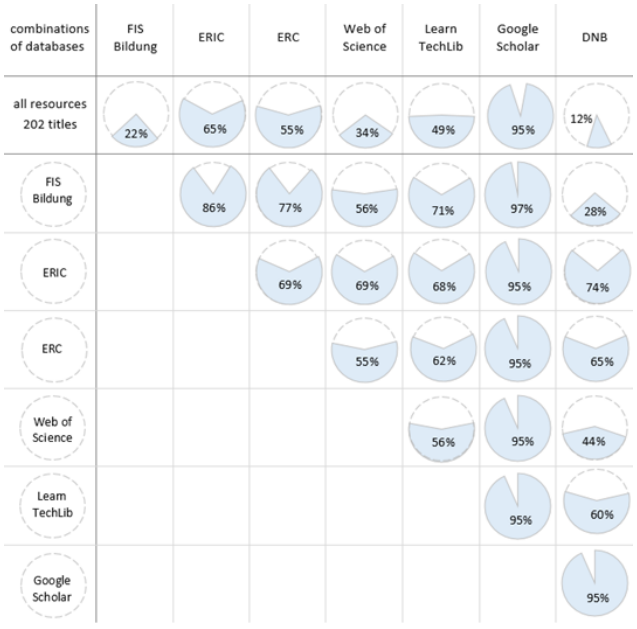

Figure 1. Coverage of database combination for *R1*.

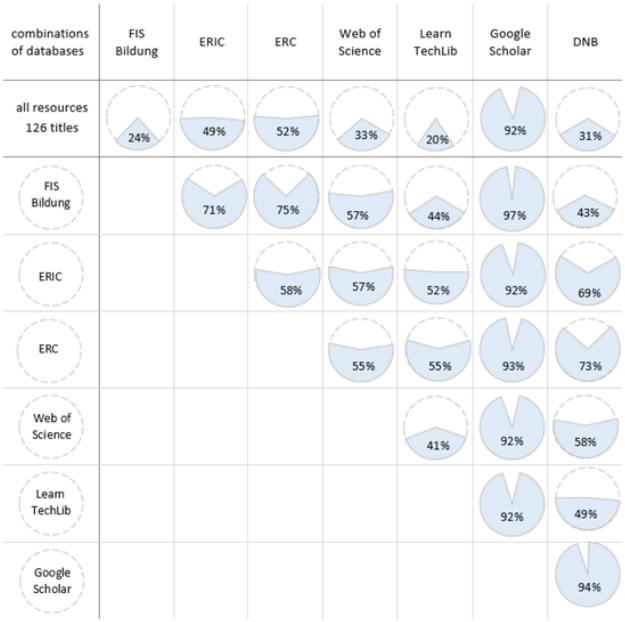

Figure 2. Coverage of database combination for *R2*.

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

and 74% for *R2* [37, 38]. In the analysis presented in this paper, we counted only literature indexed by FIS and its partners.
The policy of FIS gets clearer when looking at the outcomes for unique documents found in only one database (table 6). Here, FIS shows its relevance, specifically when we do not consider GS. DNB also has a few documents not covered by any other database.
As mentioned above, GS is not efficient for systematic review searches [17] and was omitted by the review investigators in the second search phase. In the first search, only a few documents were actually retrieved 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 individual documents is visible.

4.2 Efficiency of Database Combinations

Measurement of the coverage of relevant literature in combination of two databases (figures 1 and 2), referring to the second research question, shows that some databases seem to cover the same documents, like ERIC and ERC. In combination, the databases cover 69 % of the 202 relevant documents of *R1* only, while ERIC itself already covers 65 %. Thus, adding ERC in a search strategy would not disclose many more relevant documents. Similarly, LTL coverage seemed to be very similar to that of ERIC and ERC, more so in *R1* than in *R2*. Combining WoS with ERC would not make sense either, as in *R1* the database did not add any new relevant document.
These results reflect similarity values for the databases (tables 4 and 5). The national databases FIS and DNB, which as well contribute individual publications, show little similarity to more internationally oriented and multidisciplinary databases. Similarity of coverage is highest for ERIC and ERC in both reviews. The more overlap between two databases can be seen, the smaller the benefit of the combination. In order to search the widest possible range of sources and therefore relevant publications, it is necessary to combine as heterogeneous databases as possible.
We analyzed the most fruitful database combinations (figures 3 and 4), i.e. those that lead to the highest relative coverage. Other possible combinations do not lead to a higher coverage or adding databases to any of the combinations shown does not have any effect on the coverage. ERIC, FIS, ERC and DNB lead to 93 % coverage for *R1*. Here, WoS and LTL do not have any further effect. WoS as well does not show any effect for the second best combination for *R1* (figure 3). Interestingly, WoS has an impact for *R2* for both most efficient database combinations (figure 4). Thus, results cannot provide any practical advice for applying WoS for discipline-specific review topics. For both reviews, the more specialized databases like FIS and DNB lead to a slightly higher coverage due to their individual documents coverages.

5 Practical Implications

With regard to the coverage of 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 [17]. Discipline-specific databases like ERIC are more appropriate, whereas even more specific databases focusing on discipline-specific and national literature like FIS and DNB add individual relevant resources. Investigators should consider such databases with regard to the review questions' scope. Meta-search databases are efficient to search in multiple sources. However, investigators should be aware of the coverage and possible selection policy. Moreover, databases differ with respect to their frequency of updates. While conducting the

Table 4: Similarity of databases (total numbers and cosine coefficient) based on common relevant literature for *R1*.

	FIS	ERIC	ERC	WoS	LTL	GS	DNB
	45	131	112	69	99	191	25
FIS		2 0.03	1 0.01	0 0.00	1 0.01	40 0.43	14 0.42
ERIC			103 0.85	61 0.64	92 0.81	131 0.83	6 0.10
ERC				69 0.78	86 0.82	112 0.77	6 0.11
WoS					56 0.68	69 0.60	6 0.14
LTL						99 0.72	2 0.04
GS							25 0.36

Table 5: Similarity of databases (total numbers and cosine coefficient) based on common relevant literature for *R2*.

	FIS	ERIC	ERC	WoS	LTL	GS	DNB
	30	62	66	42	25	116	39
FIS		2 0.05	1 0.02	0 0.00	0 0.00	24 0.41	15 0.44
ERIC			55 0.86	39 0.76	22 0.56	62 0.73	14 0.28
ERC				39 0.74	22 0.54	66 0.75	13 0.26
WoS					15 0.46	42 0.60	8 0.20
LTL						25 0.46	2 0.06
GS							37 0.55

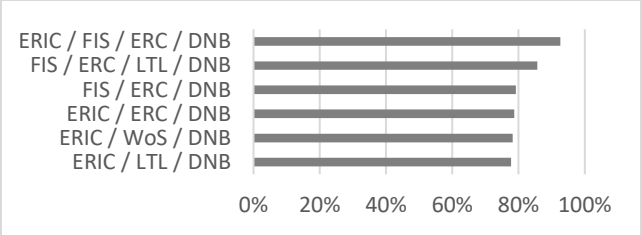

Figure 3. Most efficient database combinations for *R1*.

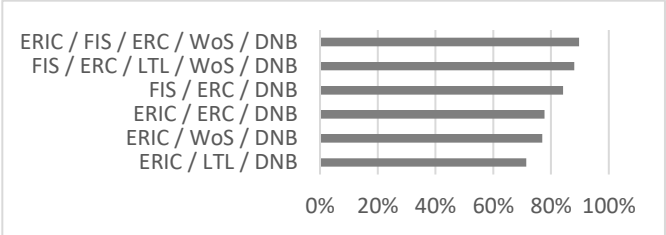

Figure 4. Most efficient database combinations for *R2*.

reviews, investigators found a lot of literature indexed in ERIC through the German Education Portal, but still not all relevant documents were retrieved.

Regarding database combination, investigators of systematic reviews should more likely choose databases that are not similar in their collection profile so they can find diverse relevant literature. They might consider choosing only one database out of two, if the sources are very similar and reveal a high number of documents they have in common.

Table 6: Number of relevant publications found in only one database of the seven databases, with and without Google Scholar.

	R1	***R1*** (without GS)	***R2***	***R2*** (without GS)
FIS	5	28	5	13
ERIC	0	15	0	2
ERC	0	0	1	2
WoS	0	0	0	2
LLT	0	1	0	0
DNB	0	5	9	11
GS	8	-	1	-

In the review search process, investigators did not find all relevant documents through database searches. In total, they identified 62 resources, which they later decided to be relevant for review synthesis, through advanced search tactics like author or citation search. A significant proportion of the titles are classified as grey literature. Yet, in our current analysis we retrieved all but eight relevant documents. One reason was that the investigators did not include DNB in *R2*, which has relevant additional documents, specifically for early childhood education literature (table 5). Therefore, review investigators should be careful to tailor databases choice according to the research topic and educational sector.

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 documents originally not found in any database. 51 % of them were not found because their bibliographic metadata did not include the applied search terms. The search syntax seemed partly too complex and investigators did not retrieve documents via databases although we proved that these documents were indexed.

Some of the eight documents not found belong to German essay collections not indexed in any database. In the original reviews, investigators identified them via author searches, which they primarily did 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 literature relevant for a review might be published by ministries, institutions or stakeholder groups that publish their reports on private websites or in repositories.

In summary, the results show that databases vary a lot and thus influence the outcomes of systematic reviews. Investigators should carefully consider database choice. Moreover, 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 [45]. Our results show that it is inappropriate to establish a concrete list of relevant databases for educational research as the efficiency of databases based on their coverage may highly depend on the scope of a review, like geographical and topical context. Thus, the results cannot be generalized to all future reviews in educational research.

Because the reviews were 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. Nevertheless, the results give insights into the relevance of discipline-specific and national databases for educational research.

6 Conclusion

We analyzed the coverage of seven databases based on 328 publications considered relevant in two larger review studies consisting of ten sub-reviews from the educational field. We

could retrieve most of the publications, though not all of those originally found via hand search. The database coverage showed high variations and clearly indicates that one source on its own does not cover a sufficient amount of relevant literature. Some databases are very similar in coverage, while national and discipline-specific databases hold publications that cannot be found elsewhere. Google Scholar outperformed all databases regarding recall. However, due to poor precision this database is considered inadequate for review purposes. Having said this, we noticed that without Google Scholar, the most efficient database combination covered 93 % of all relevant documents. Here, it has to be noted that a high number of documents was not retrieved via databases in the original review study, but found via hand-search. Further research is needed to compare more reviews in educational studies to recommend relevant databases or at least give more evidence-based information on databases that should definitely be considered for reviews in this field. Investigators should carefully consider their choice of databases and give rationales on criteria for inclusion and exclusion of sources.

REFERENCES

[1] Andreas Beelmann. 2014. Möglichkeiten und Grenzen systematischer Evidenzkumulation durch Forschungssynthesen in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft* 17, S4, 55–78. DOI: https://doi.org/10.1007/s11618-014-0509-2.

[2] Andrew Booth, Anthea Sutton, and Diana Papaioannou. 2016. *Systematic approaches to a successful literature review* (Second edition). SAGE, Los Angeles, London, New Delhi.

[3] Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond, and Katja Buntins, Eds. 2020. *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*. Springer Fachmedien Wiesbaden, Wiesbaden.

[4] Julian Higgins and James Thomas. 2021. *Cochrane Handbook for Systematic Reviews of Interventions*. *Version 6.2* (2021). Retrieved from https://training.cochrane.org/handbook/current.

[5] Janice Tripney. 2016. *Introduction to systematic reviews: evidence 2016, Pretoria, South Africa, 20-22 September 2016.* (2016). Retrieved from https://archive.org/details/tripney-2016-introduction-to-systematic-reviews.

[6] Shannon Kugley, Anne Wade, James Thomas, Quenby Mahood, Anne-Marie K. Jørgensen, Karianne Hammerstrøm, and Nila Sathe. 2017. Searching for studies. A guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews* 13, 1, 1–73. DOI: https://doi.org/10.4073/cmg.2016.1.

[7] Margaret Sampson, Berry de Bruijn, Christine Urquhart, and Kaveh Shojania. 2016. Complementary approaches to searching MEDLINE may be sufficient for updating systematic reviews. *Journal of clinical epidemiology* 78, 108–115. DOI: https://doi.org/10.1016/j.jclinepi.2016.03.004.

[8] Mark Petticrew and Helen Roberts. 2006. *Systematic reviews in the social sciences*. *A practical guide*. Blackwell, Malden.

[9] Mark Newman and David Gough. 2020. Systematic reviews in educational research: Methodology, perspectives and application. In *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*, Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond and Katja Buntins, Eds. Springer Fachmedien Wiesbaden, Wiesbaden, 3–22.

[10] Maria J. Grant and Andrew Booth. 2009. A typology of reviews. an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal* 26, 91–108. DOI: https://doi.org/10.1111/j.1471-1842.2009.00848.x.

[11] David Gough, James Thomas, and Sandy Oliver. 2012. Clarifying differences between review designs and methods. *Systematic Reviews* 1, 28. DOI: https://doi.org/10.1186/2046-4053-1-28.

[12] David Gough, Sandy Oliver, and James Thomas. 2017. Introducing systematic reviews. In *An introduction to systematic reviews*, David Gough, Sandy Oliver and James Thomas, Eds. SAGE, Los Angeles, London, New Delhi, Singapore, Washington, DC, Melbourne, 1–17.

[13] Centre for Reviews and Dissemination. 2009. *CRD's guidance for undertaking reviews in healthcare* (3. ed.). Systematic reviews. York Publ. Services, York.

[14] Chris Cooper, Andrew Booth, Jo Varley-Campbell, Nicky Britten, and Ruth Garside. 2018. Defining the process to literature searching in systematic reviews. A literature review of guidance and supporting studies. *BMC Medical Research Methodology* 18, 1, 85. DOI: https://doi.org/10.1186/s12874-018-0545-3.

[15] Arlene Fink. 2019. *Conducting research literature reviews: From the internet to paper*. Sage publications.

[16] Sanjay Patole. 2021. *Principles and Practice of Systematic Reviews and Meta-Analysis*. Springer International Publishing, Cham. DOI: https://doi.org/10.1007/978-3-030-71921-0.

[17] Martin Boeker, Werner Vach, and Edith Motschall. 2013. Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. *BMC Medical Research Methodology* 13, 1, 131. DOI: https://doi.org/10.1186/1471-2288-13-131.

[18] Julian Higgins and Sally Green. 2008. *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley Cochrane Ser. John Wiley & Sons, Newark https://training.cochrane.org/handbook/current. Retrieved from.

[19] Anne Wade, Herbert Turner, Hannah Rothstein, and Julia Lavenberg. 2006. Information retrieval and the role of the information specialist in producing high-quality systematic reviews in the social, behavioural and education sciences. *Evidence & Policy: A Journal of Research, Debate and Practice* 2, 1, 89–108. DOI: https://doi.org/10.1332/174426406775249705.

[20] Jana Schellinger, Kerry Sewell, Jamie E. Bloss, Tristan Ebron, and Carrie Forbes. 2021. The effect of librarian involvement on the quality of systematic reviews in dental medicine. *PloS one* 16, 9, e0256833. DOI: https://doi.org/10.1371/journal.pone.0256833.

[21] Mikaela Aamodt, Hugo Huurdeman, and Hilde Strømme. 2019. 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, 4, 103–127. DOI: https://doi.org/10.18438/eblip29601.

[22] Deborah Meert, Nazi Torabi, and John Costella. 2016. Impact of librarians on reporting of the literature searching component of pediatric systematic reviews. *Journal of the Medical Library Association: JMLA* 104, 4, 267.

[23] Melissa L. Rethlefsen, Ann M. Farrell, Leah C. O. Trzasko, and Tara J. Brigham. 2015. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of clinical epidemiology* 68, 6, 617–626. DOI: https://doi.org/10.1016/j.jclinepi.2014.11.025.

[24] Connie Schardt, Martha B. Adams, Thomas Owens, Sheri Keitz, and Paul Fontelo. 2007. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC medical informatics and decision making* 7, 16. DOI: https://doi.org/10.1186/1472-6947-7-16.

[25] J. McGowan, Margaret Sampson, Douglas M. Salzwedel, Elise Cogo, Vicki Foerster, and Carol Lefebvre. 2016. *PRESS - Peer Review of Electronic Search Strategies*. *2015 Guideline Explanation and Elaboration*, Ottawa https://www.cadth.ca/resources/finding-evidence/press.

[26] Cyril W. Cleverdon. 1968. The critical appraisal of information retrieval systems. In *The International Congress of the International Federation for Documentation*, Moscow.

[27] Marc Rittberger and Werner Rittberger. 1997. Measuring quality in the production of databases. *Journal of Information Science* 23, 1, 25–37.

[28] Lisa Hartling, Robin Featherstone, Megan Nuspl, Kassi Shave, Donna M. Dryden, and Ben Vandermeer. 2016. The contribution of databases to the results of systematic reviews: a cross-sectional study. *BMC Medical Research Methodology* 16, 1, 127. DOI: https://doi.org/10.1186/s12874-016-0232-1.

[29] Michael T. Lam, Christina de Longhi, Joseph Turnbull, Helen R. Lam, and Reena Besa. 2018. Has Embase replaced MEDLINE since coverage expansion? *Journal of the Medical Library Association: JMLA* 106, 2, 227. DOI: https://doi.org/10.5195/jmla.2018.281.

[30] Wichor M. Bramer, Melissa L. Rethlefsen, Jos Kleijnen, and Oscar H. Franco. 2017. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews* 6, 1, 245. DOI: https://doi.org/10.1186/s13643-017-0644-y.

[31] Ginny Brunton, Claire Stansfield, Jenny Caird, and James Thomas. 2017. Finding relevant studies. In *An introduction to systematic reviews*, David Gough, Sandy Oliver and James Thomas, Eds. SAGE, Los Angeles, London, New Delhi, Singapore, Washington, DC, Melbourne, 93–122.

[32] Thomas Heinze, Dirk Tunger, Joel E. Fuchs, Arlette Jappe, and Paul Eberhardt. 2019. *Fachliche Forschungs- und Lehrprofile staatlicher Universitäten in Deutschland. Eine Kartierung ausgewählter Fächer* (2019). Retrieved from 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_Wupperta l_BUW.

[33] Éric Archambault, Étienne Vignola-Gagné, Grégoire Côté, Vincent Larivière, and Yves Gingrasb. 2006. Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics* 68, 3, 329–342. DOI: https://doi.org/10.1007/s11192-006-0115-z.

[34] Alberto Martín-Martín, Enrique Orduna-Malea, Mike Thelwall, and Emilio Delgado López-Cózar. 2018. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *JOURNAL OF INFORMETRICS* 12, 4, 1160–1177. DOI: https://doi.org/10.1016/j.joi.2018.09.002.

[35] Kayvan Kousha and Mike Thelwall. 2007. Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology* 58, 7, 1055–1065. DOI: https://doi.org/10.1002/asi.20584.

[36] Madian Khabsa and C. L. Giles. 2014. The number of scholarly documents on the public web. *PloS one* 9, 5, e93949. DOI: https://doi.org/10.1371/journal.pone.0093949.

[37] Marc Rittberger and Carolin Keller. 2020. *Literaturrecherche für Dossier Digi EBF 1*. Retrieved from. DOI: https://doi.org/10.7477/414:1:0.

[38] Marc Rittberger and Carolin Keller. 2021. *Literaturrecherche für Dossier Digi EBF 2*. Retrieved from. DOI: https://doi.org/10.7477/414:2:0.

[39] Annika Wilmers, Carolin Anda, Carolin Keller, and Marc Rittberger. 2020. *Bildung im digitalen Wandel: Die Bedeutung für das pädagogische Personal und für die Aus-und Fortbildung*. Waxmann. DOI: https://doi.org/10.31244/9783830991991.

[40] Annika Wilmers, Michaela Achenbach, and Carolin Keller. 2021. *Bildung im digitalen Wandel: Organisationsentwicklung in Bildungseinrichtungen*. Waxmann. DOI: https://doi.org/10.31244/9783830994558.

[41] David Gough and James Thomas. 2016. Systematic reviews of research in education: Aims, myths and multiple methods. *Review of Education* 4, 1, 84–102. DOI: https://doi.org/10.1002/rev3.3068.

[42] Werner Dees. 2015. *Bildungsforschung.* Humboldt-Universität zu Berlin, Philosophische Fakultät III.

[43] 2022. *Systematic Reviews in Education. DOI: https://doi.org/*10.17605/OSF.IO/NJTWC.

[44] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*. ACM Press, New York, USA.

[45] Seperia B. Wanyama, Ronald W. McQuaid, and Markus Kittler. 2021. Where you search determines what you find: the effects of bibliographic databases on systematic reviews. *International Journal of Social Research Methodology* 0, 0, 1–13. DOI: https://doi.org/10.1080/13645579.2021.1892378.