
Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers und Michael Kerres

Zwischen Präzision und Sensitivität

Generierung eines Studienkorpus am Beispiel einer Fragestellung zu Künstlicher Intelligenz (KI) in Bildungsprozessen

Verena Köstler¹ 

¹ Universität Passau

Zusammenfassung

Neben digitalisierungsbezogenen Fragestellungen im Kontext Lehren und Lernen im Allgemeinen rücken in jüngster Zeit auch im deutschsprachigen Raum KI-bezogene Fragestellungen im Besonderen in den Fokus der Forschung. In der weitgehend interdisziplinären Bearbeitung des Themas, international aufgegriffen unter artificial intelligence in education (AIED), sind technisch geprägte Fachdisziplinen im Vergleich zu pädagogisch-didaktischen Forschungsbereichen weitaus breiter vertreten (vgl. bspw. für die Hochschuldidaktik: Zawacki-Richter et al. 2019). Zur Stärkung eines erziehungswissenschaftlichen Zugangs wird die Generierung eines Studienkorpus exemplarisch für eine Forschungssynthese mit der Zielstellung eines systematischen Überblicks zu aktuellen Entwicklungen des KI-gestützten Lehrens und Lernens im Kontext institutioneller Bildung dargelegt. Die Entwicklung einer geeigneten Suchstrategie war geprägt von Abwägungen zwischen Präzision und Sensitivität (vgl. bspw. Campbell et al. 2018). Anhand der Prozesse Auswahl von Datenbanken, der Erstellung von Konzeptgruppen und Suchstrings sowie der Spezifizierung von Inklusionskriterien werden diese spezifiziert und für das finale, 51 Treffer umfassende Studienkorpus datenbankbezogen aufgezeigt. Es wird argumentiert, dass die Darlegung von Kennwerten zu Sensitivität und Präzision der verwendeten Datenquellen bei der Bearbeitung interdisziplinärer Fragestellungen, wie sie in bildungstechnologischen Themenfeldern häufig anzutreffen sind, einen Beitrag zur methodischen Qualitätsentwicklung zukünftiger Forschungssynthesen leisten kann.

Between Sensitivity and Precision. A Systematic Literature Search Illustrated by a Research Question on Artificial Intelligence in Education

Abstract

In addition to questions of digitally enhanced teaching and learning, research has recently become more focused on AI-related questions. Internationally known as artificial intelligence in education (AIED) the topic is largely worked on by technical

disciplines while educational disciplines are mostly underrepresented (e.g. for higher education: Zawacki-Richter et al. 2019). A systematic literature research is presented as a contribution to strengthen educational perspectives and as an example referring to a research synthesis with the objective of providing a systematic overview of current developments in AI-based teaching and learning in the context of institutional education. The development of the search strategy was characterized throughout by considerations in balancing precision and sensitivity (vgl. u. a. Campbell et al. 2018). These are described by the processes of selecting databases, creating concept groups and search strings, and specifying inclusion criteria. Precision and sensitivity are calculated for the final body of 51 studies. It is argued that reporting characteristic values on sensitivity and precision in processing interdisciplinary questions, as they are frequently encountered in educational technology, can contribute to the development of methodological quality of future research syntheses.

1. Konzeptueller Rahmen und Zielstellung

Fragen zu Künstlicher Intelligenz (KI) werden in Bildungsprozessen zum einen pädagogisch bedeutsam als didaktische Fragestellungen zu *Lernen mit KI*-gestützten Bildungstechnologien, zum anderen zu *Lernen über KI* als Bildungsgegenstand unter medienpädagogischen sowie ethischen Gesichtspunkten. Künstliche Intelligenz in Bildungsprozessen meint in diesem Beitrag Ersteres und fokussiert auf KI-gestützte Lehr-Lern-Situationen im institutionellen Kontext von Schule und Hochschule.

Der Ursprung der Technologie Künstlicher Intelligenz liegt in den 1950er-Jahren. Zunächst wurde mit symbolischen KI-Systemen versucht, bis dahin nur durch menschliche Intelligenz lösbare Probleme zu bearbeiten. Die Geschichte KI-gestützter Bildungstechnologien geht zurück auf die Entwicklung adaptiver Lerntechnologien seit den 1960er-Jahren (vgl. Buiu 1999; Guan, Mou, und Jiang 2020; Kahn und Winters 2021).

Im Fokus des vorliegenden Beitrags steht KI in seiner Bedeutung als aktuelles Schlagwort. Technologische Weiterentwicklungen und die Verfügbarkeit grosser Datenmengen ermöglichen vielfältige KI-gestützte Einsatzmöglichkeiten, die an beinahe täglich erscheinenden, für die Nutzergruppen leicht zugänglichen Anwendungen sichtbar werden (vgl. bspw. <https://www.futurepedia.io/>). Als zentraler Treiber der Künstlichen Intelligenz gilt die Methode des maschinellen Lernens, wonach moderne KI-Systeme anhand von Daten trainiert werden. Das Grundprinzip liegt darin, dass Algorithmen anhand von Beispieldaten bestimmte Muster erkennen und diese zur Problemlösung anwenden (Paass und Hecker 2020, 45ff.). Mit der so genannten *Deep Learning Revolution* (Sejnowski 2018) gelang für einen besonders leistungsfähigen Teilbereich des maschinellen Lernens, der auf künstlichen neuronalen Netzen basiert, ein Durchbruch (Silver et al. 2016; Silver et al. 2017). Die

Methode des Tiefen Lernens basiert auf künstlichen neuronalen Netzen, die an die Mechanismen des menschlichen Gehirns angelehnt sind (Paass und Hecker 2020, 79ff.). Während für maschinelles Lernen vorgegeben werden muss, welche Muster (nicht) erfasst werden sollen, wird dies bei der Methode des Tiefen Lernens selbstständig durch das System bestimmt (Schiefner-Rohs und Aufenanger 2021, 5). Eine Kategorisierung von Forschungsbemühungen der Jahre 2010 bis 2019 im Bereich KI-gestützter Bildungstechnologien (Feng und Law 2021), zeigt analog zum Eintritt dieser technologischen Revolution, dass Tiefes Lernen erstmalig 2018/2019 als neuer Forschungsfokus hinzutritt. Allerdings konzentrieren sich bislang nur wenige Studien zu KI-gestütztem Lernen auf Systeme mit der zentralen Neuerung der Daten- und Algorithmenorientierung (Kabudi, Pappas, und Olsen 2021).

International werden Fragen zum Einsatz KI-gestützter Bildungstechnologien für Lernprozesse weitgehend interdisziplinär bearbeitet und in der Forschung mit dem Label *artificial intelligence in education (AIED)* aufgegriffen. Während der Fachbereich *educational technologies* unter dem Begriff *Bildungstechnologien* auch im deutschsprachigen Raum Einzug gefunden hat (H. Niegemann und Weinberger 2020), ist *AIED* als womöglich zukünftig sehr eng daran angegliederte Disziplin kaum verbreitet. Bildungstechnologisch Forschende erfahren im deutschsprachigen Raum für ihre Erkenntnisse bislang eher im Kontext der Informatik Wertschätzung als in einer erziehungswissenschaftlichen Verortung (H. Niegemann und Weinberger 2020, 11). Das Spektrum des jeweiligen disziplinären Selbstverständnisses von Forschenden im Bereich Bildungstechnologien kann von überwiegend technologisch bis hin zu vorrangig erziehungswissenschaftlich reichen. Für den hochschuldidaktischen Kontext im internationalen Raum konstatieren Zawacki-Richter et al. (2019), dass technologisch geprägte Fachdisziplinen in Forschungsarbeiten zu KI-gestützten Anwendungen im Vergleich zu pädagogisch-didaktischen Forschungsbereichen überwiegen. Eine solche von technologischen Perspektiven dominierte Interdisziplinarität birgt aus erziehungswissenschaftlicher Perspektive, die dem vorliegenden Beitrag zugrunde liegt, verschiedene Herausforderungen. Diese rahmen den vorliegenden forschungssynthetischen Ansatz konzeptuell und werden im Folgenden illustriert.

Eine erste Herausforderung liegt in der Festlegung der **konzeptuellen Ausgangspunkte** von interdisziplinären Forschungsarbeiten. Eine retrospektive Beteiligung einer erziehungswissenschaftlichen Disziplin würde bedeuten, dass diese erst hinzugezogen wird, nachdem der konzeptuelle Ausgangspunkt durch die technologische Disziplin festgelegt wurde. Dann verbleibt nurmehr die Aufgabe, retrospektiv nach eigenen Anknüpfungspunkten bspw. im Sinne pädagogisch-didaktischer Zielstellungen zu suchen. Mangelnde proaktive Einbindung pädagogischer Expertise und fehlende theoretische Grundlagen in interdisziplinären Forschungsarbeiten zu KI-gestützten Bildungstechnologien werden an verschiedenen Stellen kritisiert

(Hew et al. 2019; Zawacki-Richter et al. 2019). In der Forschungssynthese von Bartolomé, Castañeda und Adell (2018) wurden aus erziehungswissenschaftlicher Perspektive Studien der vergangenen 50 Jahre zu bildungstechnologisch gestütztem Lernen im Kontext Personalisierung nach Merkmalen von Lehr-Lern-Situationen hin analysiert. Das Analyseschema erwies sich als schwer anwendbar. Es zeigte sich, dass die für die Implementation in Lehr-Lern-Situationen relevanten Fragestellungen «[...] are not themes that occupy the authors of educational technology focused on the personalisation of learning» (Bartolomé, Castañeda, und Adell 2018, 14).

In der Erprobung von KI-gestützten Bildungstechnologien für Lehr-Lern-Situationen im pädagogischen Feld liegen weitere Herausforderungen. Erziehungswissenschaftliche Forschung zu schulischem Unterricht bzw. hochschulischer Lehre beruht genuin auf konzeptuellen Überlegungen zur Komplexität von Lehr-Lern-Situationen mit zahlreichen zu berücksichtigenden und sich teils wechselseitig beeinflussenden Faktoren (zur Systematisierung vgl. bspw. Heuristik des Angebots-Nutzungs-Modells, Vieluf et al. 2020). Die Rolle von Lehrpersonen wird insbesondere in ihrer Steuerungsfunktion in institutionellen Lehr-Lern-Situationen als zentral angesehen (Augustsson und Boström 2016). Eine von technologischen Perspektiven dominierte Interdisziplinarität läuft Gefahr, bei der Entwicklung und Erprobung KI-gestützter Bildungstechnologien die komplexen Rahmenbedingungen zur Ermöglichung von Lehr-Lern-Prozessen in institutionellen Settings zu wenig zu berücksichtigen (Holstein und Alevén 2022). Es zeigt sich, dass die Erprobung KI-gestützter Bildungstechnologien unter der Bedingung von «real world educational settings» selten ist (Verdu et al. 2014; Misiejuk und Wasson 2017, 61; Zawacki-Richter et al. 2019, 26; Kabudi, Pappas, und Olsen 2021). Die Erkenntnis, dass Interaktionen sowohl zwischen Lehrpersonen und Lernenden als auch mit KI-gestützten Systemen für die Erprobung und Implementation KI-gestützter institutioneller Lehr-Lern-Situationen bedeutsam sind, erscheint aus erziehungswissenschaftlicher Perspektive selbstverständlich. Im Forschungsfeld *artificial intelligence in education* (AIED) finden konzeptuelle Überlegungen zur Berücksichtigung grösserer Komplexität erst seit wenigen Jahren unter den Begriffen *hybrid intelligence paradigm* sowie *hybrid human-AI frameworks* (Holstein, Alevén, und Rummel 2020; Molenaar 2022) statt.

Mit dem Einsatz von KI-gestützten Bildungstechnologien werden grosse Erwartungen im Bereich der Ermöglichung von Individualisierung verbunden (Hamisch und Kruschel 2022). Heterogene Lernausgangslagen führen in institutionell organisierten Lehr-Lern-Situationen dazu, dass bereitgestellte Lernangebote in unterschiedlicher Weise genutzt werden (Interaktionseffekt; Meissner et al. 2020). Heterogenität über individualisierende Strategien zu adressieren wird begründet mit Zielsetzungen zur Verbesserung von Lernleistungen (auf der Mikroebene schulischen Unterrichts bzw. hochschulischer Lehre) bis hin zur Ermöglichung des Abbaus sozialer Ungleichheit im Bildungssystem (Makroebene). Weder Individualisierung (im englischsprachigen

Diskurs: Personalisation, Walkington, und Bernacki 2020) noch damit eng zusammenhängende Begriffe wie Adaptivität werden jeweils exklusiv im allgemeindidaktischen, bildungstechnologischen oder AIED-Diskurs verwendet, noch bedeuten sie in den verschiedenen Kontexten notwendigerweise dasselbe («messy constructs», Pajares 1992). Im erziehungswissenschaftlichen, auf allgemeindidaktische Fragestellungen fokussierten Kontext bestehen kontroverse, sowohl evidenzbasierte (Wirkt Individualisierung?) als auch normativ (Wozu Individualisierung?) geprägte Argumentationslinien zur Begründung der Notwendigkeit von Individualisierung (Lipowsky und Lotz 2015; Rabenstein, Proske, und Idel 2018). Im bildungstechnologischen Diskurs rückt der Begriff der «Adaptivität» in einer spezifischen Bedeutung in den Vordergrund. Dieser bezeichnet Merkmale des technischen Systems, das sich in der Interaktion mit dem Lernenden in Bezug auf bestimmte Variablen (wie bspw. Vorwissen, Sprache, benötigte Lernzeit) anpasst (H. M. Niegemann et al. 2008, 307; Alevén et al. 2016, 525; H. Niegemann und Heidig 2020, 362). Die bereits angeführten hybriden Ansätze im Bereich *AIED* (Holstein, Alevén, und Rummel 2020; Molenaar 2022) erweitern den Geltungsbereich des Begriffs «Adaptivität». Über das Merkmal einer KI-gestützten Bildungstechnologie hinaus wird die Rolle der Lehrperson zur Ermöglichung von Adaptivität einbezogen. Gründe, aus denen die Entwicklung adaptiver Systeme nur langsam voranschreitet (Kerres et al. in Press, 5), lassen sich mit denen vergleichen, die auch die Umsetzung individualisierender Konzepte in der pädagogischen Praxis herausfordernd erscheinen lassen. Sie beziehen sich insbesondere auf die Schwierigkeit einer akkuraten Diagnostik der Kompetenzen der Lernenden sowie auf die Bereitstellung passender Lernangebote in unterschiedlichen Schwierigkeitsgraden.

Die Interdisziplinarität der Forschung zu KI-gestützten Bildungstechnologien stellt Forschende der verschiedenen disziplinären Hintergründe vor Herausforderungen. Bislang wird das Themenfeld vorwiegend ausgehend von den verschiedenen Anwendungen KI-gestützter Bildungstechnologien strukturiert (Zawacki-Richter et al. 2019; Hwang et al. 2020; OECD 2021; Feng und Law 2021). Der vorliegende Beitrag verlagert den Schwerpunkt und versucht, den aktuellen interdisziplinären Diskurs ausgehend von einer erziehungswissenschaftlichen Perspektive zu erschliessen. Dafür eignen sich Studien, die unter grösstmöglicher Berücksichtigung der Komplexität von Lehr-Lern-Situationen der pädagogischen Praxis («real-world educational settings») angelegt sind. Allerdings sind sie im Kontext KI-gestützter Bildungstechnologie bislang selten anzutreffen (Verdu et al. 2014; Misiejuk und Wasson 2017, 61; Zawacki-Richter et al. 2019, 26; Kabudi, Pappas, und Olsen 2021). Die Identifikation dieser Studien im institutionellen Kontext schulischen Unterrichts und hochschulischer Lehre beschreibt das zentrale Leitmotiv bei der Erstellung des Studienkorpus. Dessen Erstellung und erste Analyse stehen im Fokus des vorliegenden Beitrags.

Das Studienkorpus stellt den Ausgangspunkt der sich anschließenden Analyse und Synthese dar. Es soll dazu geeignet sein, anhand der Studiendesigns der darin enthaltenen Forschungsarbeiten konkrete Lehr-Lern-Szenarien herauszuarbeiten. Dazu wird untersucht, ob und ggf. welche (latenten) lehr-lern-theoretischen Einbettungen vorliegen, ob und ggf. welche (latenten) didaktischen Intentionen mit ihnen verbunden sind, sowie ob bzw. welche kritischen Reflexionen im Hinblick auf Implementationen in schulischen Unterricht oder hochschulische Lehre enthalten sind.

2. Erstellung des Studienkorpus

Grundsätzlich bedarf jede Forschungssynthese eines für die jeweilige Fragestellung möglichst repräsentativen und systematisch erstellten Studienkorpus. In Analogie zu den methodischen Verfahren der Bildungsforschung ist dieses in seiner Bedeutung mit der Stichprobe einer empirischen Studie vergleichbar. Die Beschreibung der Stichprobe einer Studie beschränkt sich in empirischen Arbeiten üblicherweise auf einen Abschnitt des Methodenteils. Die gründliche Darstellung ihrer Generierung und Zusammensetzung als möglichst repräsentatives, verkleinertes Abbild der Grundgesamtheit ist unabdingbar dafür, die Reichweite von Rückschlüssen aufgrund der gefundenen Ergebnisse einordnen zu können.

Methodologisch lassen sich Forschungssynthesen in aggregative und konfigurative Ansätze unterscheiden (Sandelowski et al. 2012; Newman und Gough 2020). Forschungssynthetische Methoden (für eine Übersicht vgl. bspw. Grant und Booth 2009) weisen oft Elemente beider Ansätze auf (Gough, Thomas, und Oliver 2012). Die grundlegende Herangehensweise zur Generierung eines repräsentativen Studienkorpus orientiert sich an der methodologischen Logik einer Forschungssynthese.

Erfordert es die Forschungsfrage, Hypothesen auf Basis empirischer Daten zu testen (bspw. Meta-Analysen), benötigt das Studienkorpus eine ausreichend hohe Anzahl an Studien für eine möglichst unverzerrte Aggregation. Eine möglichst vollständige Suche aller relevanten Studien geht mit diesem Ansatz einher («sufficient studies for unbiased aggregation»). Davon zu unterscheiden sind Forschungssynthesen mit offener gehaltenen Forschungsfragen, die Schlussfolgerungen auf Basis der in den Studien vorhandenen Konzepte ziehen (bspw. thematische Synthese). Angelehnt an das in qualitativen Forschungsmethoden verwendete Prinzip der Sättigung ist es bei der Generierung des Studienkorpus erforderlich, eine repräsentative Bandbreite von Studien zu identifizieren («sufficient concepts for coherent configuration», Brunton et al. 2017, 95ff.).

Die Forschungsfrage des vorliegenden Beitrags zielt darauf ab, anhand aktueller Forschungsarbeiten im Kontext der Erprobung von Lernen mit KI-gestützten Bildungstechnologien (Mikroebene schulischen Unterrichts bzw. hochschulischer Lehre) ein Studienkorpus zu generieren. In der Synthese sollen Schlussfolgerungen auf

Basis der aus den Studien abgeleiteten Lehr-Lern-Szenarien gezogen werden. Die Elemente der Lehr-Lern-Szenarien können dabei in manifester (bspw. verwendete KI-Bildungstechnologie, thematische Zuordnung) und latenter Form (bspw. didaktische Zielstellung, lehr-lern-theoretische Einbettung) enthalten sein. So entsteht ein Überblick zu aktuellen Entwicklungen zu Lernen mit KI-gestützten Bildungstechnologien aus erziehungswissenschaftlicher Perspektive. Dieses Vorgehen entspricht einer überwiegend konfigurativen forschungssynthetischen Logik.

Unabhängig von der methodologischen Ausrichtung besteht ein zentrales Anliegen von Forschungssynthesen darin, mit dem generierten Studienkorpus die Forschungsfrage adäquat adressieren zu können. Mit Aussagen zu Sensitivität und Präzision (Abbildung 1) lässt sich einschätzen, wie zielführend mit der gewählten Suchstrategie insgesamt und bezogen auf die verschiedenen Datenquellen relevante Treffer generiert werden können.

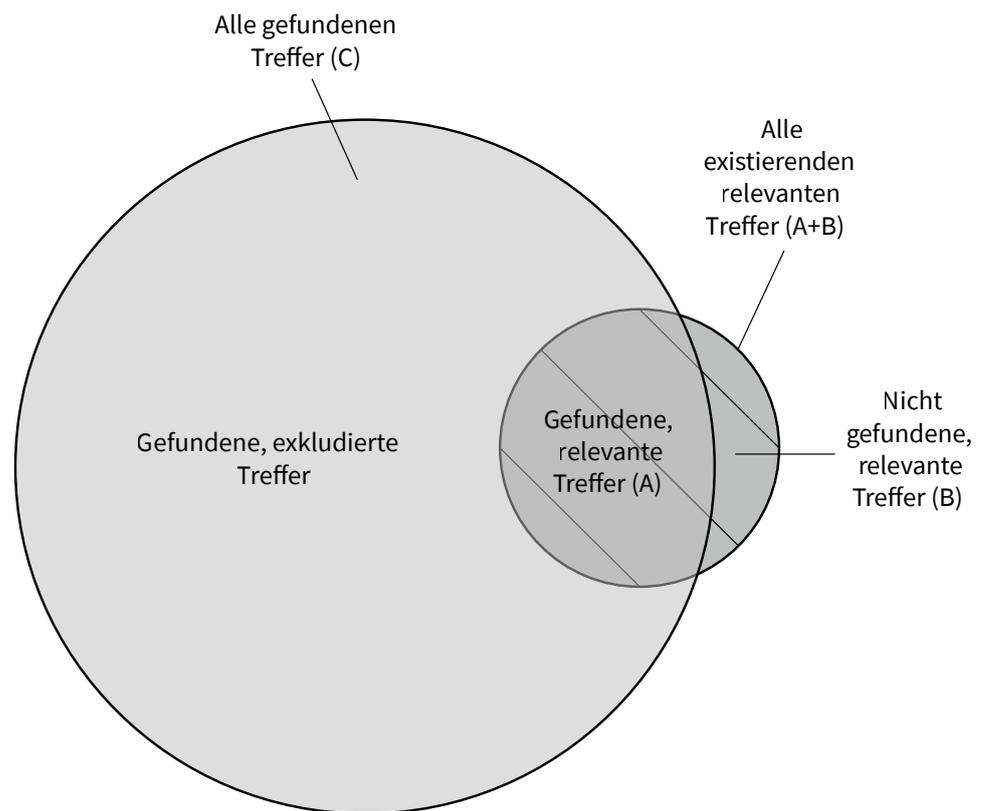


Abb. 1: Bezugsgrößen Sensitivität und Präzision, eigene Darstellung adaptiert nach Brunton et al. (2017, 112).

Sensitivität drückt als Qualitätsmerkmal einer entwickelten Suchstrategie das Verhältnis der gefundenen relevanten Treffer (A) zur nur theoretisch bestimmbaren Grundgesamtheit aller existierenden relevanten Treffer (A+B) aus (Brunton et al. 2017, 112). Um die Berechnung einer Sensitivitätskennzahl zu ermöglichen, wird die Grösse «alle existierenden relevanten Treffer (A+B)» in der Regel als Summe aller in den einbezogenen Datenbanken gefundenen relevanten Treffer operationalisiert (McFadden et al. 2012, 628 ; Best et al. 2014, 351; Campbell et al. 2018, 4). Im Sinne einer grösseren Allgemeingültigkeit des Begriffs für Forschungssynthesen, die neben Datenbanken auch weitere Datenquellen umfassen können, wird die Berechnung (Abbildung 2) unter dem Begriff Datenquellen-Sensitivität statt Datenbanken-Sensitivität dargestellt. Während auf Sensitivität schon mit Entscheidungen für oder gegen bestimmte Datenquellen Einfluss genommen wird, gibt Präzision darüber Auskunft, wie gut es mit der gewählten Suchstrategie innerhalb einer Datenquelle gelingt, irrelevante Treffer zu vermeiden (McFadden et al. 2012, 628). Dazu wird das Verhältnis aller gefundenen relevanten Treffer (A) ins Verhältnis gesetzt zur Gesamtanzahl aller gefundenen Treffer (C). Dabei gilt für das Verhältnis von Sensitivität und Präzision grundsätzlich ein gewisser Grad an «inverse relationship», wonach mit zunehmender Sensitivität (bspw. durch Einbezug zusätzlicher Datenbanken) die Präzision sinkt (Best et al. 2014, 351). Die richtige Balance zu finden, ist eine zentrale Aufgabe bei der Erstellung einer Suchstrategie zur Generierung eines Studienkorpus.

$$\text{(Datenquellen-) Sensitivität (\%)} = \frac{\text{Anzahl der gefundenen relevanten Treffer (A) einer Datenquelle}}{\text{Summe der relevanten Treffer (A) aller einbezogenen Datenquellen}} \times 100$$

$$\text{Präzision (\%)} = \frac{\text{Anzahl der gefundenen relevanten Treffer (A) einer Datenquelle}}{\text{Summe aller gefundenen Treffer (C) dieser Datenquelle}} \times 100$$

Abb. 2: Berechnung von (Datenquellen-)Sensitivität und Präzision, eigene Darstellung adaptiert nach Campbell et al. (2018, 4).

Neben den bislang in Forschungssynthesen zu erziehungswissenschaftlichen Fragestellungen kaum berichteten Qualitätsmerkmalen Sensitivität und Präzision finden sich disziplinübergreifend Qualitätsanforderungen, die sich sowohl auf die technische Ausführung, die Eignung von Methode und Studiendesign zur Adressierung der Forschungsfrage als auch auf die Bedeutung für die Forschungsfrage des Reviews beziehen. So garantiert eine methodisch-technisch tadellose Generierung eines Studienkorpus, in der alle erforderlichen Schritte eingehalten worden sind – von der Festlegung von Suchoperatoren bis zum Export von Studien – noch keine Relevanz der inkludierten Beiträge für die Zielstellung (Liabo, Gough, und Harden 2017, 253ff.).

Entsprechend transparent und umfassend müssen Publikationen zu Forschungssynthesen Aspekte erläutern, anhand derer sich Rückschlüsse auf Qualitätskriterien im Kontext des methodologischen Zugangs ziehen lassen.

Die Generierung eines Studienkorpus beinhaltet von der Zielstellung ausgehend verschiedene Schritte (vgl. bspw. Newman und Gough 2020). Sie umfassen die Entwicklung einer Suchstrategie und die Studienausswahl. Es müssen Entscheidungen zu Datenquellen, Suchstring sowie In- und Exklusionskriterien getroffen werden. Diese werden im Folgenden unter Überlegungen zu Sensitivität und Präzision am Beispiel des generierten Studienkorpus erläutert.

2.1 Auswahl von Datenbanken

Dem Studienkorpus lag eine datenbankbasierte Suchstrategie zugrunde. Diese umfasste die Datenbanken *ERIC* (via *Proquest*), *FIS-Bildung* inklusive *BASE* (via *Fachportal Pädagogik*), *LearntechLib* sowie *Web of Science* (via *Clarivate*). Durch die Auswahl dieser Datenquellen war die Grundgesamtheit aller zur Verfügung stehenden relevanten Beiträge und dadurch die Basis für Kennwerte der (Datenquellen-)Sensitivität bestimmt.

Der Weg zur Festlegung auf die genannten Datenquellen umfasste verschiedene Schritte. Ausgangspunkt war die Überlegung, andere als durch die eigene erziehungswissenschaftlich geprägte Perspektive bekannte Datenbanken einzubeziehen. In Anbetracht der Fülle zur Verfügung stehender Datenbanken war zunächst die Unterscheidung zwischen dem eigentlichen Repositorium «Datenbank» und dazugehörigen Datenbankbetreibern bedeutsam. Basierend auf Recherchen in verschiedenen methodischen Nachschlagewerken zu systematischen Reviews (Grant und Booth 2009; Newman und Gough 2020; Zawacki-Richter et al. 2020) sowie einer Sichtung der verwendeten Suchstrategien von Forschungssynthesen aus dem Kontext Digitalisierung/KI zu Fragen des Lehrens und Lernens (Tabelle 1) wurden zunächst sechs Datenbanken als Ausgangspunkt gewählt. Mit *ERIC* und *FIS-Bildung* lagen Datenquellen vor, die insbesondere Publikationen aus bildungswissenschaftlichen Disziplinen abbildeten und neben einem internationalen auch einen deutschsprachigen Fokus aufgriffen.

Review	Fokus des Reviews Einbezogener Zeitraum (Anzahl aller gefundenen Treffer/relevante in den Studienkorpus inkludierte Treffer)	Datenquellen
Zawacki-Richter et al. (2019)	AI applications in higher education 2007-2018 (1549/46)	Datenbanken (EBSCO Education Source, Web of Science, Scopus)
Bond et al. (2020)	Student engagement and educational technology 2007 – 2016 (18068/243)	Datenbanken (ERIC, Web of Science, PsycINFO, Scopus)
Bernacki, Greene und Lobczowski (2021)	Personalized learning and educational technology 2010 – 2018 (1597/376)	Datenbanken (PsychINFO, ERIC, IEEE Xplore)
Feng und Law (2021)	Artificial intelligence in education research 2010-2019 (1858/1830)	5 Zeitschriften im Themenfeld AIED
Chen, Ifenthaler und Yau (2021)	Applied educational technologies 2000-2020 (1975/38)	Zeitschriften mit hohem Impact Faktor, Google Scholar, Datenbanken (Web of Science, Science Direct), MOOC Plattformen
Kabudi, Pappas und Olsen (2021)	AI-enabled adaptive learning systems 2014-2020 (1492/147)	Datenbanken (ACM, Web of Science, EBSCO Host, Wiley, SAGE Journals, IEEE Xplore, Scopus, Taylor and Francis)

Tab. 1: Sichtung von Forschungssynthesen im Kontext Digitalisierung/KI/Bildungsprozesse: Auszugsweise Darstellung nach inhaltlichem Fokus, einbezogenem Zeitraum, inkludierten Treffern und Datenquellen.

Mit *Web of Science* und *Scopus* sowie *IEEEXplore* und *LearntechLib* lagen Paare von Datenbanken vor, mit denen sich jeweils ein ähnliches Kriterium abbilden liess. Eine im Vergleich zu *ERIC* und *FIS-Bildung* grössere multidisziplinäre Bandbreite, die auch ein grösseres Spektrum aus den informationswissenschaftlichen Disziplinen umfasste, liess sich den Datenbanken *Scopus* und *Web of Science* zuordnen. Mit der Berücksichtigung von *IEEEXplore* oder *LearntechLib* sollte versucht werden, dem insbesondere in technischen Disziplinen stärker verbreiteten Publikationsverhalten der Veröffentlichung und Listung von Konferenzbeiträgen Rechnung zu tragen. Im Verlauf der Sichtung der Datenbanken hinsichtlich Usability-Aspekten und Funktionsweise wurde zur Berücksichtigung des Kriteriums multidisziplinäre Bandbreite die Datenbank *Web of Science* sowie für das Kriterium des Publikationsverhaltens *LearntechLib* ausgewählt. Beide wurden aufgrund von Filteroptionen (*Web*

of Science) sowie ihrer spezifischeren konzeptuellen Ausrichtung (*LearntechLib*: Bildungstechnologien im Kontext Lehren und Lernen) als besser geeignet für die Generierung von Studien im Sinne der Zielstellung eingeschätzt.

Ausgehend von der Datenbank *ERIC* wurde der Suchstring in iterativen Schleifen entwickelt. Dieser Prozess steht im Fokus des folgenden Abschnitts.

Eine vergleichende Abwägung der Usability-Aspekte der verschiedenen Datenbanken war unabdingbar und soll an dieser Stelle anhand ausgewählter Beispiele verdeutlicht werden. Auch wenn dieses Kapitel zur Generierung eines Studienkorpus linear aufgebaut ist, war die Erprobung der Suchstrategie ein sehr iterativer Prozess. Dies wird daran deutlich, dass bspw. die Auswahl einer Datenbank nicht ohne eine erste Erprobung mit entsprechenden Suchoperatoren erfolgen konnte. So verfügten einige Datenbanken über ein kontrolliertes Vokabular (Thesaurus), dessen Berücksichtigung innerhalb des Suchstrings insofern empfehlenswert war, da diese Begriffe hierarchisch geordnet waren und so, je nach ihrer Stellung, eine entsprechend präzisere Suche ermöglichten. Für die Ausführung iterativer Erprobungsschleifen und die Dokumentation der entsprechenden Ergebnisse war eine Speicherungsoption sowie eine differenzierte Anzahl von Kombinations- und Filtermöglichkeiten für Suchbegriffe hilfreich. Ersteres war bei allen erprobten Datenbanken (bei *FIS-Bildung* begrenzt auf einen auf die aktuelle Sitzung bezogenen Suchverlauf) verfügbar. Zweiteres fand sich in der Regel unter der Option «erweiterte Suche / advanced search». Sofern möglich erfolgte eine personalisierte Registrierung, um Suchverläufe in chronologischer Reihenfolge zu dokumentieren und für den Export in die verwendete Literaturverwaltungssoftware *Citavi* vorzubereiten. In allen Datenbanken waren verschiedene Filteroptionen wie Sprache (Deutsch, Englisch) sowie Publikationsquellen (Zeitschriftenartikel und im Falle von *LearntechLib* zusätzlich Konferenzbeiträge) verfügbar. Der Suchzeitraum umfasste 2020 bis Mai 2022. Der enge zeitliche Fokus ist zum einen dadurch bedingt, dass das Studienkorpus aktuelle Forschungsarbeiten zu Lernen mit KI-gestützten Bildungstechnologien zum Durchführungszeitpunkt der Forschungssynthese enthalten sollte. Zum anderen dient es innerhalb der forschungspraktischen Limitationen (einjähriges Einzelstipendium, 80 Stunden pro Monat) dazu, den Trefferumfang so festzulegen, dass der gewählte forschungssynthetische Zugang exemplarisch erprobt und auf seine grundsätzliche Übertragbarkeit hin eingeschätzt werden kann. Auf die ausgewählten Datenbanken wurde über den universitären Zugang und im Falle von *LearntechLib* über eine dreimonatige kostenpflichtige Mitgliedschaft zugegriffen.

2.2 Erstellung von Konzeptgruppen und Suchstrings

Der finale und an die Besonderheiten der jeweiligen Datenbanken adaptierte Suchstring umfasste zwei Konzeptgruppen zur Erfassung der Aspekte «Künstliche Intelligenz» sowie «Adressaten in institutionellen Lehr-Lernsettings». Im iterativen Prozess zur Entwicklung passender Suchoperatoren wurde in Anlehnung an Campbell et al. (2018) versucht, verschiedene Techniken zu nutzen, um die Präzision der Suchstrategie bezogen auf die einzelnen Datenbanken zu erhöhen.

Eine möglichst präzise Suchstrategie entspricht dem methodologischen Paradigma der Forschungssynthese dieses Beitrags. Wie einleitend in diesem Kapitel beschrieben, überwiegt eine konfigurative Logik. Das Studienkorpus soll in der sich anschließenden Synthesephase dazu geeignet sein, sowohl manifeste als auch latente Elemente in Form von Lehr-Lern-Szenarien abzuleiten. Der Fokus liegt nicht auf der Generierung eines Suchstrings, mit dem eine möglichst hohe Anzahl von Studien aggregiert werden kann. Stattdessen sollen Studien mit möglichst hoher Informationsdichte («information-rich cases», Patton 1990) in Bezug auf die Komplexität von Lehr-Lern-Situationen und hoher Anschlussfähigkeit an eine Analyse und Synthese aus erziehungswissenschaftlicher Perspektive identifiziert werden.

Zunächst wurden grundsätzlich geeignete Suchoperatoren getestet. Dabei stand im Fokus, eine Kombination von Begrifflichkeiten zu finden, die weder zu eng spezifizierten noch zu weit griffen, um die gesuchten und teils nur latent vorhandenen Konstrukte in Form von Lehr-Lern-Situationen erfassen können («hard to detect evidence», O'Mara-Eves et al. 2014). Oftmals wird der PICO-Rahmen (*people, intervention, comparators, outcome*) als Unterstützungsstrategie zur Entwicklung einer Forschungsfrage im Allgemeinen (Da Santos, Pimenta, und Nobre 2007) sowie der Entwicklung des Suchstrings (bspw. bei Tai et al. 2020) vorgeschlagen. Dieser erwies sich im vorliegenden Fall als wenig hilfreich. Weder die Formulierung von Komparatoren noch von Outcomes leistete einen zielführenden Beitrag zur Konzeptionalisierung eines zur Zielstellung passenden Suchstrings.

Begriffe, für die sich in den verschiedenen disziplinären Kontexten von Erziehungswissenschaft und Informationstechnologie Verwendungen in unterschiedlicher Bedeutung zeigten, wurden im Sinne einer Erhöhung der Präzision ausgeschlossen. Dies traf bspw. auf *learn** oder *adaptiv** zu. Mit ersterem wurde versucht, die Mikroebene eines Lehr-Lernsettings als Suchoperator zu erfassen. Allerdings wird der Begriff im Kontext KI überwiegend dazu verwendet, das lernende technische System zu bezeichnen, und erzeugte daher zahlreiche irrelevante Treffer. Im Falle von *adaptiv** verhielt es sich insofern ähnlich, da auch hier in Studien im KI-Kontext das technische Merkmal der *Adaptivität eines Systems* hervortrat, nicht die aus erziehungswissenschaftlicher Perspektive beabsichtigte Erfassung von individualisierenden Lehr-Lernsettings. Auch vermeintlich mit KI-Technologie verbundene Begriffe führten in die Irre. Mit dem als KI-bezogen verstandenen Begriff *deep*

learning, der darauf abzielte, die im Bereich maschinellen Lernens verwendeten Verfahren mehrschichtiger neuronaler Netze zu erfassen (vertiefend bspw. bei Paass und Hecker 2020), wurden überwiegend Treffer erzielt, in denen der Begriff im Sinne *tiefergehender Lernprozesse lernender Personen* verwendet wurde.

Der Weg von ersten Erprobungsschleifen zum finalen Suchstring schloss eine Reduzierung der Konzeptgruppen von drei auf zwei sowie eine Reduktion einer größeren Anzahl zunächst sehr eng spezifizierender Suchoperatoren zugunsten weniger weiter greifender Begriffe ein. Begonnen wurde damit, Konzeptgruppen um die Konzepte «Künstliche Intelligenz», «Mikroebene Lehren und Lernen» sowie «institutionalisierte Bildungseinrichtung» zu bilden und über Boolesche Operatoren zu kombinieren. Das zentrale Leitmotiv des iterativen Vorgehens lag auf der Identifikation der grundsätzlich seltenen Studien (Verdu et al. 2014; Misiejuk und Wasson 2017, 61; Zawacki-Richter et al. 2019, 26; Kabudi, Pappas, und Olsen 2021), die unter grösstmöglicher Berücksichtigung der Komplexität von Lehr-Lern-Situationen der pädagogischen Praxis («real-world educational settings») angelegt sind. Die spezifischen Kriterien, mit denen sich diese Art von Studien im Sinne von In- und Exklusionskriterien (siehe nachfolgender Abschnitt) kategorisieren liessen, entstanden in diesem Prozess («iterative review», Brunton et al. 2017, 102). Bei der Erprobung der verschiedenen Konzeptgruppen und deren Kombination wurde ein qualitatives Screening der jeweils ersten 20 Treffer (Titel und Abstract) durchgeführt. Schritt für Schritt entstanden so Inklusionskriterien, mit denen sich Studien im Sinne der Zielstellung identifizieren liessen.

Das Kriterium der KI-gestützten Bildungstechnologie wurde über die Konzeptgruppe «Künstliche Intelligenz» mit den Suchoperatoren «artificial intelligen*» OR «machine learning» operationalisiert. Grundsätzlich waren zwei verschiedene Herangehensweisen möglich. So konnten sehr spezifische Suchoperatoren entlang der verschiedenen technischen Umsetzungen gebildet oder eine breitere Konzeptualisierung mit allgemeineren KI-Begrifflichkeiten fokussiert werden, die den neueren technischen Entwicklungen übergeordnet sind (Kabudi, Pappas, und Olsen 2021, 3). Letzterem wurde der Vorzug gegeben, um in möglichst grosser Bandbreite Beiträge zu erfassen, die sich zwar als Forschung zu Künstlicher Intelligenz verstehen, allerdings keine Beiträge von sehr spezifischen Communities innerhalb von KI-Technologien überproportional abbildeten.

Finale Konzeptgruppen	Suchstring
Künstliche Intelligenz	“artificial intelligence” OR “artificial intelligen*” OR “machine learning”
Adressaten in institutionellen Lehr-Lern-Settings	pupil* OR child* OR student*
Verworfenne Konzeptgruppen	
Künstliche Intelligenz	“artificial intelligence” OR “artificial intelligen*” OR “machine learning” OR “deep learning”
Mikroebene Lehren und Lernen	(individual* OR personal* OR adaptiv* OR heterogen*) AND (instruction OR teaching OR learning OR education)
Institutionalisierte Bildungseinrichtungen	“higher education” OR universit* OR college* OR undergrad* OR graduate OR postgrad* OR “K-12” OR “primary school” OR “middle school” OR school OR “post secondary” OR “post-secondary”

Abb. 3: Finale und verworfene Konzeptgruppen des Suchstrings.

Um Studien mit dem Kriterium Lehr-Lern-Situationen zu identifizieren, die grundsätzlich für die institutionellen Kontexte schulischer Unterricht bzw. hochschulische Lehre geeignet sind, wurden zunächst Konzeptgruppen zu «Mikroebene Lehren und Lernen» sowie «institutionalisierte Bildungseinrichtungen» gebildet. Orientiert an Forschungssynthesen, die Studien auf *classroom level* fokussierten (vgl. bspw. Bedenlier et al. 2020; Tai et al. 2020), wurde versucht, Suchstrings zu adaptieren. Neben den beschriebenen Uneindeutigkeiten von Begrifflichkeiten im interdisziplinären Kontext (bspw. *adaptiv**, *learn**) zeigte sich im qualitativen Screening, dass die Kombination aus den beiden Konzeptgruppen mit der Konzeptgruppe «Künstliche Intelligenz» nicht die intendierte Art von Studien generierte. Eine Lösung zur präziseren Erfassung des Aspekts von Lehr-Lern-Situationen auf der Mikroebene institutioneller Kontexte fand sich in der Verschiebung des Fokus auf die Konzeptgruppe «Adressaten in institutionellen Lehr-Lernsettings» mit den Suchoperatoren «student*» OR «child*» OR «pupil*». Zwar enthielt auch der Suchoperator «pupil*» eine gewisse Unschärfe durch seine physiologische Bedeutung «Pupille» zusätzlich zur lehr-lernbezogenen Bedeutung «Schüler:in». Diese geringere Präzision wurde in Kauf genommen und entsprechende irrelevante Treffer im Rahmen der Filterung nach In- und Exklusionskriterien unaufwendig aussortiert.

Der finale Suchstring (Abbildung 3) wurde ausgehend von der Datenbank *ERIC* erprobt und an die Spezifika der anderen Datenbanken adaptiert. Im Fall der multidisziplinären Datenbank *Web of Science* wurde die datenbankinterne Kategorisierung

der Beiträge genutzt und auf Beiträge mit dem Schwerpunkt «Education Educational Research» fokussiert. Im Mai 2022 wurden die vier Datenbanken (*ERIC*, *Web of Science*, *FIS-Bildung*, *LearntechLib*) in den Kategorien Titel, Abstract und Keywords durchsucht. Titel wurden in das Literaturverwaltungsprogramm *Citavi* importiert und die importierten Rohdaten durch Gruppenlabels mit Datenbankbezeichnung versehen. Der eingeschränkte Funktionsumfang in der Datenbank *LearntechLib* machte es nötig, den Suchstring zu splitten, zwei Importe in das Literaturverwaltungsprogramm durchzuführen und in diesem Prozess nicht zu vermeidende Mehrfachimporte wieder zu löschen. Über die Datenbanken hinweg wurde eine Dublettenprüfung durchgeführt und sieben gefundene Dubletten wurden ausgeschlossen.

2.3 Spezifizierung von Inklusionskriterien

Für die Kodierung der 442 gefundenen Treffer zur Extraktion von für die Zielstellung relevanten Studien wurden Inklusionskriterien erarbeitet. Das Suchkonzept im interdisziplinären Kontext erforderte zunächst ein Inklusionskriterium «Bildungssetting», um den Kontext Lehren und Lernen in jeglicher Fachdisziplin zu validieren. Weiterhin ging es darum, überhaupt Studien zu identifizieren, die als spezifisch im Sinne einer durchgeführten Erprobung bezeichnet werden konnten. Entscheidend war nicht der Standard eines Experimental-Kontrollgruppen-Designs oder die Art der untersuchten abhängigen Variable, sondern das Kriterium, dass eine Erprobung, bspw. auch evaluativer Art oder in Form von Aktionsforschungsansätzen, im Hinblick auf einen konkreten Fokus stattgefunden hatte. Im Hinblick auf das KI-basierte Setting war entscheidend, dass eine KI-basierte Anwendung auch wirklich eingesetzt wurde und sich diese Anwendung nicht mehr im Entwicklungsstadium befand. Fokussiert wurde auf Studien, die sich mit der Zielgruppe lernender Personen in institutionellen Kontexten auseinandersetzten, und die Settings beschrieben, die grundsätzlich im Rahmen von Planung, Durchführung und/oder Reflexion von Lehren und Lernen auf der Mikroebene schulischen Unterrichts oder hochschulischer Lehre stattfinden konnten.

Inklusionskriterien im Überblick:

- Bildungssetting/Kontext Lehren und Lernen jeglicher Fachdisziplin
- Spezifisch (Primärstudie, Erprobung, Evaluation)
- Zielgruppe: Lernende in Schule und Hochschule ab der ersten Jahrgangsstufe
- KI-basiertes Setting/Einsatz einer KI-basierten Anwendung
- Bezug zu Mikroebene Lehr-Lern-Situation im Interaktionsgeschehen von Lehrperson/Lernende

In einem ersten Screeningprozess, der auf Titel und Abstracts fokussierte, wurden im Literaturverwaltungsprogramm *Citavi* den importierten Beiträgen die Kategorien «Treffer» und «kein Treffer» zugeordnet. Die Kategorie «Treffer» wurde großzügig verwendet. Sie enthielt Beiträge, bei denen aufgrund von Abstract und Titel davon auszugehen war, dass alle Inklusionskriterien zutrafen (60). Auch unklare Fälle wurden einbezogen, bei denen mindestens eines der Inklusionskriterien nicht beurteilt werden konnte (56). Für alle «Treffer» (116) wurden Volltexte recherchiert und sie wurden einem Screening nach Inklusionskriterien unterzogen.

Die Verwendung der SFX-Funktionalität (Self eXtracting) in den Datenbanken *ERIC*, *Web of Science* sowie *FIS-Bildung* erwies sich als äusserst hilfreich für ein effizientes Auffinden der Volltexte. Sonderbestellungen für Studien, die nicht frei verfügbar oder für die keine universitäre Zugangsmöglichkeit lizenziert war, liessen sich teilweise durch Kontaktanfragen an die Autor:innen über *Researchgate* vermeiden. In *LearnTechLib* war der Zugriff auf Volltexte grösstenteils in der Registrierung inkludiert. Der Volltext eines Beitrags konnte nicht gefunden werden und wurde ausgeschlossen.

Das Screening der Volltexte nach In- und Exklusionskriterien wurde durch die Autorin durchgeführt. Zur Qualitätssicherung bot sich angesichts der Komplexität der in den Studien zu kodierenden Konstrukte ein Prozess der inhaltlichen Validierung an. Dazu wurde für jeden relevanten Beitrag (klare und unklare Fälle nach dem ersten Screening) ein kurzer Textabschnitt angefertigt, der für die Studie beschrieb, worin die spezifische Erprobung für welche Zielgruppe, in welchem KI-basierten Setting und mit welcher für das Lernen bezogenen Zielstellung bestand. Volltexte, auf die nicht alle Inklusionskriterien zutrafen, waren auch nicht im Sinne einer solchen Textpassage beschreibbar und wurden ausgeschlossen. Unklare Fälle betrafen den Aspekt des «KI-basierten Settings» und wurden mit einer Kollegin aus dem Bereich «AI studies» gelöst.

3. Ergebnisse

Das finale Studienkorpus enthielt 51 Treffer (Tabelle 2) und umfasste somit rund 12 % aller Treffer nach Datenbankrecherche. Im Schnitt mussten also 9 Beiträge kodiert werden, um einen für das Studienkorpus relevanten Treffer zu generieren («number needed to read» in Anlehnung an «number needed to treat», Best et al. 2014, 351).

Datenbank	Gefundene Treffer*	Relevante Treffer nach Screening I (Abstract, Titel)	Sensitivität(I)	Präzision(I)	Relevante Treffer nach Screening II (Volltext)	Sensitivität(II)	Präzision(II)
<i>FIS</i>	22	10	9%	45%	1	2%	5%
<i>LearnTechLib</i>	98	22	19%	22%	2	4%	2%
<i>Web of Science</i>	250	66	57%	26%	41	80%	16%
<i>ERIC</i>	79	18	16%	23%	7	14%	9%
Gesamt	449	116	100%		51	100%	

Tab. 2: Sensitivität und Präzision der verwendeten Datenbanken nach Screening I (Abstract und Titel) und II (Volltext); * 7 Dubletten enthalten (irrelevante Treffer).

Die Auflistung von Sensitivität und Präzision der verwendeten Datenbanken nach dem ersten Screening (Abstract und Titel) und dem zweiten Screeningvorgang (Volltext) zeigte grosse Unterschiede. Mit der Datenbank *Web of Science* gelang es innerhalb der Suchstrategie am sichersten, relevante Treffer für das Studienkorpus zu erzielen. Sie weist durchgehend die höchsten Werte für Sensitivität und Präzision auf. Die Datenquellen-Sensitivität stieg im zweiten Screening sogar von 57 auf 80 % an. Mit ERIC lag eine Datenbank vor, mit welcher der Anteil an relevanten Treffern für das Studienkorpus sowohl nach dem ersten als auch nach dem Screening konstant blieb (16%, 14%). Die Präzision und somit der Anteil relevanter Beiträge, die über die Datenbank gefunden wurden, hat sich nach dem Screening der Volltexte mehr als halbiert (23%, 9%). Obwohl der Suchstring in *FIS-Bildung* zu vergleichsweise wenigen Treffern führte, war die Präzision nach dem ersten Screening sehr hoch und es wurde für beinahe jeden zweiten Treffer ein Volltext importiert. Davon wurde nur ein Beitrag in das finale Studienkorpus aufgenommen. Aus der Datenbank *LearnTechLib* fanden zwei Treffer Eingang in das finale Studienkorpus.

Insgesamt wurden 326 Beiträge der gefundenen Treffer und sieben Dubletten im ersten Screening sowie 65 Treffer im weiteren Screening als für die Zielstellung irrelevant ausgeschlossen. Zwar war der Suchstring grundsätzlich auf Präzision angelegt, enthielt aber breit gefasste Begrifflichkeiten als Suchoperatoren. Dadurch fanden sich auch Beiträge, die sich nicht auf den Kontext Lehren und Lernen, sondern bspw. auf Medizin oder Neurowissenschaften («pupil» im Sinne von Pupille; maschinelles Lernen im Bereich der Diagnose von Autismus) bezogen (37). Neben spezifischen Beiträgen fanden sich zudem Überblicksarbeiten (Forschungssynthesen, Frameworks, Curricula), technische Arbeitsberichte (bspw. technische Integration in universitäre Infrastrukturen, Funktionstests) und auch Beiträge, in denen keine Erprobung stattfand (87). Ausgeschlossen wurden darüber hinaus Beiträge,

die sich auf eine nicht relevante Zielgruppe, bspw. die frühkindliche Bildung fokussierten (7). Daneben fanden sich Studien ohne KI-Setting (32) und solche, die zwar grundsätzlich einen KI-Bezug aufwiesen, allerdings nicht im Sinne des Einsatzes einer konkreten KI-Anwendung, sondern in denen bestimmte Überzeugungen von Personengruppen (bspw. was denken Lehrpersonen über Chatbots) oder medienpädagogische und ethische Aspekte (bspw. Fairness/Bias durch Algorithmen, ethische Fragen von Learning Analytics) aufgegriffen wurden (49). Das Kriterium *Mikroebene Lehr-Lern-Situation* traf auf einen Teil der Studien nicht zu (29) und solche mit einem Fokus auf *MOOCs* (*massive open online courses*) wurden ausgeklammert (14). Ein Ausschluss galt ebenso für Beiträge, deren Schwerpunkt im Bereich *educational data processing* und/oder *student modeling* lag (bspw. Generierung und/oder Überprüfung der Akkuratheit des verwendeten prädiktiven Modells; Entwicklung und/oder Erprobung von Algorithmen/Modellen) (119) oder in denen maschinelles Lernen als statistisches Analysetool zum Einsatz kam (53). Auf eine Studie konnten mehrere Exklusionskriterien zutreffen.

An den 51 Beiträgen des generierten Studienkorpus waren Autor:innen aus 19 verschiedenen Ländern beteiligt. Der Schwerpunkt liegt auf Taiwan (18), den USA (13), gefolgt von China (8). Es folgten Beiträge aus Südkorea, Japan, Deutschland, Schweden (je 3); Australien, Türkei (je 2); Saudi-Arabien, Brasilien, Algerien, Neuseeland, Yemen, Chile, England, Italien, Niederlande (je 1). Im Suchzeitraum verteilten sich die Studien auf die Jahre 2020 (12), 2021 (23) und 2022 (16). Der fachliche Kontext bezog sich überwiegend auf MINT-Themen (22), gefolgt von inhaltsunabhängigen bzw. -übergreifenden Kontexten (11), Fremdsprachenlernen (9), Medizin (3), Wirtschaftswissenschaften (3), Sprachunterricht in der Muttersprache (2) sowie Musik (1). Aus technischer Perspektive war das Kriterium der Verwendung einer KI-basierten Technologie durchgängig erfüllt. Es konnte in allen Fällen in Form der verwendeten KI-Techniken¹ validiert werden (enthalten waren: Natural Language Processing, maschinelles Lernen in Form von überwachtem, unüberwachtem oder tiefem Lernen, Computer Vision sowie Robotics). In 13 Studien liess sich zusätzlich (mindestens) ein konkretes KI-Modell spezifizieren (bspw. das im Bildungsbereich verbreitete Modell BERT [3]). Die Studien wurden im Hinblick auf die konkrete Nennung der verwendeten KI-gestützten Tools oder Systeme analysiert, in denen KI-gestützte Anwendungen integriert waren. In 31 Studien wurden 29 verschiedene gefunden: BookRoll (3), CASUS (2) sowie LOsMonitor, AR Pianist, CILLE, SoLearn, Roboter Zenbo, DACMs, MIT App Inventor, IR-App, TPR AI, NetLogo, ALEKS, Hybrid-AQG, MaTHiSiS, Mosoteach, AVW-Space, LIS system, EconBot, CAPELS, Squirrel AI Learning, MOCA, Algebra Nation, HASbot, Custom Vision, Replika App, Chatbot created with «collect-chat», Chatbot «Ellie», e-rater-ML (je 1). Die enthaltenen KI-gestützten

1 Zu beachten ist, dass sich KI-Techniken meist nicht streng voneinander abgrenzen lassen, sondern ineinander greifen.

Settings lassen sich im Hinblick auf ihre Funktion für Lehr-Lern-situationen in vier verschiedenen, sich teils überschneidenden Bereichen beschreiben: *Assessment* in vielfältigen Umsetzungsformen (26) in der Phase der Durchführung, Planung oder Evaluation von Lehr-Lern-Situationen; die *Bereitstellung* von Scaffolding (7); der *Ein-satz von KI-gestützten Systemen* als Partner für Kommunikation und/oder Demonstration (17); die *Erprobung von KI* durch Lernende im Sinne medienpädagogischer Reflexion über KI (3) (Beispielreferenzen siehe Tabelle 3).

Funktion des KI-Settings für Lehr-Lern-Situationen	Beispielreferenz	Fachlicher Kontext/KI-Tool
Assessment (Phase der Durchführung)	Hunte et al. 2021 Hege, Kieseewetter und Adler 2020	Sprachunterricht (Muttersprache)/- Medizin/CASUS
Assessment (Phase der Planung)	Alammary 2021	MINT (Informatik)/LOs Monitor
Assesement (Phase der Evaluation)	Mubarak, Cao und Zhang 2022	Inhaltsunabhängig-/übergreifend/-
Partner für Kommunikation und Demonstration	Cui 2022 Yang et al. 2022	Musik/AR Pianist EFL/Chatbot «Ellie»
Scaffolding	Y. P. Cheng, S. C. Cheng und Y. M. Huang 2022	MINT (Informatik)/DACMs
KI-Erprobung zur medienpädagogischen Reflexion	Khaddage und Lattemann 2021	MINT (Informatik)/NetLogo

Tab. 3: Beispielreferenzen aus dem Studienkorpus.

4. Diskussion

In diesem Beitrag wurden Prozesse zur systematischen Generierung eines Studienkorpus im Kontext KI-gestützter Lehr-Lern-Situationen auf der Mikroebene schulischen Unterrichts bzw. hochschulischer Lehre dargestellt. Der konzeptuelle Rahmen beinhaltet zum einen ein Verständnis von KI vor dem Hintergrund der aktuellen technologischen Entwicklungen. Zum anderen stellt eine interdisziplinäre, aber technologisch dominierte Bearbeitung von Fragestellungen zu *Lernen mit KI-gestützten Bildungstechnologien* erziehungswissenschaftlich Forschende vor Herausforderungen. Im Sinne einer konfigurativen Forschungslogik sollte eine repräsentative Breite von Studien entstehen, die das Merkmal der grösstmöglichen Berücksichtigung der Komplexität pädagogischer Praxis aufwiesen. Im weiteren Verlauf wird sich eine Synthese des Studienkorpus (Thomas et al. 2017) anschliessen.

Darin werden Szenarien KI-gestützten Lehrens und Lernens anhand manifester und ggf. latenter Merkmale abgeleitet und in Form eines Überblicks zu aktuellen Entwicklungen KI-gestützten Lehrens und Lernens systematisiert.

Das forschungssynthetische Vorgehen umfasste verschiedene Schritte. Diese wurden für einen interdisziplinären, von technischen Perspektiven geprägten Kontext illustriert: die Entwicklung einer Suchstrategie mit vier Datenbanken, ein zwei Konzepte umfassender Suchstring sowie In- und Exklusionskriterien, mit denen die relevanten Treffer in zwei Kodiervorgängen (Abstract und Titel, Volltext) ausgewählt wurden. Kennwerte der *Präzision* und *Sensitivität* verdeutlichten, wie gut es mit der Suchstrategie gelang, aus den einbezogenen Datenbanken relevante Treffer zu generieren. Diese Vorgehensweise ist auf zukünftige Forschungssynthesen im Kontext des Lernens mit digitalen Medien gut übertragbar. Besonders der konzeptuelle Fokus des generierten Studienkorpus ist anschlussfähig. Er zeichnet sich durch das Merkmal der grösstmöglichen Berücksichtigung der Komplexität pädagogischer Praxis aus. Die entwickelte Suchstrategie lässt sich adaptieren. Dafür würde die Konzeptgruppe «Künstliche Intelligenz» durch die im Fokus der Zielstellung stehende Bildungstechnologie ersetzt und mit dem übergreifenden Konzept der «Adressatengruppe» kombiniert. Mit dem Inklusionskriterium «KI-basiertes Setting/Einsatz einer KI-basierten Anwendung» könnte man in analoger Weise verfahren.

Die Balance zwischen Präzision und Sensitivität zu finden, ist im Rahmen von Forschungssynthesen eine Herausforderung bei der Entwicklung von Suchstrategien. Zielt eine Fragestellung im Sinne einer aggregativen Methodologie darauf ab, möglichst alle existierenden relevanten Treffer zu finden, ist eine maximal sensitive Suche erforderlich (wie bspw. in Cochrane Reviews oder Reviews der Campbell Collaboration praktiziert). Für das vorliegende Studienkorpus sollten Studien möglichst präzise erfasst werden. Im Vordergrund stand, eine repräsentative Bandbreite an Fällen zu erzeugen, in denen sich eine Vielfalt von «real-world educational settings» für Lernen mit KI-gestützten Bildungstechnologien zeigte. In der schrittweisen Entwicklung der Suchstrategie wurde klar, wie solche Studien identifiziert werden konnten. Aus qualitativen Screenings der generierten Treffer wurden Schritt für Schritt In- und Exklusionskriterien spezifiziert. Der finale Suchstring enthielt die beiden Konzepte «Adressaten in Lehr-Lern-Settings» und «Künstliche Intelligenz». Obwohl die verwendeten Suchoperatoren sehr breit angelegt waren, führten sie in der Kombination der beiden Konzepte zu Präzisionskennwerten zwischen 22 und 45% in den vier durchsuchten Datenbanken (Screening I, Abstract und Titel). Dies stützt die Aussage, dass Studien im Kontext von Lehren und Lernen mit KI-gestützten Bildungstechnologien, die möglichst nah an die Komplexität von Lehr-Lern-Situationen in der pädagogischen Praxis heranreichen, insgesamt selten sind (Verdu et al. 2014; Misiejuk und Wasson 2017, 61; Zawacki-Richter et al. 2019, 26; Kabudi, Pappas, und Olsen 2021).

Die Präzision reduzierte sich nach dem Screening der Volltexte noch einmal deutlich (zwei bis 16%). Aus zwei Datenbanken (*FIS-Bildung*, *LearntechLib*) fanden nur noch ein respektive zwei Treffer Eingang in den Studienkorpus. Die Herausforderung einer technologisch geprägten Interdisziplinarität für erziehungswissenschaftliche Blickwinkel wurde im Screening der Volltexte sichtbar. Die Beschreibung des Einsatzes einer KI-gestützten Anwendung in den Abstracts enthielt oft wenige Informationen zum Inklusionskriterium «Bezug zu Mikroebene Lehr-Lern-Situation im Interaktionsgeschehen von Lehrperson/Lernende». Eine Inklusion im ersten Screening (Abstract, Titel) erfolgte immer auch dann, wenn die Erprobung der KI-gestützten Anwendung im Kontext einer Lehr-Lern-Situation aus erziehungswissenschaftlicher Perspektive grundsätzlich plausibel war.

Die Grundgesamtheit aller existierenden Treffer für ein Studienkorpus ist nur theoretisch bestimmbar und wird operationalisiert über die Auswahl der Datenquellen. Vier Datenbanken mit unterschiedlichen, für den konzeptuellen Fokus der Zielstellung geeigneten Schwerpunkten wurden einbezogen (*ERIC*, *FIS Bildung*, *Web of Science*, *LearntecLib*). Die wenigen Dubletten (7) zeigten, dass alle einbezogenen Datenbanken für das Studienkorpus relevante Treffer in Form von «unique hits» (Best et al. 2014, 353) generierten. *Web of Science* wies die höchste Sensitivität (80%) auf. Die Kennwerte der Präzision und Sensitivität wurden nach dem ersten und zweiten Screening berechnet. Grundsätzlich wäre auch eine Berechnung der Kennwerte in der iterativen Entwicklung eines Suchstrings interessant. Dann könnten bspw. Aussagen quantifiziert werden, wie sich das Hinzufügen, Weglassen oder die Neukombination von Konzeptgruppen in Suchstrings oder die Aufnahme weiterer Datenquellen auf die Balance von Präzision und Sensitivität auswirkt. Dazu müssten die generierten Treffer in jeder iterativen Suchschleife kodiert werden. Dies war für den vorliegenden Beitrag weder beabsichtigt noch möglich. Der iterative Entwicklungsprozess der Suchstrategie war erforderlich, um In- und Exklusionskriterien final zu spezifizieren («iterative review», Brunton et al. 2017, 102). Ausgehend vom nun vorliegenden Suchstring und Kodierschema könnten weitere Datenquellen (bspw. einschlägige Journals im Kontext *AIED*; Feng und Law 2021) erprobt werden.

Die Analyse der Zusammensetzung des Studienkorpus zeigt, dass 51 Studien aus 19 Ländern, innerhalb 8 verschiedener fachlicher Kontexte und unter Verwendung der KI-Technologien Natural Language Processing, maschinelles Lernen (überwacht, unüberwacht, tiefes Lernen), Computer Vision und Robotics enthalten sind. Induktiv wurden vier Grob-Kategorien nach der Funktion der KI-Unterstützung für Lehr-Lern-Situationen abgeleitet (Assessment, Partner für Kommunikation und Demonstration, KI-Erprobung zur medienpädagogischen Reflexion, Scaffolding). Durch das Screening der Volltexte entstanden zur inhaltlichen Validierung des Kodierprozesses Beschreibungen zur Lehr-Lern-Situation im Sinne der Inklusionskriterien. Im Prozess der Generierung der Textpassagen wurde deutlich, dass in überwiegender

Zahl Lehrpersonen und Lernende jeweils für sich im Interaktionsgeschehen mit KI-gestützten Systemen agieren. Die konzeptuellen Überlegungen hybrider Ansätze im Kontext *AIED* (Molenaar 2022; Holstein, Alevin, und Rummel 2020), in denen die Rollen der verschiedenen Stakeholder eines Lehr-Lern-Geschehens ineinandergreifen, werden nicht adressiert.

Literatur

- Alammary, Ali. 2021. «LOsMonitor: A Machine Learning Tool for Analyzing and Monitoring Cognitive Levels of Assessment Questions». *IEEE Trans. Learning Technol.* 14 (5): 640–52. <https://doi.org/10.1109/TLT.2021.3116952>.
- Alevin, Vincent, Elizabeth A. McLaughlin, Amos R. Glenn, und Kenneth Koedinger. 2016. «Instruction Based on Adaptive Learning Technologies». In *Handbook of Research on Learning and Instruction*, herausgegeben von Richard E. Mayer, und Patricia A. Alexander. 2nd ed., 522–60. Educational Psychology Handbook Ser. Florence: Taylor and Francis.
- Augustsson, Gunnar, und Lena Boström. 2016. «Teachers' Leadership in the Didactic Room: A Systematic Literature Review of International Research». *ADNO* 10 (3): 7. <https://doi.org/10.5617/adno.2883>.
- Bartolomé, Antonio, Linda Castañeda, und Jordi Adell. 2018. «Personalisation in educational technology: the absence of underlying pedagogies». *Int J Educ Technol High Educ* 15. <https://doi.org/10.1186/s41239-018-0095-0>.
- Bedenlier, Svenja, Melissa Bond, Katja Buntins, Olaf Zawacki-Richter, und Michael Kerres. 2020. «Learning by Doing? Reflections on Conducting a Systematic Review in the Field of Educational Technology». In *Systematic Reviews in Educational Research*, herausgegeben von Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond und Katja Buntins, 111–27. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27602-7_7.
- Bernacki, Matthew L., Meghan J. Greene, und Nikki G. Lobczowski. 2021. «A Systematic Review of Research on Personalized Learning: Personalized by Whom, to What, How, and for What Purpose(s)?». *Educ Psychol Rev* 33 (4): 1675–1715. <https://doi.org/10.1007/s10648-021-09615-8>.
- Best, Paul, Brian Taylor, Roger Manktelow, und Janice McQuilkin. 2014. «Systematically retrieving research in the digital age: Case study on the topic of social networking sites and young people's mental health». *Journal of Information Science* 40 (3): 346–56. <https://doi.org/10.1177/0165551514521936>.
- Bond, Melissa, Katja Buntins, Svenja Bedenlier, Olaf Zawacki-Richter, und Michael Kerres. 2020. «Mapping research in student engagement and educational technology in higher education: a systematic evidence map». *Int J Educ Technol High Educ* 17. <https://doi.org/10.1186/s41239-019-0176-8>.

- Brunton, Ginny, Claire Stansfield, Jenny Caird, und James Thomas. 2017. „Finding Relevant Studies.“ In *An Introduction to Systematic Reviews*, herausgegeben von David Gough, Sandy Oliver und James Thomas. 2nd edition, 93–122. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.
- Buiu, Catalin. 1999. «Artificial Intelligence in education: state of the art and perspectives». *ZIFF-Papiere* (111). <https://nbn-resolving.org/urn:nbn:de:hbz:708-dh1799>.
- Campbell, Allison, Brian Taylor, Jessica Bates, und Una O'Connor-Bones. 2018. «Developing and Applying a Protocol for a Systematic Review in the Social Sciences». *New Review of Academic Librarianship* 24: 1–22. <https://doi.org/10.1080/13614533.2017.1281827>.
- Chen, Li, Dirk Ifenthaler, und Jane Yin-Kim Yau. 2021. «Online and blended entrepreneurship education: a systematic review of applied educational technologies». *Entrep Educ* 4 (2): 191–232. <https://doi.org/10.1007/s41959-021-00047-7>.
- Cheng, Y. P., S. C. Cheng, und Y. M. Huang. 2022. «An Internet Articles Retrieval Agent Combined With Dynamic Associative Concept Maps to Implement Online Learning in an Artificial Intelligence Course». *International Review of Research in Open and Distributed Learning* 23.
- Cui, Kangxu. 2022. «Artificial intelligence and creativity: piano teaching with augmented reality applications». *Interactive Learning Environments*, 1–12. <https://doi.org/10.1080/10494820.2022.2059520>.
- Da Santos, Cristina Mamédio Costa, Cibele Andruccioli de Mattos Pimenta, und Moacyr Roberto Cuce Nobre. 2007. «The PICO Strategy for the Research Question Construction and Evidence Search». *Revista latino-americana de enfermagem* 15 (3): 508–11. <https://doi.org/10.1590/s0104-11692007000300023>.
- Feng, Shihui, und Nancy Law. 2021. «Mapping Artificial Intelligence in Education Research: a Network-based Keyword Analysis». *Int J Artif Intell Educ* 31 (2): 277–303. <https://doi.org/10.1007/s40593-021-00244-4>.
- Gough, David, Sandy Oliver, und James Thomas, Hrsg. 2017. *An Introduction to Systematic Reviews*. 2nd edition. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.
- Gough, David, James Thomas, und Sandy Oliver. 2012. «Clarifying Differences Between Review Designs and Methods». *Systematic reviews* 1: 28. <https://doi.org/10.1186/2046-4053-1-28>.
- Grant, Maria J., und Andrew Booth. 2009. «A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies». *Health information and libraries journal* 26 (2): 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>.
- Guan, Chong, Jian Mou, und Zhiying Jiang. 2020. «Artificial intelligence innovation in education: A twenty-year data-driven historical analysis». *International Journal of Innovation Studies* 4 (4): 134–47. <https://doi.org/10.1016/j.ijis.2020.09.001>.
- Hamisch, Katharina, und Robert Kruschel. 2022. «Zwischen Individualisierungsversprechen und Vermessungsgefahr. Die Rolle der Schlüsseltechnologie Künstliche Intelligenz in der inklusiven Schule». <https://doi.org/10.25656/01:23821>.

- Hege, Inga, Isabel Kiesewetter, und Martin Adler. 2020. «Automatic Analysis of Summary Statements in Virtual Patients – a Pilot Study Evaluating a Machine Learning Approach». *BMC medical education* 20: 366. <https://doi.org/10.1186/s12909-020-02297-w>.
- Hew, Khe Foon, Min Lan, Ying Tang, Chengyuan Jia und Chung Kwan Lo. 2019. «Where is the <theory> within the field of educational technology research?». *Br J Educ Technol* 50 (3): 956–71. <https://doi.org/10.1111/bjet.12770>.
- Holstein, Kenneth, und Vincent Alevan. 2022. «Designing for human – AI complementarity in K-12 education». *AI Magazine* 43 (2): 239–48. <https://doi.org/10.1002/aaai.12058>.
- Holstein, Kenneth, Vincent Alevan, und Nikol Rummel. 2020. «A Conceptual Framework for Human–AI Hybrid Adaptivity in Education», 240–54. https://doi.org/10.1007/978-3-030-52237-7_20.
- Hunte, Melissa R., Samantha McCormick, Maitree Shah, Clarissa Lau, und Eunice Eunhee Jang. 2021. «Investigating the potential of NLP-driven linguistic and acoustic features for predicting human scores of children’s oral language proficiency». *Assessment in Education: Principles, Policy & Practice* 28 (4): 477–505. <https://doi.org/10.1080/0969594X.2021.1999209>.
- Hwang, Gwo-Jen, Haoran Xie, Benjamin W. Wah, und Dragan Gašević. 2020. «Vision, challenges, roles and research issues of Artificial Intelligence in Education». *Computers and Education: Artificial Intelligence* 1:100001. <https://doi.org/10.1016/j.caeai.2020.100001>.
- Kabudi, Tumaini, Ilias Pappas, und Dag Håkon Olsen. 2021. «AI-enabled adaptive learning systems: A systematic mapping of the literature». *Computers and Education: Artificial Intelligence* 2:1–12. <https://doi.org/10.1016/j.caeai.2021.100017>.
- Kahn, Ken, und Niall Winters. 2021. «Constructionism and AI: A history and possible futures». *Br J Educ Technol* 52 (3): 1130–42. <https://doi.org/10.1111/bjet.13088>.
- Kerres, Michael, Katja Buntins, J. Buchner, Hendrik Drachslers, und Olaf Zawacki-Richter. 2023 im Druck. «Adaptive und künstliche-intelligente Lernprogramme: Eine kritische Analyse aus Sicht der Mediendidaktik». In *Künstliche Intelligenz in der Bildung*. herausgegeben von de Witt, Claudia, Christina Gloerfeld und Silke Elisabeth Wrede. Wiesbaden: Springer VS. <https://learninglab.uni-due.de/publikationen/14147>.
- Khaddage, Ferial, und Christoph Lattemann. 2021. «Artificial Intelligence and Cloud-based Technologies to Empower Learning <Active Experiments via NetLogo>». In *EdMedia + Innovate Learning 2021*, herausgegeben von Theo J. Bastiaens, 15–21. United States: Association for the Advancement of Computing in Education (AACE). <https://www.learntechlib.org/p/219633>.
- Liabo, Kristin, David Gough, und Angela Harden. 2017. «Developing Justifiable Evidence Claims». In *An Introduction to Systematic Reviews*, herausgegeben von David Gough, Sandy Oliver und James Thomas. 2nd edition, 252–77. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.

- Lipowsky, Frank, und Miriam Lotz. 2015. «Ist Individualisierung der Königsweg zum erfolgreichen Lernen? Eine Auseinandersetzung mit Theorien, Konzepten und empirischen Befunden». In *Begabungen entwickeln & Kreativität fördern*, herausgegeben von Mehlhorn, Gerlinde, Karola Schöppe und Frank Schulz, 155–219. KREApplus Band 8. München: kopaed.
- McFadden, Paula, Brian J. Taylor, Anne Campbell, und Janice McQuilkin. 2012. «Systematically Identifying Relevant Research». *Research on Social Work Practice* 22 (6): 626–36. <https://doi.org/10.1177/1049731512453209>.
- Meissner, Sibylle; Anna-Katharina Praetorius, Katrin Rakoczy, Katrin Kleinknecht, und Marucs Pietsch. 2020. «Differenzielle Effekte der Unterrichtsqualität auf die aktive Lernzeit». In *Empirische Forschung zu Unterrichtsqualität: Theoretische Grundfragen und quantitative Modellierungen*, herausgegeben von Anna-Katharina Praetorius, Juliane Grünkorn, und Eckhard Klieme, 81–94. Zeitschrift für Pädagogik Beiheft 66. Beiheft (April 2020). Weinheim, Basel: Beltz Juventa.
- Misiejuk, Kamila, und Barbara Wasson. 2017. *State of the Field Report on Learning Analytics*. Bergen: Centre for the Science of Learning & Technology.
- Molenaar, Inge. 2022. «The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning». *Computers and Education: Artificial Intelligence* 3: 100070. <https://doi.org/10.1016/j.caeai.2022.100070>.
- Mubarak, Ahmed A., Han Cao, und Weizhen Zhang. 2022. «Prediction of students' early dropout based on their interaction logs in online learning environment». *Interactive Learning Environments* 30 (8): 1414–33. <https://doi.org/10.1080/10494820.2020.1727529>.
- Newman, Mark, und David Gough. 2020. «Systematic Reviews in Educational Research: Methodology, Perspectives and Application». In *Systematic Reviews in Educational Research*, herausgegeben von Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond und Katja Buntins, 3–22. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-27602-7>.
- Niegemann, Helmut, und Steffi Heidig. 2020. «Interaktivität und Adaptivität in multimedialen Lernumgebungen». In *Handbuch Bildungstechnologie: Konzeption und Einsatz digitaler Lernumgebungen*, herausgegeben von Helmut Niegemann und Armin Weinberger, 343–67. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-54368-9_33.
- Niegemann, Helmut, und Armin Weinberger, Hrsg. 2020. *Handbuch Bildungstechnologie: Konzeption und Einsatz digitaler Lernumgebungen*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-54368-9>.
- Niegemann, Helmut, und Armin Weinberger. 2020. «Was ist Bildungstechnologie?» In *Handbuch Bildungstechnologie: Konzeption und Einsatz digitaler Lernumgebungen*, herausgegeben von Helmut Niegemann und Armin Weinberger, 3–16. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-54368-9_1.
- Niegemann, Helmut M., Steffi Domagk, Silviav Hessel, Alexandra Hein, Matthias Hupfer, und Annett Zobel. 2008. *Kompodium multimediales Lernen*. X.media.press. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-540-37226-4>.

- OECD. 2021. «OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots». <https://doi.org/10.1787/589b283f-en>.
- O'Mara-Eves, Alison, Ginny Brunton, David McDaid, Josephine Kavanagh, Sandy Oliver, und James Thomas. 2014. «Techniques for Identifying Cross-Disciplinary and «Hard-to-Detect» Evidence for Systematic Review». *Research synthesis methods* 5: 50–59. <https://doi.org/10.1002/jrsm.1094>.
- Paass, Gerhard, und Dirk Hecker. 2020. *Künstliche Intelligenz: Was steckt hinter der Technologie der Zukunft?* Wiesbaden, Heidelberg: Springer Vieweg. <https://doi.org/10.1007/978-3-658-30211-5>.
- Pajares, M. Frank. 1992. «Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct». *Review of Educational Research* 62 (3): 307–32. <https://doi.org/10.3102/00346543062003307>.
- Patton, M. 1990. *Qualitative evaluation and Research methods*. Beverly Hills: CA: Sage Publications.
- Rabenstein, Kerstin, Matthias Proske, und Till-Sebastian Idel. 2018. «Individualisierung schulischen Lehrens und Lernens als Reformstrategie». *Zeitschrift für Pädagogik* 64 (2): 147–58.
- Sandelowski, Margarete, Corrine I. Voils, Jennifer Leeman, und Jamie L. Crandell. 2012. «Mapping the Mixed Methods-Mixed Research Synthesis Terrain». *Journal of mixed methods research* 6 (4): 317–31. <https://doi.org/10.1177/1558689811427913>.
- Schiefner-Rohs, Mandy, und Stefan Aufenanger. 2021. «Vorbereiten auf die neue Welt. Künstliche Intelligenz und Schule». *on – Lernen in der digitalen Welt* (5): 4–7.
- Sejnowski, Terrence J. 2018. *The Deep Learning Revolution*. Cambridge, London: The MIT Press.
- Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser et al. 2016. «Mastering the Game of Go with Deep Neural Networks and Tree Search». *Nature* 529 (7587): 484–89. <https://doi.org/10.1038/nature16961>.
- Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert et al. 2017. «Mastering the Game of Go Without Human Knowledge». *Nature* 550 (7676): 354–59. <https://doi.org/10.1038/nature24270>.
- Tai, Joanna, Rola Ajjawi, Margaret Bearman, und Paul Wiseman. 2020. «Conceptualizations and Measures of Student Engagement: A Worked Example of Systematic Review». In *Systematic Reviews in Educational Research*, herausgegeben von Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond und Katja Buntins, 91–110. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-27602-7>.
- Thomas, James, Alison O'Mara-Eves, Angela Harden, und Mark Newman. 2017. «Synthesis Methods for Combining and Configuring Textual or Mixed Methods Data». In *An Introduction to Systematic Reviews*, herausgegeben von Oliver Gough, Sandy Oliver, und James Thomas, 181–209. London: Sage.

- Thomas, James, Alison O'Mara-Eves, Angela Harden, und Mark Newman. 2017. «Synthesis Methods for Combining and Configuring Textual or Mixed Methods Data». In *An Introduction to Systematic Reviews*, herausgegeben von David Gough, Sandy Oliver, und James Thomas. 2nd edition, 181–209. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.
- Verdu, E., L. M. Regueras, M. J. Verdu, J. P. de Castro, D. Kohen-Vacs, E. Gal und M. Ronen. 2014. «Intelligent tutoring interface for technology enhanced learning in a course of computer network design». *IEEE Frontiers in Education Conference (FIE) Proceedings*, 1–7. <https://doi.org/10.1109/FIE.2014.7044139>.
- Vieluf, Svenja, Anna-Katharina Praetorius, Katrin Rakoczy, Marc Kleinknecht, und Marcus Pietsch. 2020. «Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts. Ein kritischer Vergleich verschiedener Modellvarianten». In *Empirische Forschung zu Unterrichtsqualität: Theoretische Grundfragen und quantitative Modellierungen*, herausgegeben von Anna-Katharina Praetorius, Juliane Grünkorn, und Eckhard Klieme, 63–80. Zeitschrift für Pädagogik Beiheft 66. Beiheft (April 2020). Weinheim, Basel: Beltz Juventa.
- Walkington, Candace, und Matthew L. Bernacki. 2020. «Appraising research on personalized learning: Definitions, theoretical alignment, advancements, and future directions». *Journal of Research on Technology in Education* 52 (3): 235–52. <https://doi.org/10.1080/15391523.2020.1747757>.
- Yang, Hyejin, Heyoung Kim, Jang Ho Lee, und Dongkwang Shin. 2022. «Implementation of an AI chatbot as an English conversation partner in EFL speaking classes». *ReCALL* 34 (3): 327–43. <https://doi.org/10.1017/S0958344022000039>.
- Zawacki-Richter, Olaf, Michael Kerres, Svenja Bedenlier, Melissa Bond, und Katja Buntins, Hrsg. 2020. *Systematic Reviews in Educational Research*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-27602-7>.
- Zawacki-Richter, Olaf, Victoria I. Marín, Melissa Bond, und Franziska Gouverneur. 2019. «Systematic review of research on artificial intelligence applications in higher education – where are the educators?». *Int J Educ Technol High Educ* 16. <https://doi.org/10.1186/s41239-019-0171-0>.