



54

MedienPädagogik

Zeitschrift für Theorie und Praxis der Medienbildung

**Forschungssynthesen in der Mediendidaktik.
Ansätze und Herausforderungen**

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

Themenheft Nr. 54

Forschungssynthesen in der Mediendidaktik

Ansätze und Herausforderungen

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

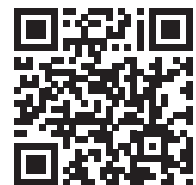


Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Titel: Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen
Herausgebende: Svenja Bedenlier, Katja Buntins, Annika Wilmers und Michael Kerres
Cover & Design: Klaus Rummler
Produktion: Klaus Rummler
Verlag: *OAPublishing Collective Genossenschaft* für die Zeitschrift MedienPädagogik, hrsg. durch die Sektion Medienpädagogik (DGfE)
Herstellung: Books on Demand GmbH, Norderstedt, Deutschland
Reihe: Themenhefte
Nummer: 54

ISBN (print): 978-3-03978-009-9
ISBN (online): 978-3-03978-064-8
DOI-URL: <https://doi.org/10.21240/mpaed/54.X>
ISSN: 1424-3636



© Zürich, August 14, 2023. Creative Commons Attribution 4.0 International (CC BY 4.0), alle Rechte liegen bei den Autor:innen

Das Werk und jeder seiner Beiträge, sind urheberrechtlich geschützt. Sie dürfen das Material in jedwedem Format oder Medium vervielfältigen und weiterverbreiten, das Material remixen, verändern und darauf aufbauen und zwar für beliebige Zwecke. Unter folgenden Bedingungen: Namensnennung – Sie müssen angemessene Urheber- und Rechteangaben machen, einen Link zur Lizenz einschl. Original-DOI beifügen und angeben, ob Änderungen vorgenommen wurden. Diese Angaben müssen den üblichen wissenschaftlichen Zitierformaten folgen.

Diese Publikation wurde unterstützt durch die Open-Access-Publikationsfonds der Universität Duisburg-Essen und der Friedrich-Alexander-Universität Erlangen-Nürnberg.

Inhalt

Editorial: Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen Svenja Bedenlier, Katja Buntins, Annika Wilmers und Michael Kerres	i
Zwischen Präzision und Sensitivität. Generierung eines Studienkorpus am Beispiel einer Fragestellung zu Künstlicher Intelligenz (KI) in Bildungsprozessen Verena Köstler	1
Forschungssynthesen und der Researcher-Practitioner-Gap. Ein Diskussionsbeitrag zu Potenzialen und Grenzen anhand einer Forschungssynthese aus der Medienpädagogik Simon Munk, Lisa Ziernwald, Kaley Lesperance und Doris Holzberger	28
Publication Status as a Common Inclusion Criterion in Systematic Reviews. On the Use of Digital Media in the Context of Teaching Pupils with SEN in Inclusive, Integrative or Segregated Settings Claudia Mertens, Carolin Quenzer-Alfred, Anna-Maria Kamin und Daniel Mays	51
Potenziale und Herausforderungen von Forschungssynthesen im Kontext der Digitalisierung im Bildungsbereich am Beispiel des Critical Reviews Anna Heinemann, Jan Koschorreck, Annika Wilmers, Pia Sander und Jens Leber	79
Cognitive Biases in Screening Processes – Search Strategies in Educational Technology Research. A Systematic Review on Learning with Virtual Reality Katja Buntins, Miriam Mulders, und Nadine Schröder	103
Was meinst du eigentlich, wenn du von «Feedback» sprichst? Chancen und Grenzen qualitativ-inhaltsanalytischer Scope-Reviews zur Herausarbeitung von Taxonomien zur Beschreibung didaktischer Szenarien am Beispiel «Feedback» Annika Brück-Hübner und Jennifer Schluer	125
Methodological Approaches to Evidence Synthesis in Educational Technology. A Tertiary Systematic Mapping Review Katja Buntins, Svenja Bedenlier, Victoria Marín, Marion Händel und Melissa Bond	167





Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

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

Editorial: Forschungssynthesen in der Mediendidaktik

Ansätze und Herausforderungen

Svenja Bedenlier¹ , Katja Buntins² , Annika Wilmers³  und Michael Kerres² 

¹ Friedrich-Alexander-Universität Erlangen-Nürnberg

² Universität Duisburg-Essen

³ DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Einleitung

Die Herbsttagung 2023 der Sektion Medienpädagogik in der Gesellschaft für Erziehungswissenschaft steht unter dem Tagungsthema «Mit Medienpädagogik in die Zukunft. Entwürfe, Begründungen und (inter-)disziplinäre Begegnungen» und fragt dabei auch nach den methodologischen und methodischen Zugängen zum Forschungsfeld (Hartung et al. 2014). Vor dem Hintergrund der internationalen Forschungsstränge zu *Digital Education* bzw. *Educational Technology* und dem gesellschaftlichen Bedarf nach evidenzbasierter bzw. -informierter Bildungspolitik zu Fragen der Digitalisierung ist die lange Zeit primär theoretisch angelegte Diskussion im deutschsprachigen Raum zunehmend durch eine empirische Perspektive erweitert worden (Buntins et al. 2018). Dabei kommen sowohl qualitative, quantitative als auch mixed methods Ansätze zum Einsatz, und es stellt sich die Frage, wie die Vielzahl der vorliegenden Befunde systematisch erfasst, synthetisiert und für weitere Forschung nutzbar gemacht werden können. Forschungssynthesen gehen über einfache Literaturlauswertungen hinaus und basieren in ihren verschiedenen Ausprägungen auf systematischen, d.h. intersubjektiv nachvollziehbaren Verfahren (Sutton et al. 2019). Sie werden zunehmend auch in der medienpädagogischen Forschung herangezogen und finden Beachtung in der Bildungspraxis und -politik (s.a. Delere 2020; Müller et al. 2022; Schmidt 2023).

Die historische und disziplinäre Entwicklung von Forschungssynthesen – und hier im Besonderen in der Form des Systematic Review – stellt jedoch eine besondere Situation dar. Mit ihrem Ursprung in der medizinischen und pharmakologischen Forschung der 1970er Jahre wurden Systematic Reviews in den 1990er Jahren verstärkt auch im Kontext von Public Policy aufgegriffen und fanden Eingang in die Sozialwissenschaften (Oakley et al. 2005). Der Einsatz von Forschungssynthesen in diesem Feld wurde begleitet von einer vertiefenden Diskussion zu ihrer methodischen Anlage (Ades et al. 2005; Afshari et al. 2017; Bohlin 2012; Borrego et al. 2014; Chen und Tseng 2011; Esteves et al. 2017; Petticrew 2003) sowie der Erweiterung um qualitative Zugänge und komplexerer Ansätze, die die bislang vorherrschenden quantitativen Verfahren ergänzen (Grant und Booth 2009).

Forschungssynthesen zum Lernen mit digitalen Medien und zum Feld der Educational Technology sind hiervon ebenso betroffen, da sich zum ersten sogenannte «Goldstandards», die in den Ursprungsdisziplinen von Forschungssynthesen angelegt werden können, nicht problemlos auf (medien)pädagogische und (medien) didaktische Fragestellungen und Forschungsarbeiten anwenden lassen. Zum zweiten sind es komplexe Interaktionen und Konditionen unter denen Auswirkungen bildungsbezogener Interventionen entstehen, was einfache Wertungen ausschließt (Berliner 2002; Hammersley 2020). Angesichts dieser Gemengelage und der zunehmenden Verwendung und aktueller Prominenz von Forschungssynthesen in der Forschung zu *Educational Technology* (Kimmons und Rosenberg 2022) stellen sich somit Fragen nach Grenzen und Potenzialen dieses Ansatzes.

Als Folge bedarf es im Kontext einer stärker forschungsmethodologisch und -methodisch ausgerichteten Debatte einer kritischen Bestandsaufnahme der Anlage und Ausgestaltung von Forschungssynthesen in diesem Themenfeld mit Bezug auf die spezifischen Fragen der medienerzieherischen bzw. -didaktischen Forschung. Dies schließt die Fragen ein, wie Forschungssynthesen auf den Kontext der Disziplin, ihres Forschungsgegenstandes und ihrer Traditionen anzupassen sind (Zawacki-Richter et al. 2020). Das vorliegende Themenheft gibt einen Einblick in die Vielzahl der aktuell diskutierten Fragestellungen, die gleichermaßen die methodische Ausgestaltung von Arbeitsprozessen und Formaten, inhaltlicher Spezifika im Themenfeld Digitalisierung und Medienpädagogik oder auch die Darstellung und den Transfer von synthetisierten Forschungsbefunden betreffen. Das Themenheft möchte damit einen Beitrag leisten, um den Ansatz der Forschungssynthesen mit Blick auf medienpädagogische und mediendidaktische Fragestellungen zu reflektieren und Denkanstöße für ihre Weiterentwicklung zu setzen.

Übersicht über die Beiträge

Das Themenheft vereint sieben Beiträge, die sich in unterschiedlicher Weise mit methodologisch-methodischen Herausforderungen von Forschungssynthesen im medienpädagogischen und -didaktischen Bereich auseinandersetzen. Hierbei wird teilweise direkt anhand eines konkreten medienpädagogischen oder mediendidaktischen Themas auf besondere Herausforderungen verwiesen (z. B. der Beitrag von Buntins und Mulders 2023), teilweise wird jedoch auch ein konkretes methodologisch-methodisches Problem diskutiert und an einem medienpädagogischen Beispiel illustriert (z. B. der Beitrag von Munk, Ziernwald, Lesperance, und Holzberger 2023). In der Gesamtschau verweisen die Beiträge des Themenheftes auf eine große Diversität an identifizierten Herausforderungen, denen Forschungssynthesen bildungswissenschaftlicher Forschung in methodologisch-methodischer Hinsicht Rechnung tragen sollten, um ihren untersuchten Gegenstand angemessen zu adressieren und einen distinkten methodischen Zugang darstellen zu können.

Der Diskussionsbeitrag von **Munk, Ziernwald, Lesperance und Holzberger (2023)** greift die besondere Problematik des Researcher-Practitioner-Gaps auf, der gerade dann an Bedeutung gewinnt, wenn eine zunehmend gewünschte evidenzbasierte Bildungspraxis proklamiert wird. Gleichzeitig gestaltet sich der Transfer von Forschungsergebnissen in die Praxis komplex. Im Beitrag wird dies konkret mit Bezug auf die Übertragung wissenschaftlicher Erkenntnisse in den Bildungskontext Schule gezeigt und an einem medienpädagogischen Beispiel illustriert. Der Beitrag verweist auf die Möglichkeiten und Grenzen, die sich durch Forschungssynthesen für den Erkenntnistransfer ergeben und den Researcher-Practitioner-Gap fokussieren.

Buntins, Mulders und Schröder (2023) illustrieren in ihrem Beitrag zu Suchstrategien in Forschungssynthesen am Beispiel des Themas Lernen mit Virtual Reality, welche grundsätzlichen Herausforderungen sich nicht zuletzt bezüglich des cognitive loads der Reviewer:innen durch den Einsatz unterschiedlicher Suchstrategien ergeben. Entlang der Parameter Präzision, Sensitivität und Genauigkeit zeigen sie für zwei Suchen mit den Operatoren AND und OR deren jeweilige Möglichkeiten und Grenzen auf. Implikationen ergeben sich vor allem für die weitere Forschung zur kognitiven Beanspruchung und Güte im Reviewprozess.

Der Beitrag von **Mertens, Quenzer-Alfred, Kamin und Mays** (2023) beschäftigt sich mit der Frage, inwiefern die Festlegung auf bestimmte Publikationsformen als Ein- beziehungsweise Ausschlusskriterium sinnvoll ist. Dabei wird überprüft, ob sich die finale Literaturlauswahl einer Forschungssynthese zum Einsatz digitaler Medien in inklusiven, integrativen und exkludierenden Unterrichtssettings verändert, wenn graue Literatur und Literatur ohne Peer-Review nicht von vornherein ausgeschlossen werden. Eine solche Vorauswahl orientiert sich am Vorgehen von Forschungssynthesen in der Medizin, muss aber – wie der Beitrag zeigt – in einem Fach mit anderen Forschungstraditionen und fachlichen Gegebenheiten nicht gleichermassen sinnvoll sein.

Buntins, Bedenlier, Marín, Händel und Bond (2023) gehen in ihrem Beitrag der Frage nach, wie transparent und reproduzierbar verschiedene Formen von Forschungssynthesen ihr methodisches Vorgehen in ihrer Publikation dokumentieren. Die Autor:innen argumentieren, dass Transparenz und Replizierbarkeit nicht nur für das Verständnis des Reviews wichtig sind, sondern auch für mögliche Änderungen oder Erneuerungen des Reviews. Zu diesem Zweck kodieren sie die Dokumentation in den verschiedenen Schritten der Synthese und vergleichen Metaanalysen, Systematic Reviews und Literaturreviews. Sie stellen fest, dass in fast allen untersuchten Synthesen ein grosses Transparenzdefizit besteht, das bei Literaturreviews deutlich grösser ist als bei den beiden anderen Reviewtypen. Die Arbeit der Autor:innen zeigt, dass ein methodischer Diskurs, vor allem aber Standards für die Erstellung und Verschriftlichung von Reviews notwendig sind, um forschungsmethodischen Qualitätsstandards zu genügen.

Der Beitrag von **Köstler** (2023) nimmt sich am Beispiel von Künstlicher Intelligenz im Bildungsbereich (AIED) der Frage an, wie die Abdeckung verschiedener Datenbanken in diesem speziellen Forschungsfeld aussieht. Sie untersucht die Präzision und Sensitivität der Datenbanken FIS Bildung, LearntechLib, Web of Science und ERIC. Sensitivität beschreibt, wie gut es gelingt, eine möglichst grosse Anzahl relevanter Artikel aufzufinden. Es geht darum, keine wichtigen Informationen zu übersehen. Präzision hingegen konzentriert sich auf die gezielte Suche nach relevanten Dokumenten, ohne dabei eine Flut von irrelevanten Ergebnissen zu generieren. Mit anderen Worten, es geht darum, die Spreu vom Weizen zu trennen und nur die wirklich relevanten Informationen zu extrahieren. Köstler macht in ihrem Beitrag deutlich, dass die Wahl der Datenbank einen erheblichen Einfluss auf die Aussagekraft und Generalisierbarkeit von wissenschaftlichen Überblicksartikeln hat. Die richtige Auswahl und der richtige Mix der Datenbank in Systematic Reviews ist von essentieller Bedeutung, da sie darüber entscheiden, welche Studien und Informationen

in die Analyse einfließen und welche nicht. Diese Entscheidung kann erhebliche Auswirkungen auf die Qualität der Ergebnisse haben und sollte daher mit Bedacht getroffen werden.

Heinemann, Koschorreck, Wilmers, Sander und Leber (2023) gehen in ihrem Beitrag anhand mehrerer Beispiele der Frage nach, wie Critical Reviews gestaltet sein können, welche methodischen und inhaltlichen Herausforderungen damit einhergehen und welche Potenziale in dieser Form der Synthese liegen. Die Autor:innen arbeiten heraus, dass viele Schritte innerhalb eines Critical Reviews mit denen eines Systematic Reviews vergleichbar sind, jedoch in der Synthese mehr Freiheiten bieten und somit einerseits pointierter und andererseits zeiteffizienter sein können. Darüber hinaus bieten sie die Möglichkeit, bestimmte Aspekte kritisch zu hinterfragen und damit neue Perspektiven zu eröffnen. Beide Fälle zeigen, dass eine weitere kritische Auseinandersetzung mit dieser Methode notwendig ist und die Autor:innen mit ihrem Artikel einen vielversprechenden Anfang machen.

Brück-Hübner und Schluer (2023) gehen am Beispiel des Begriffs Feedback der Frage nach, ob systematische inhaltsanalytische Literaturreviews eine geeignete Methode zur Erstellung einer Begriffstaxonomie darstellen. Ausgangspunkt der Autor:innen ist dabei, dass in einem komplexen Wechselwirkungsfeld mediendidaktischer Forschung die Gefahr besteht, dass synthetisierende Methoden die Vielfalt zentraler, facettenreicher und heterogener Nutzendenkonstrukte nicht abbilden und damit verzerrende Effekte untersuchen. Sie argumentieren, dass vor einer inhaltlichen Synthese im eigentlichen Sinne vielmehr eine vertiefte Auseinandersetzung mit den verwendeten Konstrukten notwendig ist. Zu diesem Zweck untersuchen sie, ob sich systematische inhaltsanalytische Literaturreviews dafür eignen und welche Schritte dafür notwendig sind. Auf diese Weise gelingt es den Autor:innen, eine umfassende Taxonomie des Begriffs zu erstellen, die sich als Einführung in den Begriff gut eignet und damit zeigt, dass dieser Ansatz gelingen kann.

Literatur

- Ades, A. E., G. Lu, und J. P. T. Higgins. 2005. «The Interpretation of Random-Effects Meta-Analysis in Decision Models». *Medical decision making: an international journal of the Society for Medical Decision Making* 25 (6): 646–54. <https://doi.org/10.1177/0272989X05282643>.
- Afshari, A., J. Wetterslev, und A. F. Smith. 2017. «Can Systematic Reviews with Sparse Data Be Trusted?». *Anaesthesia* 72 (1): 12–16. <https://doi.org/10.1111/anae.13730>.
- Berliner, David C. 2002. «Comment: Educational Research: The Hardest Science of All». *Educational Researcher* 31 (8): 18-20. <https://doi.org/10.3102/0013189X031008018>.
- Bohlin, Ingemar. 2012. «Formalizing Syntheses of Medical Knowledge: The Rise of Meta-Analysis and Systematic Reviews». *Perspectives on Science* 20 (3): 273–309. https://doi.org/10.1162/POSC_a_00075.
- Borrego, Maura, Margaret J. Foster, und Jeffrey E. Froyd. 2014. «Systematic Literature Reviews in Engineering Education and Other Developing Interdisciplinary Fields». *Journal of Engineering Education* 103 (1): 45–76. <https://doi.org/10.1002/jee.20038>.
- Brück-Hübner, Annika, und Jennifer Schluer. 2023. «Was meinst du eigentlich, wenn du von «Feedback» sprichst? Chancen und Grenzen qualitativ-inhaltsanalytischer Scope-Reviews zur Herausarbeitung von Taxonomien zur Beschreibung didaktischer Szenarien am Beispiel «Feedback»». Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers, und Michael Kerres. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 54 (Research Syntheses): 125–66. <https://doi.org/10.21240/mpaed/54/2023.11.29.X>.
- Buntins, Katja, Svenja Bedenlier, Melissa Bond, Michael Kerres, und Olaf Zawacki-Richter. 2018. «Mediendidaktische Forschung aus Deutschland im Kontext der internationalen Diskussion. Eine Auswertung englischsprachiger Publikationsorgane von 2008 bis 2017». In *Digitalisierung und Hochschulentwicklung: Proceedings zur 26. Tagung der Gesellschaft für Medien in der Wissenschaft e.V.* Bd. 74, herausgegeben von Barbara Getto, Patrick Hintze, und Michael Kerres, 246–63. *Medien in der Wissenschaft* 74. Münster, New York: Waxmann.
- Buntins, Katja, Svenja Bedenlier, Victoria Marín, Marion Händel, und Melissa Bond. 2023. «Methodische Ansätze zu Evidenzsynthesen in der Bildungstechnologie: Eine tertiäre Übersichtsarbeit». Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers, und Michael Kerres. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 54 (Research Syntheses): 167–91. <https://doi.org/10.21240/mpaed/54/2023.12.20.X>.
- Buntins, Katja, Miriam Mulders, und Nadine Schroeder. 2023. «Kognitive Prozesse in Screening-Prozessen – Suchstrategien in der Bildungstechnologieforschung: Systematische Übersichtsarbeiten zum Lernen mit virtueller Realität». Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers, und Michael Kerres. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 54 (Research Syntheses): 103–24. <https://doi.org/10.21240/mpaed/54/2023.11.19.X>.
- Chen, Chien Chin, und You-De Tseng. 2011. «Quality evaluation of product reviews using an information quality framework». *Decision Support Systems* 50 (4): 755–68. <https://doi.org/10.1016/j.dss.2010.08.023>.

- Delere, Malte. 2020. «Konzepte medienpädagogischer Kompetenz von Lehramtsstudierenden in deutschsprachigen und internationalen Studien – ein systematisches Literaturreview». *Medienimpulse* 58 (02). <https://doi.org/10.21243/mi-02-20-16>.
- Esteves, Sandro C., Ahmad Majzoub, und Ashok Agarwal. 2017. «The Problem of Mixing ‘Apples and Oranges’ in Meta-Analytic Studies». *Translational andrology and urology* 6 (Suppl 4): 412-413. <https://doi.org/10.21037/tau.2017.03.23>.
- 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>.
- Hammersley, Martyn. 2020. «Reflections on the Methodological Approach of Systematic Reviews». In *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*, herausgegeben von Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond, und Katja Buntins, 23–39. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-27602-7_2.
- Hartung, Anja, Bernd Schorb, Horst Niesyto, Heinz Moser, und Petra Grell, Hrsg. 2014. *Jahrbuch Medienpädagogik 10. Methodologie und Methoden medienpädagogischer Forschung*. Bd. 10. Jahrbuch Medienpädagogik. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-04718-4>.
- Heinemann, Anna, Jan Koschorreck, Annika Wilmers, Pia Sander, und Jens Leber. 2023. «Potenziale und Herausforderungen von Forschungssynthesen im Kontext der Digitalisierung im Bildungsbereich am Beispiel des Critical Reviews». Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers, und Michael Kerres. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 54 (Research Synthesen): 79–102. <https://doi.org/10.21240/mpaed/54/2023.08.13.X>.
- Kimmons, Royce, und Joshua M. Rosenberg. 2022. «Trends and topics in educational technology», 2022 edition. *TechTrends* 66 (2): 134-140.
- Köstler, Verena. 2023. «Zwischen Präzision und Sensitivität: Generierung eines Studienkorpus am Beispiel einer Fragestellung zu Künstlicher Intelligenz (KI) in Bildungsprozessen». Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers, und Michael Kerres. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 54 (Research Synthesen): 1–27. <https://doi.org/10.21240/mpaed/54/2023.08.10.X>.
- Mertens, Claudia, Carolin Quenzer-Alfred, Anna-Maria Kamin, und Daniel Mays. 2023. «Publication Status as a Common Inclusion Criterion in Systematic Reviews: On the Use of Digital Media in the Context of Teaching Pupils with SEN in Inclusive, Integrative or Segregated Settings». Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers, und Michael Kerres. *MedienPädagogik: Zeitschrift Für Theorie Und Praxis Der Medienbildung* 54 (Research Synthesen): 51–78. <https://doi.org/10.21240/mpaed/54/2023.08.12.X>.

- Munk, Simon, Lisa Ziernwald, Kaley Lesperance, und Doris Holzberger. 2023. «Forschungssynthesen und der Researcher-Practitioner-Gap: Ein Diskussionsbeitrag zu Potenzialen und Grenzen anhand einer Forschungssynthese aus der Medienpädagogik». Herausgegeben von Svenja Bedenlier, Katja Buntins, Annika Wilmers, und Michael Kerres. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 54 (Research Syntheses): 28–50. <https://doi.org/10.21240/mpaed/54/2023.08.11.X>.
- Müller, Jane, Moritz Tischer, Mareike Thumel, und Paul Petschner. 2022. «Unboxing Digitale Souveränität: Ein Scoping Review Zu Digitaler Souveränität Von Individuen». *Medienimpulse* 60 (4): 37 Seiten. <https://doi.org/10.21243/mi-04-22-19>.
- Oakley, Ann, David Gough, Sandy Oliver, und James Thomas. 2005. «The politics of evidence and methodology: lessons from the EPPI-Centre». *Evidence & Policy: A Journal of Research, Debate and Practice* 1 (1): 5–32. <https://doi.org/10.1332/1744264052703168>.
- Petticrew, Mark. 2003. «Why Certain Systematic Reviews Reach Uncertain Conclusions». *BMJ (Clinical research ed.)* 326 (7392): 756–58. <https://doi.org/10.1136/bmj.326.7392.756>.
- Schmidt, Carlo. 2023. «Schule, Medien Und Gesellschaft: Ein Systematisches Review Der medienpädagogischen Zuschreibungen an Die Funktionen Der Schule». *MedienPädagogik: Zeitschrift für Theorie Und Praxis Der Medienbildung* 19 (Jahrbuch Medienpädagogik): 281–310. <https://doi.org/10.21240/mpaed/jb19/2023.03.11.X>.
- Sutton, Anthea, Mark Clowes, Louise Preston, und Andrew Booth. 2019. «Meeting the review family: exploring review types and associated information retrieval requirements». *Health Information & Libraries Journal* 36 (3): 202–222.
- Zawacki-Richter, Olaf, Michael Kerres, Katja Buntins, Melissa Bond, und Svenja Bedenlier, Hrsg. 2020. *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*. Springer Nature. <https://doi.org/10.1007/978-3-658-27602-7>.

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 (Schiefer-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.

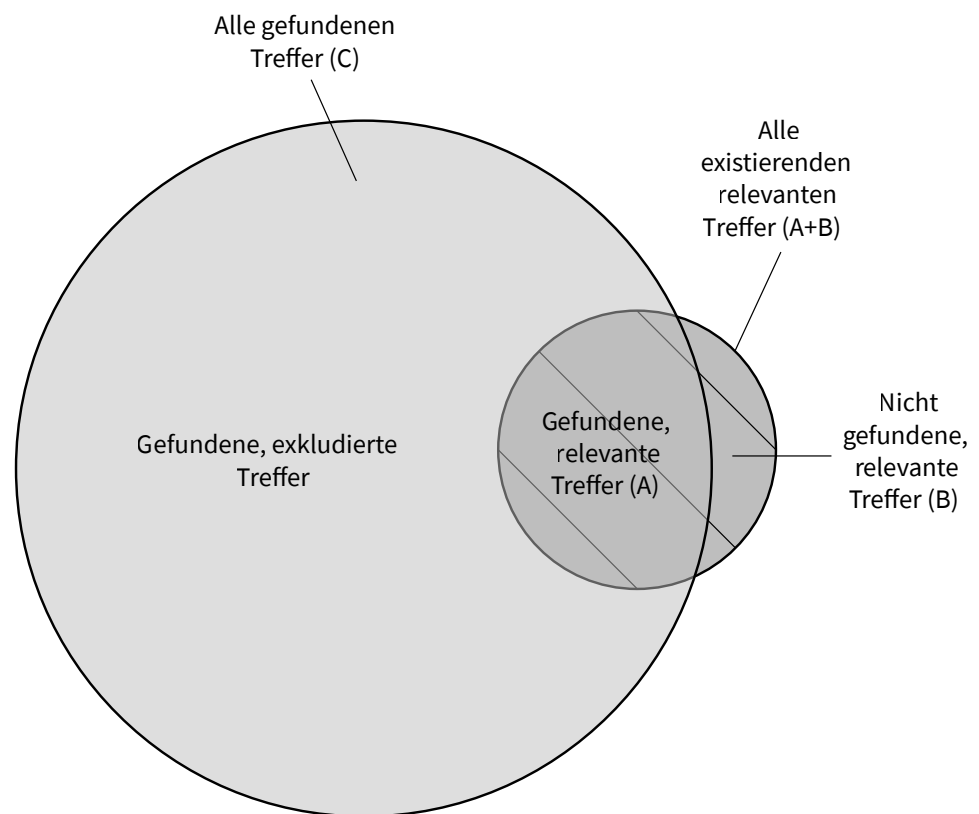


Abb. 1: Bezugsgrößen Sensitivität und Präzision, eigene Darstellung adaptiert nach Brunton et al. (2017, 112).

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 Educa- tion Source, Web of Science, Scopus)
Bond et al. (2020)	Student engagement and educational tech- nology 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 Im- pact Faktor, Google Scholar, Datenbanken (Web of Science, Science Direct), MOOC Platt- formen
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 intelligenz*» 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 Marc 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>.

Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

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

Forschungssynthesen und der Researcher-Practitioner-Gap

Ein Diskussionsbeitrag zu Potenzialen und Grenzen anhand einer Forschungssynthese aus der Medienpädagogik

Simon Munk^{1*} , Lisa Ziernwald^{1*} , Kaley Lesperance¹  und Doris Holzberger¹ 

¹ Technische Universität München, TUM School of Social Sciences and Technology; Zentrum für internationale Vergleichsstudien (ZIB)

* Der vorliegende Beitrag wurde in geteilter Erstautorenschaft verfasst.

Zusammenfassung

In den letzten Jahren hat die Berücksichtigung wissenschaftlicher Befunde im Bildungsbereich an Bedeutung gewonnen. Dabei zeigt sich jedoch, dass zwischen Wissenschaft und Praxis eine Lücke besteht, die dazu führt, dass wissenschaftliche Befunde nicht immer im Klassenzimmer ankommen (Researcher-Practitioner-Gap). Der vorliegende Artikel ist ein Diskussionsbeitrag, der Potenziale und Grenzen von Forschungssynthesen zur Verringerung des Researcher-Practitioner-Gaps beleuchtet. Es ergeben sich drei Hauptargumente, die zeigen, dass Forschungssynthesen in der Lage sind, den Researcher-Practitioner-Gap zu verkleinern: Überblick erleichtern, Generalisierbarkeit untersuchen und Transparenz und damit das Vertrauen in die Wissenschaft fördern. Trotz dieser Potenziale von Forschungssynthesen, den Researcher-Practitioner-Gap zu verkleinern, sollen auch Grenzen des Nutzens von Forschungssynthesen für die Praxis betrachtet werden. Hierzu zählen die Abstraktheit der Ergebnisse sowie die Schwierigkeit, Überzeugungen zu verändern. Um diesen Grenzen von Forschungssynthesen zu begegnen, spielt der Wissenschafts-Praxis-Transfer eine entscheidende Rolle. Ein wechselseitiger Austausch ermöglicht, dass empirische Evidenz in der Bildungspraxis ankommt und damit das Lehren und Lernen optimiert werden kann. Anhand eines Beispiels aus der Medienpädagogik wird die konkrete Umsetzung des Wissenschafts-Praxis-Transfers abschliessend näher beleuchtet.



Research Syntheses and the Researcher-Practitioner-Gap. A Contribution to the Discussion on Potentials and Limitations Based on a Research Synthesis on the Use of Educational Technology

Abstract

In recent years, attention to scientific findings in education has grown in importance. However, a gap has become apparent between science and practice, meaning scientific findings do not always reach the classroom (researcher-practitioner-gap). This article contributes to the discussion on this issue and sheds light on the potentials and limitations of research syntheses to reduce the researcher-practitioner-gap. We propose three main arguments that show that research syntheses are able to reduce the researcher-practitioner gap: facilitate an overview of the research, investigate generalizability, and increase transparency and therefore trust in science. Despite the potential of research syntheses to reduce the researcher-practitioner-gap, the limits of using research syntheses for practice should also be considered. These include the abstract nature of the results and the difficulty in changing teacher beliefs. To address the limitations of research syntheses, we argue that a science-practice transfer is of central importance. A two-way exchange enables empirical evidence to reach educational practice and thus can optimize teaching and learning. Finally, using an example from media education, the concrete implementation of science-practice transfer is examined in more detail.

1. Researcher-Practitioner-Gap und die Bedeutung von Forschungssynthesen

Die Bildungsforschung gilt als eine Disziplin, die durch ihre Arbeit auch Impulse für die Bildungspraxis setzen möchte: Neben Implikationen für die Forschung benennen wissenschaftliche Publikationen häufig auch mögliche Anknüpfungspunkte für die Praxis (Gräsel 2010). Trotz dieses Anspruchs und der festgelegten Zielsetzung, die Unterrichtspraxis an wissenschaftlichen Befunden zu orientieren (Kultusministerkonferenz 2019a, 2019b), scheint der Transfer zwischen Wissenschaft und Praxis nicht durchgehend zu gelingen. Es zeigt sich, dass zwischen Wissenschaft und Praxis eine Lücke besteht, die dazu führt, dass relevante wissenschaftliche Ergebnisse nicht immer im Klassenzimmer ankommen (Broekkamp und van Hout-Wolters 2007; Joyce und Cartwright 2020).

Diese Lücke, die als Researcher-Practitioner-Gap bekannt ist, ist problematisch, da Lehrpersonen so aufgrund von weitergeführten Traditionen häufig nicht die Methoden anwenden, die entsprechend wissenschaftlicher Befunde am lernwirksamsten wären (Bauer, Prenzel, und Renkl 2015; Everton, Galton, und Pell 2000; Knogler, Hetmanek, und Seidel 2022). Dies ist als besonders kritisch anzusehen, da wiederholt gezeigt werden konnte, dass bei Lehrpersonen Fehlvorstellungen

bezüglich pädagogischen Wissens vorliegen (Ferrero et al. 2020; Menz, Spinath, und Seifried 2021). Wissenschaft kann zwar nicht die Entscheidung abnehmen, welche konkrete Unterrichtsmethode im Einzelfall am ehesten zielführend ist. Jedoch kann Forschung eine Orientierungsgrundlage liefern, auf deren Basis dann konkrete Entscheidungen in einer bestimmten Unterrichtsstunde von der Lehrperson gefällt werden können. So können Annahmen abgebaut werden, die nicht vom derzeitigen Forschungsstand gedeckt werden (vgl. Bauer, Prenzel, und Renkl 2015; Menz, Spinath, und Seifried 2021).

Dabei ist die Integration von Wissen aus unterschiedlichen Quellen (z. B. Erfahrungswissen, Theorien, empirische Evidenz) in der Lehrpersonenbildung eine fortwährende Herausforderung (Renkl 2022). Besonders die Gewinnung von empirischer Evidenz in der Wissenschaft folgt einem bestimmten Ablauf, wobei am Ende dieses Prozesses häufig keine Informationen stehen, die von Lehrpersonen unmittelbar zur Problemlösung eingesetzt werden können (Knogler, Hetmanek, und Seidel 2022). Daher ist es zentral zu überlegen, welche Art von empirischer Evidenz und welche Prozesse im Rahmen des Wissenschafts-Praxis-Transfers hilfreich sein können, um Lehrpersonen evidenzbasiertes Handeln zu erleichtern.

Eine vielversprechende Quelle empirischer Evidenz, um den Researcher-Practitioner Gap zu verkleinern, sind Forschungssynthesen (Gräsel und Fussangel 2022; Levin 2004; Wissenschaftsrat 2016; Woerfel, Zimmer, und Witte 2021). Dabei handelt es sich um Überblicksarbeiten, die eine Vielzahl von Forschungsbefunden aus Einzelstudien systematisch zusammenfassen und analysieren (Holzberger und Ziernwald 2020). Beelmann (2014) nimmt dabei an, dass Forschungssynthesen als Überblicksarbeiten im Vergleich zu Einzelstudien von Bildungspraktiker:innen häufiger wahrgenommen und gelesen werden und damit eine wichtige Transferfunktion erfüllen. Die Bedeutsamkeit von Forschungssynthesen greift auch die Kultusministerkonferenz auf und äussert den Wunsch, «die stetig wachsende Anzahl von Forschungsergebnissen systematisch zu sichten, aufzubereiten und sowohl für die Bildungspolitik als auch für die Bildungspraxis bereitzustellen» (Kultusministerkonferenz 2016, 5f.).

Dabei können verschiedene Arten von Forschungssynthesen mit unterschiedlichen methodischen Vorgehensweisen unterschieden werden (Booth, Sutton, und Papaioannou 2016). Der vorliegende Beitrag fokussiert auf Metaanalysen und systematische Reviews, da diese im Vergleich zu anderen Arten von Forschungssynthesen und solchen anderer wissenschaftlicher Evidenz (z. B. Einzelstudien) besonders gut geeignet sind, um wissenschaftliche Evidenz zu konsumieren, anzuwenden und daraus Implikationen für die Bildungspraxis zu generieren (ebd.; Murad et al. 2016). Die konkreten Potenziale und Vorteile von Forschungssynthesen zur Verringerung des Researcher-Practitioner-Gaps im Vergleich zu anderer wissenschaftlicher Evidenz werden in Kapitel 3 diskutiert.

Zusammenfassend ist der vorliegende Artikel ein Diskussionsbeitrag, der zunächst einen Überblick über den Ablauf einer Forschungssynthese gibt und anschliessend Potenziale und Grenzen dieser Publikationsform zur Verringerung des Researcher-Practitioner-Gaps beleuchtet. Ferner widmen wir uns dem Wissenschafts-Praxis-Transfer und beschreiben, wie dieser definiert werden kann, welche Gelingensbedingungen und Hindernisse zu berücksichtigen sind und wie dieser gestaltet werden kann. Abschliessend beschreiben und reflektieren wir die Bemühungen zur Verringerung des Researcher-Practitioner-Gaps anhand eines konkreten Beispiels – einer Forschungssynthese zum Einsatz digitaler Medien im mathematisch-naturwissenschaftlichen Unterricht der Sekundarstufe (Hillmayr et al. 2020) – und leiten daraus Empfehlungen für den zukünftigen Wissenschafts-Praxis-Transfer zur Verringerung des Researcher-Practitioner-Gaps ab.

2. Ablauf einer Forschungssynthese

Bevor mit der methodischen Umsetzung einer Forschungssynthese begonnen wird, steht die Entscheidung über das Thema im Vordergrund. Das Finden eines praxisrelevanten Themas für eine Forschungssynthese ist entscheidend, da die Forschung nur so Bedarfe der Praxis passgenau adressieren und die Relevanz der Ergebnisse für die Bildungspraxis ermöglichen kann (Knogler, Hetmanek, und Seidel 2022). Dies stellt einen entscheidenden Beitrag zur Verringerung des Researcher-Practitioner-Gaps dar. Das gefundene Thema wird dann in eine wissenschaftliche Fragestellung überführt, bevor die methodische Umsetzung beginnt. Der Ablauf einer Forschungssynthese im Kontext des Wissenschafts-Praxis-Transfers kann in fünf übergeordnete Schritte unterteilt werden (s. Abb. 1).

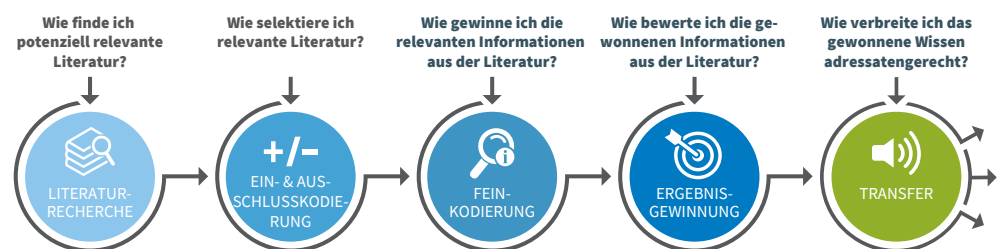


Abb. 1: Ablauf einer Forschungssynthese.

Im ersten Schritt, der Literaturrecherche, ist das Ziel, systematisch möglichst alle Studien zum untersuchten Thema der Forschungssynthese zu finden. Dafür sollen verschiedene Suchbegriffe inklusive Synonyme (z. B. computer, pc, digital media etc.), Literaturquellen (z. B. Datenbanken) und Suchmethoden (z. B. Call for Studies über Verteiler von Forschungsgesellschaften) verwendet werden (Bramer et al. 2018; Gusenbauer und Haddaway 2020). Im zweiten Schritt, der Ein- und

Ausschlusskodierung (Screening), werden alle relevanten Studien auf der Basis von Titel und Kurzzusammenfassung für die Forschungsfrage systematisch ausgewählt. Hier sollten alle Kriterien kommuniziert werden, welche für die Ein- und Ausschlussentscheidung verwendet werden, sodass transparent und nachvollziehbar wird, anhand welcher Kriterien die Studien ausgewählt wurden. Ziel des dritten Schritts, der Feinkodierung, ist, alle relevanten Informationen systematisch aus den Studien zu gewinnen. Die Ein- und Ausschlusskodierung sowie die Feinkodierung werden (zu einem gewissen Teil) von mindestens zwei unabhängigen Wissenschaftler:innen durchgeführt, um die Qualität der Ratings sicherzustellen und die objektive Nachvollziehbarkeit des Prozesses durch die Interrater-Reliabilität zu beurteilen (Lefebvre et al. 2019; Li, Higgins, und Deeks 2019). Anschliessend werden im vierten Schritt, der Ergebnispengewinnung, die gewonnenen Informationen qualitativ und/oder quantitativ zusammengefasst und aufbereitet. Hier gibt es zahlreiche Methoden, welche in den letzten Jahren weiter ausgebaut und verbessert wurden (Booth, Sutton, und Papaioannou 2016; Borenstein 2009; Siddaway, Wood, und Hedges 2019). Verfolgt eine Forschungssynthese beispielsweise das Ziel, den Forschungsstand inklusive existierender Forschungslücken zu einem Thema der Bildungsforschung aufzubereiten, kann die Ergebnisdarstellung der in Schritt 4 gewonnenen Erkenntnisse in wissenschaftlichen Zeitschriften das Ende des Prozesses einer gelungenen Forschungssynthese markieren. Der fünfte Schritt, der Transfer, ist jedoch aus unserer Sicht für die Reduktion des Researcher-Practitioner-Gaps unerlässlich, da wissenschaftliche Forschungsberichte «normalerweise nicht auf eine Rezeption durch eine fachfremde, wissenschaftsexterne Öffentlichkeit angelegt» sind (Bromme, Prenzel, und Jäger 2014, 8). Um dennoch die Verwendung von Ergebnissen einer Forschungssynthese zu ermöglichen, beschreiben wir deshalb gesondert in Kapitel 5, wie der Transfer der Ergebnisse von Forschungssynthesen konkret gestaltet werden kann.

3. Potenziale von Forschungssynthesen zur Verringerung des Researcher-Practitioner Gaps

Aus unserer Sicht sind Forschungssynthesen aufgrund von drei zentralen Potenzialen in der Lage, den Researcher-Practitioner-Gap zu verringern: Forschungssynthesen können dabei helfen, (a) einen Überblick über die aktuelle Studienlage sowie aktuelle Theorien zu schaffen, (b) die Generalisierbarkeit und Widersprüchlichkeit von Forschungsergebnissen zu untersuchen und (c) die Transparenz und damit das Vertrauen in die Wissenschaft zu fördern.

Forschungssynthesen können dabei helfen, einen *schnellen Überblick* über den aktuellen Forschungsstand zu gewinnen (Holzberger und Ziernwald 2020; Weckmann, Chenot, und Reber 2015). Neben der Übersicht über den empirischen Forschungsstand können sie auch als Ergebnis einen Überblick über die theoretische

Fundierung bildungsrelevanter Konzepte geben und diese dadurch weiterentwickeln (z. B. Booth, Sutton, und Papaioannou 2016; Seel 2017; Webster und Watson 2002). So können beispielsweise die von Renkl (2022) genannten Vorteile von Theorien (z. B. Darstellung eines kohärenten Rahmens) auch in Forschungssynthesen genutzt werden. Durch die Zusammenfassung einer Vielzahl an Einzelstudien und theoretischen Fundierungen kann die Übersichtsarbeit anstelle vieler Einzelstudien gelesen werden, wodurch dem Hindernis ‹fehlende Zeit der Bildungspraktiker:innen› (z. B. Ratcliffe et al. 2005; Thomm et al. 2021) begegnet werden kann.

Als ein weiteres Potenzial kann die Überprüfung der *Generalisierbarkeit* von Forschungsergebnissen genannt werden. Einzelstudien zum gleichen Thema kommen immer wieder zu unterschiedlichen Ergebnissen. Dadurch fragen sich Lehrpersonen beispielsweise, ob ein Studienergebnis auch für ihr Land oder ihre Schulart anwendbar ist (Joram, Gabriele, und Walton 2020; Joyce und Cartwright 2020). Da der Gesamteffekt der Forschungssynthese durch die Zusammenfassung von Einzelstudien aus verschiedenen Kontexten auf einer grösseren und breiteren Datengrundlage beruht, verfügt er einerseits meist im Vergleich zu Einzelstudien über präzisere Schätzungen (Cohn und Becker 2003). Nichtsdestotrotz setzt sich andererseits der Gesamteffekt häufig aus heterogenen Studienergebnissen zusammen. Mithilfe von Forschungssynthesen kann die Grösse der Heterogenität in den Ergebnissen bestimmt werden. So können neben dem Gesamteffekt, welcher einen guten Überblick über das Gesamtbild liefert, auch Einflüsse von Kontextfaktoren auf die Ergebnisse untersucht werden (sog. Moderatoranalysen). Unter Kontextfaktoren versteht man Rahmenbedingungen, die das Ergebnis möglicherweise beeinflussen. Beispiele für Kontextfaktoren sind das Alter der Schüler:innen, das Unterrichtsfach oder die Schulart. Dadurch lassen sich widersprüchlich erscheinende Ergebnisse aus Einzelstudien erklären und Ergebnisse für spezifischere Anwendungskontexte (z. B. Effekte für Schüler:innen einer bestimmten Schulart) gewinnen.

In den letzten Jahren haben weitere methodische Verfahren (z. B. die Multiversums- und Spezifikationskurvenanalyse) im Bereich der Forschungssynthesen an Bedeutung gewonnen, um die Heterogenität zwischen den eingeschlossenen Einzelstudien genauer zu untersuchen (Voracek, Kossmeier, und Tran 2019). Damit wird das Potenzial von Forschungssynthesen, Unterschiede und mögliche Widersprüche zwischen Einzelstudien zu diskutieren und aufzulösen, weiter gestärkt. Anhand einer separaten Betrachtung von Kontexten (z. B. Schüler:innen nur aus einer bestimmten Schulart oder Altersgruppe) können Praktiker:innen das Ergebnis auswählen, welches zu ihrem Anwendungskontext am besten passt.

Das dritte zentrale Potenzial ist, dass Forschungssynthesen dazu beitragen können, *Transparenz zu schaffen*. Die Methode der Forschungssynthese eignet sich besonders, Transparenz und damit Vertrauen herzustellen, da sie ein sehr systematisches Vorgehen auszeichnet, das von Dritten nachvollzogen und überprüft werden

kann (Pant 2014; Woerfel, Zimmer, und Witte 2021). Durch das schrittweise Vorgehen, welches über verschiedene Forschungssynthesen hinweg ähnlich umgesetzt wird, ist eine Beurteilung der Qualität durch Dritte einfacher möglich. Dass die Lesenden sich selbst ein Urteil über die Qualität und damit Vertrauenswürdigkeit der Forschungssynthese machen können, kann zu mehr Vertrauen in das Forschungsergebnis beitragen. Dieses Vertrauen kann als essenziell angesehen werden, damit wissenschaftliche Befunde in der Bildungspraxis aufgegriffen und umgesetzt werden (Herzog 2022) und somit der Researcher-Practitioner-Gap verkleinert wird. Dieses Potenzial der Transparenz wird auch in sogenannten *Community-Augmented Meta-Analyses* (CAMAs) verwirklicht. Hierbei werden Metaanalysen inklusive der eingeflossenen Daten öffentlich verfügbar gemacht. Dadurch können Analysen repliziert und beispielsweise bestimmte Rahmenbedingungen von Forschungsergebnissen weiter untersucht werden (Burgard, Bošnjak, und Studtrucker 2021).

Diese drei zentralen Potenziale von Forschungssynthesen erleichtern den Einsatz von Forschungsergebnissen in der Bildungspraxis und ermöglichen dadurch, Hindernisse wie zum Beispiel fehlende Zeit zu überwinden. Diese Potenziale bieten damit also die Möglichkeit, den Researcher-Practitioner-Gap zu verkleinern, sind jedoch nicht automatisch hinreichend, um eine Veränderung in der Praxis zu bewirken. Zentrale Voraussetzung, um diese Potenziale möglichst gewinnbringend in der Praxis nutzen zu können, ist der Wissenschafts-Praxis-Transfer.

4. Grenzen von Forschungssynthesen zur Verringerung des Researcher-Practitioner Gaps

Trotz der Potenziale von Forschungssynthesen sollen auch Grenzen ihres Nutzens für die Praxis diskutiert werden: die Abstraktheit ihrer Ergebnisse und die Schwierigkeit, bestehende Überzeugungen zu verändern. Hinsichtlich der *Abstraktheit* ist festzuhalten, dass Details einer Einzelstudie (z. B. bezüglich ihrer Intervention) beim Aggregieren zu einer Forschungssynthese nicht bestehen bleiben können und Bildungspraktiker:innen dadurch möglicherweise Schwierigkeiten haben, eine bestimmte Massnahme konkret umzusetzen (Renkl 2022). Auch wenn Forschungssynthesen häufig versuchen, durch Moderatoranalysen die Heterogenität der Einzelstudien zu erklären, ist das nicht immer eine ausreichende Lösung für die Bildungspraxis. Lehrpersonen können durch eine Vielzahl an Effektstärken (z. B. overall-Effekt, Moderatoranalysen) überwältigt sein (ebd.).

Eine weitere Grenze von Forschungssynthesen bei der Verringerung des Researcher-Practitioner Gaps ist, dass die alleinige Darstellung von Forschungsergebnissen nicht automatisch zu einer *Veränderung bestehender Überzeugungen* von Lehrpersonen führt (Menz, Spinath, und Seifried 2021). Es konnte jedoch gezeigt werden, dass sogenannte *refutation texts*, welche die eigenen bestehenden (Fehl-)Vorstellungen

aufgreifen und Inhalte enthalten, die diesen widersprechen (Lassonde, Kendeou, und O'Brien 2016), Fehlvorstellungen bei angehenden Lehrpersonen reduzieren können. Dies war insbesondere der Fall, wenn es sich um Texte mit Ergebnissen aus empirischen Studien, nicht um solche mit anekdotischen Inhalten handelte (Menz, Spinath, und Seifried 2021). Gleichzeitig waren die Veränderungen der Fehlvorstellungen jedoch nicht langfristig. Des Weiteren wurde keine Verringerung der Verwendungsabsicht bestimmter Unterrichtsmethoden gefunden, welche auf den Fehlvorstellungen basieren (Ferrero et al. 2020). Aus diesem Grund ist es nicht ausreichend, die Forschungsergebnisse zu veröffentlichen und darauf zu warten, dass sie in der Bildungspraxis Änderungen erzeugen und Verwendung finden werden.

Um diese Grenzen von Forschungssynthesen abzubauen, ist es aus unserer Sicht sinnvoll und notwendig, dass einerseits Lehrpersonen im Umgang mit Forschungssynthesen geschult, andererseits Wissenschaftler:innen ermutigt werden, die Ergebnisse von Forschungssynthesen ansprechend in die Bildungspraxis zu transferieren. Daher wird im nächsten Abschnitt die Bedeutsamkeit des Wissenschafts-Praxis-Transfers zur Verringerung des Researcher-Practitioner-Gaps diskutiert.

5. Transfer der Ergebnisse von Forschungssynthesen

5.1 Begriffsklärung

Damit die Ergebnisse von Forschungssynthesen in der Unterrichtspraxis ankommen und der Researcher-Practitioner-Gap verkleinert werden kann, ist der Transfer zwischen Bildungsforschung und Bildungspraxis bedeutsam. Während zwar bereits ein intuitives Verständnis des Begriffs *Wissenschafts-Praxis-Transfer* vorliegt, wollen wir diesen zunächst vom verwandten Begriff *Wissenschaftskommunikation* abgrenzen. Im Rahmen des *Wissenschafts-Praxis-Transfers* sollen

«wissenschaftliche Erkenntnisse in bildungspraktische Handlungsfelder (Bildungspraxis, -administration und -politik) einfließen und relevante Fragestellungen sowie Wissen aus diesen Bereichen an die Wissenschaft herangetragen und dort berücksichtigt werden» (Dehmel 2017, 1).

Transfer ist damit ein wechselseitiger Prozess, der von beiden Seiten mit ihren jeweiligen Perspektiven und Expertisen bereichert wird. *Wissenschaftskommunikation* wiederum ist «der Transfer von wissenschaftlichem Wissen [...] in außerwissenschaftliche Bereiche und der Austausch darüber mit der Gesellschaft, Kultur, Wirtschaft und Politik» (Wissenschaftsrat 2021, 7). Wissenschaftskommunikation kann also als Überbegriff verstanden werden und konzentriert sich im Gegensatz zum *Wissenschafts-Praxis-Transfer* nicht nur auf den Austausch mit der Fachpraxis,

sondern strebt auch einen gesamtgesellschaftlichen Diskurs an. Für die Verringerung des Researcher-Practitioner-Gaps konzentrieren wir uns in diesem Beitrag auf den Wissenschafts-Praxis-Transfer von Forschungssynthesen und schildern im Folgenden, wie dieser Transfer zur Verringerung des Gaps beitragen kann.

5.2 Wichtige Einflussfaktoren für den Transfer

Um genauere Angaben zur Ausgestaltung des Transfers von Ergebnissen aus Forschungssynthesen abzuleiten, ist zunächst wichtig zu klären, welche Barrieren einem gelungenen Wissenschafts-Praxis-Transfer im Wege stehen, um diese entsprechend zu überwinden: Häufig wird hierbei die *Sprache der Wissenschaft* genannt (z. B. Vanderlinde und van Braak 2010). Aus einer zu technischen und komplexen Sprache mit vielen Fachbegriffen resultieren Verständnisschwierigkeiten bei Bildungspraktiker:innen. Ein blosses Zerlegen der wissenschaftlichen Publikation in kleinere Teile erscheint deshalb als nicht ausreichend für den Wissenschafts-Praxis-Transfer (De Corte 2000).

Auch die *fehlende Zeit* sich mit Forschungsergebnissen auseinanderzusetzen, vergrößert die Lücke zwischen Wissenschaft und Praxis (Levin 2004; Ratcliffe et al. 2005; Thomm et al. 2021). Zusätzlich zur ohnehin hohen Zahl an Aufgaben, die Lehrpersonen in ihrer Arbeitszeit bewältigen müssen, kann kaum Zeit für das Studieren von wissenschaftlichen Einzelstudien bleiben. Knogler, Hetmanek und Seidel (2022, 139) fassen diese Problematik wie folgt zusammen:

«Der Zugang zu den oftmals lizenzierten Quellen sowie die notwendige Zeit und Expertise bei Suche, Auswahl und Interpretation stellen für viele Lehrkräfte und Personen in der Lehrerinnen- und Lehrerbildung eine nahezu unüberwindbare Hürde dar.»

Ferner berichten auch Forschende von fehlenden zeitlichen Ressourcen als Hindernis für den Wissenschafts-Praxis-Transfer (Blatter, Schelle, und Michl 2022).

Auch *Einstellungen der Praktiker:innen* vergrößern den Researcher-Practitioner-Gap: Je stärker Lehrpersonen zustimmten, dass ihnen Zeit und Fähigkeiten fehlen, um wissenschaftliche Studien zu suchen und zu lesen, desto eher bewerteten sie wissenschaftliche Ergebnisse als für ihre Praxis irrelevant (Thomm et al. 2021). Häufig ist ihnen das Verbesserungspotenzial für Unterrichtspraktiken durch das Hinzuziehen wissenschaftlicher Befunde nicht bewusst (Levin 2004). Damit sehen sie keinen Bedarf, etwas an ihrer bestehenden Unterrichtspraxis zu verändern. Des Weiteren stehen mitunter wissenschaftliche Befunde dem Erfahrungswissen von Lehrpersonen entgegen, was Frustration über diesen Zwiespalt auslösen kann (Levin 2004). Ermutigend für den Wissenschafts-Praxis-Transfer ist jedoch zum einen, dass eine Änderung der Einschätzung auf Basis empirischer Evidenz möglich ist

(Menz, Spinath, und Seifried 2021). Zum anderen ist die Nachfrage nach adressatengerecht dargestellter wissenschaftlicher Evidenz in den letzten Jahren im Bildungswesen angestiegen (Woerfel, Zimmer, und Witte 2021).

Doch auch die *Einstellung der Forschenden* kann den Researcher-Practitioner-Gap vergrößern: Indem sich Forschende stärker auf Grundlagenforschung konzentrieren, ist der Transfer des Wissens in den Anwendungsbereich schwieriger (Gräsel 2010). Auch wenn Forschende den Wissenschafts-Praxis-Transfer überwiegend zu ihren Aufgaben zählen, nennen sie hinderliche Rahmenbedingungen des Wissenschaftssystems, die den Transfer erschweren. Hierzu zählt die fehlende Reputation des Wissenschafts-Praxis-Transfers für eine wissenschaftliche Karriere (Blatter, Schelle, und Michl 2022), für die häufig Indikatoren wie wissenschaftliche Publikationen oder Drittmittelwerbung eine bedeutendere Rolle spielen.

Damit wird deutlich, dass es eine Reihe von Stellschrauben gibt, an denen man ansetzen kann, um die Lücke zwischen Wissenschaft und Praxis zu verkleinern (Farley-Ripple et al. 2018). Diese geben Hinweise, wie die konkrete Gestaltung des Transfers gelingen kann.

5.3 Gestaltung des Transfers

Bevor wir die Gestaltung des Transfers genauer beleuchten, ist zunächst dessen Ziel zu klären. Dieses ist, wie er im vorliegenden Beitrag diskutiert wird, der Transfer von Wissen, nicht gleichzusetzen mit der Vermittlung von konkreten Handlungsweisen oder Verhaltenstrainings. Um die Gestaltung dieses Transfers genauer zu beleuchten, orientieren wir uns an den folgenden Fragen: (a) Wer sind die Akteur:innen? (b) Was wird kommuniziert? (c) Wie wird kommuniziert? (d) Wo wird kommuniziert? Bei der Beantwortung dieser Fragen ist hervorzuheben, dass die Abläufe des Transfers sich oftmals nicht danach unterscheiden, ob das Ergebnis einer Einzelstudie oder einer Forschungssynthese kommuniziert wird. Ausgehend von den zuvor genannten Potenzialen von Forschungssynthesen (schneller Überblick, Überprüfung der Generalisierbarkeit, Schaffen von Transparenz) für den Abbau des Researcher-Practitioner-Gaps leiten wir jedoch ab, dass der Transfer von Forschungssynthesen für den Austausch zwischen Bildungsforschung und Bildungspraxis in besonderem Mass vielversprechend ist.

(a) Wer sind die Akteur:innen?

Die Frage nach den Akteur:innen ist zentral für die Betrachtung des Wissenschafts-Praxis-Transfers und seines Beitrags zur Verringerung des Researcher-Practitioner-Gaps. Aus unserer Sicht sind es fünf Gruppen von Akteur:innen, die miteinander in den Austausch treten: Forschende, Bildungspraxis, Bildungspolitik, Clearingstellen sowie Multiplikator:innen, wobei einzelne Personen auch mehrere Rollen

einnehmen können. *Forschende* können als Urheber:innen der Forschungssynthese authentisch über die Ergebnisse berichten und ihr Wissen weitergeben. Vonseiten der *Bildungspraxis* stehen im vorliegenden Beitrag zum einen Lehrpersonen und der Kontext Schule im Fokus. Grundsätzlich können Lehrpersonen in sämtlichen Phasen der Lehrpersonenbildung (Studium, Vorbereitungsdienst und im Anschluss an ihre Ausbildung) aus der Bildungsforschung Impulse für ihre Arbeit erhalten. Allerdings unterscheiden sich Art und Ort der Kommunikation zwischen den Phasen der Lehrpersonenbildung erheblich (siehe unter c und d). Umgekehrt können sie aber auch offene Probleme aus ihrem unterrichtspraktischen Alltag an die Forschenden adressieren, sodass diese Probleme zu Forschungsfragen für Forschungssynthesen werden. Häufig liegen die Entscheidungen jedoch nicht nur in der Bildungspraxis. Es ist Aufgabe der *Bildungspolitik*, die Weichen für gelingende Bildung zu stellen und dabei zu unterstützen, dass empirisches Wissen auch in konkretes Unterrichtshandeln münden kann. Diese Entscheidungen können und sollten ebenfalls auf der Basis wissenschaftlicher Erkenntnisse erfolgen.

Vor dem Hintergrund jedoch, dass Forschende berichten, kaum zeitliche Ressourcen für den Wissenschafts-Praxis-Transfer zu haben und ihre Anstrengungen in diesem Bereich für Qualifikationsprozesse wenig bedeutsam sind, rücken zudem *Clearingstellen* wie das *Clearing House Unterricht* (Diery et al. 2020) in den Fokus. Dieses Vorgehen hat Vor- und Nachteile: Kommunizieren die Forschenden selbst, so können sie authentisch von ihrer Arbeit berichten, ohne dass eine weitere Person oder Institution für bessere Verständlichkeit die ursprüngliche Kernaussage abwandelt. Kommunizieren dagegen Clearing-Stellen, können diese gezielt eine Vermittlerfunktion an der Schnittstelle zwischen Wissenschaft und Bildungspraxis übernehmen und sich stärker auf die sprachliche Aufbereitung wissenschaftlicher Befunde konzentrieren. Sie nehmen ausserdem eine neutrale Position in der Vermittlung von Forschungsergebnissen ein und können so auch objektiv die Qualität der wissenschaftlichen Befunde bewerten. Unter Abwägung dieser Vor- und Nachteile sind Clearingstellen aus unserer Sicht neben den Forschenden wichtige Protagonist:innen des Transferprozesses. Des Weiteren gibt es auch eine Reihe von *Multiplikator:innen*, die an der Schnittstelle zwischen Wissenschaft und Praxis arbeiten und eine Vermittlerrolle einnehmen können. Zum einen sind hiermit Landesinstitute und Qualitätseinrichtungen der Länder gemeint, die in ihrer Arbeit ihrerseits Transferbezüge aufzeigen oder auf bestehende Formate hinweisen können. Zum anderen zählen auch Lehrerbildende hierzu, die in der Ausbildung von Lehrpersonen Wissen vermitteln.

Zusammenfassend ist anzumerken, dass es das Zusammenspiel sämtlicher Akteur:innen ist, das zu einem gelungenen Wissenschafts-Praxis-Transfer führt. Hierfür sind gemeinsame Bemühungen aller Akteur:innen notwendig.

(b) Was wird kommuniziert?

Um das Interesse der Bildungspraxis zu wissenschaftlichen Ergebnissen herzustellen und damit den Researcher-Practitioner-Gap zu verringern, sind die kommunizierten Inhalte (Vorgehen, Ergebnisse, Bewertung und Einordnung sowie Reflexion) eine entscheidende Stellschraube. Zentrale Voraussetzung für die Kommunikation sämtlicher Inhalte ist die Anwendbarkeit des Forschungsergebnisses in der Praxis (Sibanda und Begede 2015). Dies unterstreicht erneut die Wichtigkeit der Suche nach einer geeigneten Forschungsfrage für die Forschungssynthese. Da Forschungssynthesen den Überblick über ein Thema erleichtern können, eignen sie sich besonders, um Lehrpersonen den Wunsch nach umfassendem Wissen zu einem Thema erfüllen zu können. Während eine Einzelstudie häufig nur eine kleine Facette eines Themas abdeckt, können Forschungssynthesen einen breiteren Zugang bieten und mehr Perspektiven eines Problems abdecken.

Trifft die Forschungsfrage den Bedarf der Praxis, so ist ferner für eine aufgeschlossene Einstellung der Bildungspraxis zur Bildungsforschung wichtig, dass das *Vorgehen* transparent und verständlich dargestellt wird. So wird deutlich, dass die Befunde aus einem nachvollziehbaren Prozess entstanden sind. Durch ihr schrittweises und übersichtliches Vorgehen eignen sich Forschungssynthesen hierfür in besonderem Mass.

Neben dem Vorgehen bleibt aber das *Ergebnis* das Herzstück der Kommunikation. Aus unserer Sicht ist hierbei ratsam, sich klar auf wenige zentrale Botschaften der Forschungssynthese zu beschränken. Hierfür eignen sich Forschungssynthesen gut, da sie allgemeinere Aussagen (z. B. über den overall-Effekt) sowie spezifischere Aussagen (z. B. zu Moderatoranalysen) zulassen.

Ferner kann sich beim Transfer durch Clearing-Stellen auch noch eine *Bewertung und Einordnung* der Ergebnisse anschliessen. Der Vorteil dieses Vorgehens besteht darin, dass der Bildungspraxis nicht nur die Selektion der Befunde, sondern auch deren Bewertung erleichtert wird. Somit kann wiederum die Darstellung der Stärken und Schwächen einer Forschungssynthese durch eine unabhängige Clearing-Stelle auch die Transparenz von Forschungssynthesen und dadurch das Vertrauen in die Forschung erhöhen. Ein Nachteil kann möglicherweise darin bestehen, dass Lehrpersonen durch die vorgegebene Bewertung die eigene kritische Reflexion der Ergebnisse vernachlässigen.

Neben dem Vorgehen, den Ergebnissen und der Bewertung kann noch ein weiteres Werkzeug ein Schritt auf dem Weg des Transfers von der wissenschaftlichen Erkenntnis zur unterrichtspraktischen Entscheidung sein. Mithilfe von *Reflexionsfragen* nennen und beschreiben Forschende nicht nur das Ergebnis der Forschungssynthese, sondern geben auch einen Impuls zum Weiterdenken. Dieser kann von Praktiker:innen aufgenommen werden, sodass sie selbstständig die Relevanz für ihr Tätigkeitsfeld erarbeiten können.

(c) Wie wird kommuniziert?

Betrachtet man das «Wie» der Kommunikation, so ist zwischen Formaten und deren Eigenschaften zu unterscheiden. Im Folgenden präsentieren wir verschiedene Formate (deutschsprachige Publikationen, Erklärvideos, Podcasts) und gehen jeweils auf Eigenschaften ein, die einen gelungenen Transfer ermöglichen. Um den Research-Practitioner-Gap zu verringern und das Hindernis Sprache (Vanderlinde und van Braak 2010) abzubauen, lässt sich schlussfolgern, dass es für die Bildungspraxis neben einem wissenschaftlichen, häufig englischsprachigen Zeitschriftenartikel mit vielen Fachbegriffen auch eine *deutschsprachige Publikation* geben sollte. Eine solche Publikation – etwa ein Themenheft, Kurzreview oder Online-Dossier – sollte nur unerlässliche Fachbegriffe enthalten und diese adressatengerecht erklären. Diese Formen haben gemeinsam, dass sie den Forschungsstand zu einem praxisrelevanten Thema auswählen und adressatengerecht aufbereiten. Im Unterschied zur wissenschaftlichen Publikation erscheint auch eine kleinschrittigere Beschreibung von Methoden, Modellen und Ergebnissen wichtig. Möglichkeiten, dies zu gestalten, sind das Hervorheben wichtiger Take-Home-Messages oder die Verbildlichung der Ergebnisse durch kleine Schaubilder. Forschungssynthesen eignen sich hier in besonderem Mass, da sie überblicksartige Botschaften senden können. Solche lassen sich besonders gut in Schaubildern darstellen, da von den Betrachtenden zum Verständnis der Botschaft weniger Hintergrundwissen nötig ist als beispielsweise für das Verständnis eines spezifischen Ergebnisses einer Einzelstudie.

Da die Akteur:innen der Bildungspraxis häufig von wenig Zeit berichten (Levin 2004; Ratcliffe et al. 2005; Thomm et al. 2021) stellen auch kurze *Erklärvideos* oder *Podcasts* ein geeignetes Format des Transfers dar. Inhaltlich ähnlich wie ein Themenheft oder Kurzreview aufgebaut, liefern Erklärvideos oder Podcasts eine komprimierte Darstellung der Inhalte. So lassen sich die Inhalte im Vergleich zum Lesen eines Themenhefts in noch kürzerer Zeit und in verschiedenen Situationen ansprechend transportieren und konsumieren. Aufgrund des Potenzials von Forschungssynthesen, kurze, überblicksartige Botschaften zu senden, eignen sich auch kürzere Transferformate in besonderer Weise.

(d) Wo wird kommuniziert?

Neben dem Kommunikationsinhalt ist auch die Wahl des Kommunikationskanals für die Verringerung des Research-Practitioner-Gaps entscheidend. Wissenschaftliche Publikationen sind häufig nicht frei zugänglich und können deshalb nur gegen Gebühr gelesen werden. Vor diesem Hintergrund ist es wichtig, Transferformate wie Themenhefte oder Erklärvideos kostenlos zum Download zur Verfügung zu stellen (Open Access). Besonders geeignet erscheinen hierfür Social-Media-Plattformen, da hier ohnehin eine Vielzahl an Lehrpersonen aktiv ist. Beispielsweise finden sich auf Twitter unter dem Hashtag *twitterlehrerzimmer* täglich neue Fragen,

Rückmeldungen und Anregungen, um den unterrichtspraktischen Alltag effektiv zu gestalten. Hier kann dann beispielsweise auf eine Website mit einem neuen Themenheft oder Erklärvideo verlinkt werden. Da soziale Medien kurze Botschaften bevorzugen, können Kernbotschaften von Forschungssynthesen hier gewinnbringend distribuiert werden. Auch die orts- und zeitunabhängige Verfügbarkeit der Inhalte scheint zentral. Durch einen kostenfreien rund-um-die-Uhr-Zugriff zu Plattformen können auch Zugfahrten oder Wartezeiten für die Auseinandersetzung mit Forschungsergebnissen genutzt werden.

Bei all diesen Bemühungen, Hindernisse, die zu einer Vergrößerung des Researcher-Practitioner-Gaps beitragen, zu überwinden, bleibt der Dialog ein unerlässlicher Bestandteil (Mortari, Girelli, und Silva 2016). Wissenschafts-Praxis-Transfer sollte deshalb nie als Einbahnstrasse sondern immer als wechselseitiger Austausch verstanden werden. Transferveranstaltungen sind eine Gelegenheit, miteinander in den Austausch zu treten. Aus unserer Sicht ist hierbei wichtig, dass beide Perspektiven (die der Wissenschaft und die der Praxis) ausreichend gehört und wertgeschätzt werden. Dies kann gelingen, indem nicht nur Forschende sondern auch Praktiker:innen kurze Impulsvorträge vorbereiten. Im Zentrum einer solchen Veranstaltung sollte jedoch die gemeinsame Diskussion stehen. Damit sich Wissenschaft und Praxis annähern, ist der stetige Dialog unerlässlich. Hierbei erfahren Wissenschaftler:innen, welche Begriffe klar und welche gegebenenfalls erläutert werden müssen. Gleichzeitig erhält auch die Praxis tiefergehende Einblicke in die Fachsprache.

6. Ein Anwendungsbeispiel aus der Medienpädagogik für den Transfer der Ergebnisse einer Forschungssynthese und mögliche Erfolgsindikatoren

Der Transfer der Ergebnisse von Forschungssynthesen zur Verringerung des Researcher-Practitioner-Gaps wird nun anhand einer Metaanalyse zum Einsatz digitaler Medien im mathematisch-naturwissenschaftlichen Unterricht der Sekundarstufe (Hillmayr et al. 2020) dargestellt. Zunächst präsentieren wir ausgewählte Ergebnisse der Forschungssynthese und gehen anschliessend auf die konkret gewählten Transferformate ein. Wie sehr diese in der Lage sind, den Researcher-Practitioner-Gap zu verkleinern, reflektieren wir u. a. anhand von Erfolgsindikatoren wie Downloadzahlen und Evaluationsergebnissen.

Die Metaanalyse beschäftigt sich mit den Fragen (a) welchen Effekt der Einsatz digitaler Medien auf die Leistung und Einstellung von Schüler:innen hat und (b) unter welchen Bedingungen digitales Lernen besonders effektiv ist. Die Analysen zeigten, dass der Einsatz digitaler Medien einen mittelgrossen positiven Effekt ($g = 0.65$) auf die schulische Leistung und einen kleinen positiven Effekt ($g = 0.45$) auf die Einstellung der Schüler:innen zum Unterrichtsfach hatten. Darüber hinaus stellten die

Autor:innen fest, dass Schüler:innen, deren Lehrperson ein Lehrpersonentraining besucht hat, signifikant stärker vom Einsatz digitaler Medien profitierten. Moderatoren wie das Schulfach oder die Jahrgangsstufe zeigten keine signifikanten Effekte. Diese Ergebnisse sind für die Bildungspraxis bedeutsam, da sie eine Reihe von Kernbotschaften implizieren. Zunächst ist zu betonen, dass sich den Ergebnissen zufolge der Einsatz digitaler Medien lohnt. Dies ist eine zentrale Erkenntnis, da gerade bei Lehrpersonen diesbezüglich häufig Unsicherheit besteht. Mit dieser Metaanalyse kann nun ein entscheidendes Argument hervorgebracht werden, um mögliche Skeptiker:innen zu überzeugen. Weiter zeigte sich, dass der positive Effekt des Einsatzes digitaler Medien robust festgestellt werden konnte. Das Schulfach oder die Jahrgangsstufe spielten keine signifikante Rolle bezüglich der Größe des positiven Effekts. Eine ebenfalls zentrale Botschaft für die Bildungspraxis ist, dass sich die Schulung der Lehrpersonen zu lohnen scheint. Dies ist zum einen eine wichtige Nachricht für Multiplikator:innen, welche Lehrpersonentrainings anbieten, da sie so die Teilnahme fördern können. Zum anderen ist es auch eine wichtige Botschaft für die Lehrpersonen. Ihr Mehraufwand aufgrund des Besuchs einer solchen Veranstaltung scheint sich zu lohnen: Ihre Schüler:innen profitieren vom Einsatz digitaler Medien mehr als Schüler:innen, deren Lehrpersonen keine Fortbildung erhalten haben.

Die Ergebnisse der Forschungssynthese sind also relevant für die Bildungspraxis, da sie bedarfsgerecht Informationen über praktische Herausforderungen liefern (vgl. Knogler, Hetmanek, und Seidel 2022). Nun stellt sich die Frage, wie der Transfer gelingen kann, damit die Befunde den Researcher-Practitioner-Gap überwinden. Ausgehend von den in den vorherigen Kapiteln dargestellten Erkenntnissen, möchten wir einige Umsetzungsbeispiele zum Wissenschafts-Praxis-Transfer entsprechend den oben genannten Botschaften verdeutlichen.

Die Ergebnisse der Forschungssynthese wurden neben der wissenschaftlichen Publikation (Hillmayr et al. 2020) auch in einer Reihe weiterer Transferformate aufbereitet: Hierzu zählen Themenheft (Hillmayr et al. 2017), Kurzreview (Hetmanek, Munk, und CHU Research Group 2021), Erklärvideo¹ sowie Transferveranstaltungen. Im Folgenden stellen wir nun die Erfahrungen mit diesen Formaten zur Verringerung des Researcher-Practitioner-Gaps dar. Im Themenheft nennen und erklären die Autor:innen neben den Ergebnissen der Forschungssynthese unter anderem bekannte Lerntheorien (z. B. die Kognitive Theorie des multimedialen Lernens nach Mayer 2014), die einen positiven Effekt des Einsatzes digitaler Medien vermuten lassen. Zusätzlich bieten Interviews mit Expert:innen aus der Bildungspraxis zu den Forschungsfragen eine Möglichkeit, die Perspektive der Praktiker:innen zu hören und die Relevanz der Ergebnisse für die Praxis zu unterstreichen. Durch Gestaltungsvorschläge für Unterrichtssequenzen mit digitalen Unterrichtsmaterialien wird die Relevanz der Ergebnisse für den unterrichtspraktischen Alltag weiter unterstrichen.

1 <https://www.edu.sot.tum.de/suf/fuer-lehrkraefte/digitale-medien/>.

Das Themenheft steht zusammen mit einem Erklärvideo zur kostenlosen Nutzung auf der Website der Arbeitsgruppe zur Verfügung. Darüber hinaus wurden beide Formate über Social-Media-Kanäle beworben. Wie zielführend diese Transferformate waren, sieht man beispielsweise an den Downloadzahlen des Themenhefts. Über 65.000 Downloads zwischen 2017 und 2022 zeigen, dass sowohl das Thema als auch der Inhalt häufig nachgefragt werden. Zum anderen bestätigt die sich an die Veröffentlichung anschließende umfangreiche Berichterstattung in der Presse, dass das Themenheft ein geeignetes Format der Kommunikation ist und von der breiten Öffentlichkeit aufgegriffen wurde.

Das Clearing House Unterricht als Clearingstelle hat darüber hinaus die Forschungssynthese in einem Kurzreview (Hetmanek, Munk, und CHU Research Group 2021) aufbereitet und bewertet. Teil der Bewertung ist beispielsweise, wie differenziert die Ergebnisse dargestellt oder wie methodisch verlässlich die Befunde sind. Abschliessend beschreiben die Autor:innen des Kurzreviews in einem Fazit für die Unterrichtspraxis die zentralen Botschaften. Befragungen der Lesenden zeigen, dass Kurzreviews von Lehrerbildenden als nützlich und unterstützend für die Vermittlung von Wissen an Lehrpersonen wahrgenommen werden (Diery et al. 2020).

Wie im vorherigen Kapitel angesprochen, ist der Wissenschafts-Praxis-Transfer essenziell, damit die Ergebnisse der Forschungssynthese im Klassenzimmer ankommen. Deshalb wurde neben den zeitunabhängigen und rezeptiven Formaten Themenheft und Erklärvideo auch eine Transferveranstaltung entworfen, die den Dialog über das Thema ins Zentrum stellt. Zielgruppe waren Lehrpersonen unterschiedlicher Schularten. Hierbei präsentierten die Forschenden zunächst die Methode und Ergebnisse der Forschungssynthese. Sie konzentrieren sich hierbei auf die Formulierung klarer Take-Home-Messages. Gerade komplexe Zusammenhänge kurz und verständlich darzustellen ist eine Schwierigkeit, da häufig die Verkürzung und Vereinfachung von Aussagen auch dazu führt, dass Aussagen falsch verstanden oder ausgelegt werden können. Empfehlenswert ist dann, eine grosse Aussage in mehrere kleine zu zerlegen. Ein erster Schritt hinein in den Dialog sind Reflexionsfragen: Was zeichnet aus Sicht der Lehrpersonen einen gelungenen Einsatz digitaler Medien aus? Erleben Lehrpersonen auch in ihrer täglichen Arbeit die positiven Effekte des Einsatzes digitaler Medien im Unterricht? Und welche konkreten Praxisbeispiele zum Einsatz digitaler Medien sind bereits bekannt? Nach einer Kleingruppendiskussion, in der sich die Teilnehmenden aus der Praxis untereinander abstimmen können, kommen Wissenschaft und Praxis in einer Abschlussdiskussion miteinander ins Gespräch. Evaluationsergebnisse dieses Formats zeigen, dass die Lehrpersonen, die an diesem Format teilgenommen haben, es überwiegend als gut und empfehlenswert bewerten und angeben, dass dadurch der Austausch zwischen Wissenschaft und Praxis gefördert werden kann. Gleichzeitig berichten die Teilnehmenden aber auch von Schwierigkeiten bei der Herstellung von Bezügen in den schulischen

Alltag. Da aus unserer Sicht die Aufgabe der Wissenschaft, durch einen erfolgreichen Wissenstransfer und das Anstossen von Reflexionen erfüllt ist, braucht es für konkrete Handlungsveränderungen im Klassenzimmer weitere Partner:innen wie beispielsweise Landesinstitute, welche dann das vorhandene Wissen gemeinsam mit den Lehrpersonen nutzbar machen und Handlungswissen erarbeiten.

7. Fazit

Aus diesen Überlegungen ergeben sich aus unserer Sicht drei zentrale Botschaften, die wir in die Diskussion einbringen: (1) Forschungssynthesen können den Researcher-Practitioner-Gap verkleinern. (2) Der Wissenschafts-Praxis-Transfer ist hierbei essenziell. (3) Es kommt auf Wissenschaft *und* Praxis an.

Forschungssynthesen haben das Potenzial, den Researcher-Practitioner-Gap zu verkleinern. Indem sie schnell einen Überblick zu einem Thema liefern, differenzierte Befunde zur Generalisierbarkeit geben und durch ihr systematisches Vorgehen Transparenz und damit Vertrauen in die Wissenschaft schaffen, können sie einen Beitrag leisten, dass bildungswissenschaftliches Wissen seinen Weg ins Klassenzimmer findet. Dieser Beitrag soll daher dazu aufrufen, dass das Potenzial von Forschungssynthesen in der Forschung aber auch in der Bildungspraxis stärker wahrgenommen wird, wobei deren Aufbereitung und der explizite Wissenschafts-Praxis-Transfer von entscheidender Bedeutung ist.

Hindernissen wie der Abstraktheit oder der Änderung von Fehlvorstellungen können Forschungssynthesen vor allem mit geeigneten Transferformaten begegnen. Neben adressatengerechten Publikationen oder Erklärvideos tragen auch Transferveranstaltungen zu einem Austausch zwischen Wissenschaft und Praxis – und damit einer Verkleinerung des Researcher-Practitioner-Gaps – bei. Wichtig ist hierbei, dass Formate gefunden werden, die zur Zielgruppe der Bildungspraxis passen. Nur wenn die Bedürfnisse der Akteur:innen ausreichend abgebildet werden, können Überzeugungen geändert und Hindernisse abgebaut werden. Mit diesem Beitrag möchten wir deshalb dazu ermutigen, beim Transfer viele unterschiedliche Formate auszuprobieren, aus denen beispielsweise Lehrpersonen das für sie passende Format auswählen können. So kann die Heterogenität innerhalb der Zielgruppe bestmöglich berücksichtigt werden.

Weder das Wissen aus der Bildungspraxis noch das Wissen aus der Bildungsforschung alleine ist ausreichend, um gute Entscheidungen zu treffen (Gräsel und Fussangel 2022; Sackett et al. 1996). Entscheidend für einen guten Unterricht ist daher ein gelungener Austausch zwischen Bildungsforschung *und* Bildungspraxis. Daher sehen wir sowohl Wissenschaft als auch Praxis in der Verantwortung, sich auf die andere Seite einzulassen, deren Bedürfnisse zu sehen und sich schrittweise aufeinander zuzubewegen. Uns ist dabei bewusst, dass beide Seiten Restriktionen

ausgesetzt sind, die sie nicht einfach ablegen können: Die Fragen des Wissenschafts-Praxis-Transfers kommen häufig zum ohnehin hohen Arbeitspensum hinzu. Die damit verbundenen Aufgaben werden schnell als lästig und nicht notwendig erachtet. Widmen sich Forschende oder Lehrpersonen dennoch dem Wissenschafts-Praxis-Transfer, müssen sie nicht zuletzt Hürden in ihrer eigenen Community überwinden. Mit diesem Beitrag möchten wir aufzeigen, dass sich der zusätzliche Aufwand lohnt und ein wichtiger Nutzen, nämlich die Reduktion des Researcher-Practitioner-Gaps, daraus resultieren kann. Aus unserer Sicht haben Forschungssynthesen und der Wissenschafts-Praxis-Transfer ein erhebliches Potenzial, zukünftige Lern- und Arbeitsprozesse zu optimieren, wodurch der gegenwärtige Mehraufwand als Investition für später gesehen werden kann. Dieser Beitrag unterstützt daher den aktuellen Trend in Projektausschreibungen, welche gezielt den Wissenschafts-Praxis-Transfer als Anliegen aufnehmen und den Transfer damit fest in der Projektplanung verankern.

Zur Unterstützung dieses Wissenschafts-Praxis-Transfers spielen insbesondere auch die Landesinstitute und Qualitätseinrichtungen der Länder in Deutschland eine entscheidende Rolle (Bieber et al. 2018; Kultusministerkonferenz 2016). Als Expert:innen für Schule mit einem klaren Blick auf die Wissenschaft können sie eine Brücke bilden. Aus unserer Sicht wäre es wünschenswert, dass Landesinstitute und Qualitätseinrichtungen einzeln, aber auch in Zusammenarbeit miteinander fortlaufend Anstrengungen unternehmen, um Akteur:innen aus Wissenschaft und Bildungspraxis miteinander ins Gespräch zu bringen. Ein Beispiel hierfür ist die Veranstaltungsreihe Wissenschaft im Dialog des Instituts für Bildungsanalysen Baden-Württemberg.

Nutzt man die Potenziale, beugt Grenzen vor und stellt die Voraussetzungen für einen gelungenen Wissenschafts-Praxis-Transfer sicher, können Forschungssynthesen einen wertvollen Beitrag zur Verkleinerung des Researcher-Practitioner-Gaps leisten und damit einen Grundstein für guten Unterricht liefern.

Literatur

- Bauer, Johannes, Manfred Prenzel, und Alexander Renkl. 2015. «Evidenzbasierte Praxis – im Lehrerberuf?! Einführung in den Thementeil». *Unterrichtswissenschaft* 43 (3): 188–92.
- Beelmann, Andreas. 2014. «Möglichkeiten und Grenzen systematischer Evidenzkumulation durch Forschungssynthesen in der Bildungsforschung». *Zeitschrift für Erziehungswissenschaft* 17 (Supplement): 55–78. <https://doi.org/10.1007/s11618-014-0509-2>.
- Bieber, Götz, Eugen L. Egyptien, Günter Klein, Karin Oechslein, und Birgit Pikowsky. 2018. «Positionspapier der Landesinstitute und Qualitätseinrichtungen der Länder zum Transfer von Forschungswissen». https://www.qua-lis.nrw.de/cms/upload/aktuelles/Positionspapier_Transfer_31.10.18.pdf.

- Blatter, Kristine, Regine Schelle, und Stefan Michl. 2022. «Wissenschaft im Elfenbeinturm? Wissenstransfer im Elementarbereich aus Sicht der Forschenden», Hildesheim, 15. September.
- Booth, Andrew, Anthea Sutton, und Diana Papaioannou. 2016. *Systematic approaches to a successful literature review*. 2nd ed. Los Angeles: Sage.
- Borenstein, Michael. 2009. *Introduction to meta-analysis*. Chichester, U. K. John Wiley & Sons.
- Bramer, Wichor M., Gerdien B. de Jonge, Melissa L. Rethlefsen, Frans Mast, und Jos Kleijnen. 2018. «A systematic approach to searching: An efficient and complete method to develop literature searches». *Journal of the Medical Library Association: JMLA* 106 (4): 531–41. <https://doi.org/10.5195/jmla.2018.283>.
- Broekkamp, Hein, und Bernadette van Hout-Wolters. 2007. «The gap between educational research and practice: A literature review, symposium, and questionnaire». *Educational Research and Evaluation* 13 (3): 203–20. <https://doi.org/10.1080/13803610701626127>.
- Bromme, Rainer, Manfred Prenzel, und Michael Jäger. 2014. «Empirische Bildungsforschung und evidenzbasierte Bildungspolitik». *Zeitschrift für Erziehungswissenschaft* 17 (S4): 3–54. <https://doi.org/10.1007/s11618-014-0514-5>.
- Burgard, Tanja, Michael Bošnjak, und Robert Studtrucker. 2021. «Community-augmented meta-analyses (CAMAs) in psychology». *Zeitschrift für Psychologie* 229 (1): 15–23. <https://doi.org/10.1027/2151-2604/a000431>.
- Cohn, Lawrence D., und Betsy J. Becker. 2003. «How meta-analysis increases statistical power». *Psychological Methods* 8 (3): 243–53. <https://doi.org/10.1037/1082-989X.8.3.243>.
- De Corte, Erik. 2000. «Marrying theory building and the improvement of school practice: a permanent challenge for instructional psychology». *Learning and Instruction* 10 (3): 249–66. [https://doi.org/10.1016/S0959-4752\(99\)00029-8](https://doi.org/10.1016/S0959-4752(99)00029-8).
- Dehmel, Alexandra. 2017. *Bildungswissenschaftlicher und praktischer Transfer: Hintergrundpapier*.
- Diery, Annika, Maximilian Knogler, Claudia Mazziotti, Annika Schneeweiss, Andreas Hetmanek, Doris Holzberger, und Tina Seidel. 2020. «Das Clearing House Unterricht. Ein Service für die Lehrer*innenbildung?!». *Journal für LehrerInnenbildung*, 42–51. https://doi.org/10.35468/jlb-02-2020_03.
- Everton, Tim, Maurice Galton, und Tony Pell. 2000. «Teachers' perspectives on educational research: Knowledge and context». *Journal of Education for Teaching* 26 (2): 167–82. <https://doi.org/10.1080/02607470050127081>.
- Farley-Ripple, Elizabeth, Henry May, Allison Karpyn, Katherine Tilley, und Kalyn McDonough. 2018. «Rethinking connections between research and practice in education: A conceptual framework». *Educational Researcher* 47 (4): 235–45. <https://doi.org/10.3102/0013189X18761042>.

- Ferrero, Marta, Tom Elis Hardwicke, Emmanouil Konstantinidis, und Miguel A. Vadillo. 2020. «The effectiveness of refutation texts to correct misconceptions among educators». *Journal of Experimental Psychology: Applied* 26(3): 411-21. <https://doi.org/10.31234/osf.io/ehybj>.
- Gräsel, Cornelia. 2010. «Stichwort: Transfer und Transferforschung im Bildungsbereich». *Zeitschrift für Erziehungswissenschaft* 13 (1): 7–20. <https://doi.org/10.1007/s11618-010-0109-8>.
- Gräsel, Cornelia, und Kathrin Fussangel. 2022. «Modelle und Prozesse des Transfers – Verbindung von Theorie und Praxis». In *Sprachliche Bildung im Transfer: Konzepte der Sprach- und Schriftsprachförderung weitergeben*, herausgegeben von Michael Becker-Mrotzek, Hans-Joachim Roth, Johanna Griessbach, Nora von Dewitz, und Christiane Schöneberger, 23–36. Bildung durch Sprache und Schrift Band 6. Stuttgart: W. Kohlhammer.
- Gusenbauer, Michael, und Neal R. Haddaway. 2020. «Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources». *Research Synthesis Methods* 11 (2): 181–217. <https://doi.org/10.1002/jrsm.1378>.
- Herzog, Lisa. 2022. «Know-how für die Wissenschaftskommunikation: Wie Barrieren überbrückt werden können». [Know-how for science communication: How barriers can be bridged]. *Forschung & Lehre* 7: 538–39. https://www.wissenschaftsmanagement-online.de/system/files/downloads-wimoarticle/f%26l7-22_Know-how_fuer_die_Wissenschaftskommunikation_Herzog.pdf.
- Hetmanek, Andreas, Simon Munk, und CHU Research Group. 2021. «Digitale Tools im Unterricht. Welche Typen gibt es und wie kommen sie effektiv zum Einsatz? Kurzreview 27». www.clearinghouse-unterricht.de.
- Hillmayr, Delia, Frank Reinhold, Lisa Ziernwald, und Kristina Reiss. 2017. *Digitale Medien im mathematisch-naturwissenschaftlichen Unterricht der Sekundarstufe: Einsatzmöglichkeiten, Umsetzung und Wirksamkeit*. Münster: Waxmann.
- Hillmayr, Delia, Lisa Ziernwald, Frank Reinhold, Sarah I. Hofer, und Kristina Reiss. 2020. «The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis». *Computers & Education* 153:103897. <https://doi.org/10.1016/j.compedu.2020.103897>.
- Holzberger, Doris, und Lisa Ziernwald. 2020. «Forschungsflut – den Überblick behalten!». *Pädagogische Führung* 1: 25–27.
- Joram, Elana, Anthony J. Gabriele, und Katherine Walton. 2020. «What influences teachers' «buy-in» of research? Teachers' beliefs about the applicability of educational research to their practice». *Teaching and Teacher Education* 88:102980. <https://doi.org/10.1016/j.tate.2019.102980>.

- Joyce, Kathryn E., und Nancy Cartwright. 2020. «Bridging the gap between research and practice: Predicting what will work locally». *American Educational Research Journal* 57 (3): 1045–82. <https://doi.org/10.3102/0002831219866687>.
- Knogler, Maximilian, Andreas Hetmanek, und Tina Seidel. 2022. «Bestimmung und Bereitstellung der ‹best available› Evidenz für bestimmte Praxisfelder im Bildungsbereich». In *Optimierung schulischer Bildungsprozesse – What works?* herausgegeben von Nele McElvany, Michael Becker, Fani Laueremann, Hanna Gaspard, und Annika Ohle-Peters, 135–44. Dortmunder Symposium der Empirischen Bildungsforschung 6.
- Kultusministerkonferenz. 2016. «Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring». https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf.
- Kultusministerkonferenz. 2019a. «Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung: Beschluss der Kultusministerkonferenz vom 16.10.2008 in der Fassung vom 16.05.2019». https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf.
- Kultusministerkonferenz. 2019b. «Standards für die Lehrerbildung: Bildungswissenschaften: Beschluss der Kultusministerkonferenz vom 16.12.2004 in der Fassung vom 16.05.2019». https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf.
- Lassonde, Karla A., Panayiota Kendeou, und Edward J. O'Brien. 2016. «Refutation texts: Overcoming psychology misconceptions that are resistant to change». *Scholarship of Teaching and Learning in Psychology* 2 (1): 62–74. <https://doi.org/10.1037/stl0000054>.
- Lefebvre, C., J. Glanville, S. Briscoe, A. Littlewood, C. Marshall, M-I Metzendorf, A. Noel-Storr et al. 2019. «Searching for and selecting studies». In *Cochrane handbook for systematic reviews of interventions*, herausgegeben von Julian P. T. Higgins, James Thomas, Jaqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew J. Page, und Vivian A. Welch, 66–107: The Cochrane Collaboration; Wiley Blackwell.
- Levin, Ben. 2004. «Making research matter more». *EPAA* 12:56. <https://doi.org/10.14507/epaa.v12n56.2004>.
- Li, Tianjing, Julian P. T. Higgins, und Jonathan J. Deeks. 2019. «Collecting data». In *Cochrane handbook for systematic reviews of interventions*, herausgegeben von Julian P. T. Higgins, James Thomas, Jaqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew J. Page, und Vivian A. Welch, 109–42: The Cochrane Collaboration; Wiley Blackwell.
- Mayer, Richard. 2014. «Cognitive theory of multimedia learning». In *The Cambridge handbook of multimedia learning*, herausgegeben von Richard Mayer, 43–71: Cambridge University Press.
- Menz, Cordelia, Birgit Spinath, und Eva Seifried. 2021. «Misconceptions die hard: prevalence and reduction of wrong beliefs in topics from educational psychology among preservice teachers». *Eur J Psychol Educ* 36 (2): 477–94. <https://doi.org/10.1007/s10212-020-00474-5>.

- Murad, M. Hassan, Noor Asi, Mouaz Alsawas, und Fares Alahdab. 2016. «New Evidence Pyramid». *Evidence-based medicine* 21 (4): 125–27. <https://doi.org/10.1136/ebmed-2016-110401>.
- Pant, Hans Anand. 2014. «Aufbereitung von Evidenz für bildungspolitische und pädagogische Entscheidungen: Metaanalysen in der Bildungsforschung». *Zeitschrift für Erziehungswissenschaft* 17 (S4): 79–99. <https://doi.org/10.1007/s11618-014-0510-9>.
- Ratcliffe, Mary, Hannah Bartholomew, Vicky Hames, Andy Hind, John Leach, Robin Millar, und Jonathan Osborne. 2005. «Evidence-based practice in science education: the researcher-user interface». *Research Papers in Education* 20 (2): 169–86. <https://doi.org/10.1080/02671520500078036>.
- Renkl, Alexander. 2022. «Meta-analyses as a privileged information source for informing teachers' practice?». *Zeitschrift für Pädagogische Psychologie* 36 (4): 217–31. <https://doi.org/10.1024/1010-0652/a000345>.
- Sackett, D. L., William M. C. Rosenberg, J. A. Muir Gray, R. Brian Haynes, und W. Scott Richardson. 1996. «Evidence based medicine: what it is and what it isn't». *BMJ* 312: 71–72.
- Seel, Norbert M. 2017. «Model-based learning: a synthesis of theory and research». *Education Tech Research Dev* 65 (4): 931–66. <https://doi.org/10.1007/s11423-016-9507-9>.
- Sibanda, Jabulani, und Begede, Martin P. 2015. «Extent of ESL teachers' access to, utilisation and production of research». *South African Journal of Education* 35 (2): 1–12. <https://doi.org/10.15700/saje.v35n3a1080>.
- Siddaway, Andy P., Alex M. Wood, und Larry V. Hedges. 2019. «How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses». *Annual Review of Psychology* 70: 747–70. <https://doi.org/10.1146/annurev-psych-010418-102803>.
- Thomm, Eva, Christine Sälzer, Manfred Prenzel, und Johannes Bauer. 2021. «Predictors of teachers' appreciation of evidence-based practice and educational research findings». *Zeitschrift für Pädagogische Psychologie* 35 (2-3): 173–84. <https://doi.org/10.1024/1010-0652/a000301>.
- Vanderlinde, Ruben, und Johan van Braak. 2010. «The gap between educational research and practice: views of teachers, school leaders, intermediaries and researchers». *British Educational Research Journal* 36 (2): 299–316. <https://doi.org/10.1080/01411920902919257>.
- Voracek, Martin, Michael Kossmeier, und Ulrich S. Tran. 2019. «Which data to meta-analyze, and how? A specification-curve and multiverse-analysis approach to meta-analysis». *Zeitschrift für Psychologie* 227 (1): 64–82. <https://doi.org/10.1027/2151-2604/a000357>.
- Webster, Jane, und Richard T. Watson. 2002. «Analyzing the past to prepare for the future: Writing a literature review». *MIS Quarterly* 26 (2): 13–23.

Weckmann, Gesine, Jean-François Chenot, und Katrin C. Reber. 2015. «Metaanalysen lesen und interpretieren: eine praktische Anleitung». *Zeitschrift für Allgemeinmedizin* 91 (11): 469–73.

Wissenschaftsrat. 2016. *Wissens- und Technologietransfer als Gegenstand institutioneller Strategien: Positionspapier*. (Drs. 5665-16).

Wissenschaftsrat. 2021. *Wissenschaftskommunikation: Positionspapier*. (Drs. 9367-21).

Woerfel, Till, Karin Zimmer, und Annika Witte. 2021. «Evidenzinformierte Bildungspolitik und Bildungspraxis: Das Potenzial von Forschungssynthesen». In *Vermessen? Zum Verhältnis von Bildungsforschung, Bildungspolitik und Bildungspraxis*, herausgegeben von David Kethofer, Johannes Reiting, und Katharina Soukup-Altrichter, 161–73. Beiträge zur Bildungsforschung Band 7. Münster, New York: Waxmann.

Förderhinweis

Das dieser Publikation zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung, und Forschung sowie des Sekretariats der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland unter dem Förderkennzeichen ZIB2022 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor*innen.

Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

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

Publication Status as a Common Inclusion Criterion in Systematic Reviews

On the Use of Digital Media in the Context of Teaching Pupils with SEN in Inclusive, Integrative or Segregated Settings

Claudia Mertens^{1*} , Carolin Quenzer-Alfred^{2*} , Anna-Maria Kamin¹  und Daniel Mays² 

¹ Universität Bielefeld

² Universität Siegen

*In geteilter Erstautorenschaft

Abstract

This paper critically discusses the use of publication status as a common inclusion criterion for systematic reviews, a method adopted from the field of medicine into education and media-related education studies. Two systematic reviews exploring the use of digital media in inclusive, integrative or segregated teaching settings are compared. By adding peer-reviewed grey and non-peer reviewed literature in a second review, the initial corpus of 15 studies (articles in journals with peer review) was increased by another 19 studies (without peer review). The advantages and disadvantages of including research articles with a different publication status in systematic reviews is discussed based on the comparison of both reviews. Overall, in both reviews, the focus was on individual support with digital media while teaching, especially within technology-based learning classes, mostly from a quantitative perspective. Both reviews revealed a didactic focus (learning with digital media) and hardly any study dealt with learning about digitalization.

Der Publikationsstatus als gängiges Einschlusskriterium in systematischen Übersichtsarbeiten. Über den Einsatz digitaler Medien in inklusiven, integrativen und exkludierenden Unterrichtssettings für Schüler:innen mit sonderpädagogischem oder zusätzlichem Förderbedarf

Zusammenfassung

Der Beitrag setzt sich kritisch mit der Nutzung von Publikationsmodi als Einschlusskriterium aus der Medizin für systematische Übersichtsarbeiten in der (Medien-) Pädagogik auseinander. Es werden zwei systematische Übersichtsarbeiten zum Einsatz digitaler Me-



dien in inklusiven, integrativen oder exkludierenden Settings verglichen. Durch die Hinzunahme peer-reviewter grauer und nicht peer-reviewter weisser Literatur in einem zweiten Review wurde der ursprüngliche Korpus von 15 Studien (Forschungsartikel in Fachzeitschriften mit Peer Review) um weitere 19 Studien (ohne Peer Review) erweitert. Insgesamt lag bei beiden Studientypen der Fokus auf der individuellen Unterstützung mit digitalen Medien im Unterricht – insbesondere im Rahmen von technologiebasiertem Unterricht, meist aus einer quantitativen Perspektive. Beide Übersichtsarbeiten zeigten einen didaktischen Fokus (Lernen mit digitalen Medien) und kaum eine Studie beschäftigte sich mit dem Lernen über Medien.

1. Introduction¹

Systematic reviews identify, evaluate and summarize findings according to pre-specified criteria in order to answer a research question. They originate from the medical discipline and aim to minimize bias by using explicit, systematic methods documented *a priori* with an operationalizable research concept of the research question and methods used – including search strategy, selection and evaluation criteria (documented in a protocol) (Gough, James and Sandy 2012). The method thus tries to strive for systematic reproducible research strategies so that, ideally, all empirical studies that meet the inclusion criteria are identified via the search string. It tries to give a quick and objective overview on the ever-growing amount of literature available on specific research topics (Pollock et al. 2022).

A key feature that distinguishes systematic reviews from literature reviews is the pre-established list of criteria for inclusion and exclusion of literature into the corpus (McKenzie et al. 2022; Haddaway et al. 2015). These criteria must be so transparent and consistent that the review could be replicated or updated at any time. Aside from the PICO elements – PICO stands for Population, Intervention, Comparison, and Outcomes (Higgins and Green 2013) – a common inclusion criterion is based on publication status. Although most guidelines on systematic reviews stress the importance of including results reported in all types of publications to reduce bias (cf. Cochrane, chapter 7), grey literature and research published without peer review are often excluded in the research body (i.e. La Paro and Pianta 2000; Hartling et al. 2017; Haddaway and Bayliss 2015). In library and information science, the term «grey literature» refers to academic information that is not commercially published and often not widely disseminated like government documents or unpublished reports (Haddaway and Bayliss 2015). It is typically not peer-reviewed, but there may be some exceptions (i.e. conference proceedings and dissertations). Grey literature can provide valuable information, particularly in fields where traditional journal

1 The authors wish to thank Melissa Bond for her valuable feedback and comments on the article.

publication is uncommon. In contrast, «white literature» refers to commercially published and widely disseminated scientific information, including peer-reviewed and non-peer-reviewed journal articles, monographs, and book chapters (Tab. 1).

	Peer review	No peer review
White literature	Research articles in journals with external peer review system	Journal articles, monographs, book chapters in edited volumes
Grey literature	Dissertations	Working reports, unpublished research reports

Tab. 1: Examples of grey/white literature (table based on Schulte and Zimmer, unpublished).

Beyond the challenges of finding grey literature and non-peer reviewed studies, it is often argued that journal articles published with peer review are generally of the highest quality standard. However, including only those publications is controversially discussed by what is called «non-reporting biases» (Page, Higgins, and Sterne 2022). Non-reporting bias refers to the phenomenon of not reporting or publishing certain findings due to lack of significance or unexpected results. Common guidelines for systematic reviews, like the Cochrane Handbook, also strongly recommend including grey literature.

For the field of education and media education studies, another problem arises if only peer reviewed studies are included, as the discipline may rely more heavily on grey and non-peer reviewed white literature than others (Birukou et al. 2011; Shaw, Phillips and Gutiérrez 2022). Excluding this type of literature in systematic reviews in the field may lead to a lack of representation and understanding of a research field (Birukou et al. 2011; Schulte and Zimmer unpublished). This article therefore explores how the output of systematic reviews differs if grey and non-peer reviewed research is included, by investigating and comparing two systematic reviews undertaken by the research team. The topic of these reviews was the use of digital media² for pupils with additional or special educational needs in inclusive, integrative or segregated teaching settings. When conducting the systematic reviews, the question arose of whether the criterion of peer review, stemming from the natural sciences such as medicine with a strong focus on evidence-based intervention studies, was 1:1 transferrable to education sciences with their more explorative approach to science (Birukou et al. 2011). In the end, it was decided to split the research process into two parts: review one would include only journal articles with peer review and review two would only include grey or non-peer reviewed studies to allow for a systematic comparison of the inclusion criterion. The research questions of this article can therefore be summarized as follows:

² We use the term «digital media» throughout the text knowing that terms vary culturally. In the German-speaking context the term goes beyond «digital tools». We refer to «digital tools» as well as to digital learning and to critical thinking about media etc.

- Is peer review in journal articles a valid and appropriate inclusion criterion in systematic reviews to comprehensively synthesise the research in education studies?
- How do the results of two systematic reviews differ when using an identical research string, when one review only includes journal articles with peer review and the other includes all types of literature?

2. Selection criteria in systematic reviews

The systematic selection of literature is carried out in a systematic review consistently, according to clearly determined rules, which must be defined in advance based on the research question and based on the theoretical-conceptual references. The determination of these criteria is part of the scientific process to concentrate the screening of scientific texts on the set topic and must therefore also be presented in the paper as transparently as possible. The list of inclusion criteria describes the characteristics that a study should fulfill in order to be included into the results (Stern, Jordan and McArthur 2014). The exclusion criteria describe the characteristics that lead to the exclusion of publications (ibid.). A balance must be found between selection rules that are too narrow and too specific (with the risk of excluding relevant studies) and criteria that are too broad (which would lead to the inclusion of too many irrelevant studies). During the search process, these criteria may be refined; however, adaptations must be protocolled. A change in criteria usually occurs when the search string delivers hits that are excessive, insufficient, or irrelevant to the research topic.

One of the main advantages of using clear and specific selection criteria in educational systematic reviews is that they help to minimize biases in the process (McKenzie et al. 2022; Haddaway and Bayliss 2015). Biases can occur when studies that do not meet the established criteria are included in a review, or when studies that meet the criteria are excluded. Using selection criteria also helps to ensure that the studies included in the review have similar characteristics and that the intervention, or for example the teaching method being evaluated, is the same across all studies (Cherry and Dickson 2014). This increases the internal validity of the systematic review, which means that the results of the review can more confidently be attributed to the subject being evaluated, rather than to other factors.

Common selection criteria in educational systematic reviews involve the use of the PICO framework (Brunton et al. 2017). Other common selection criteria refer to language, publication date, study design, type of publication (i.e. only original studies) and publication status (McKenzie et al. 2022). The latter refers to the selection rule of whether only peer reviewed journal articles or also studies with different publication modi are included. Since peer review by independent field-specific

and methodological experts commonly serves as a control for journal articles to achieve the highest possible quality through review and helps editors to judge the relevance of scientific manuscripts for high-impact scientific professional journals (Müller 2009), peer review is a frequently used and widely accepted inclusion criterion (Hartling et al. 2017). Its aim is to «ensure the validity of research, opinions regarding the reliability, thoroughness, and appropriateness of reports on research findings» (Hall et al. 2019: 295; cf. Katzav and Vaesen 2017; cf. Jukola 2017). All in all, the aim of peer review is a «check for accuracy, appropriate methodology, novelty, and relevance to the journal» (Rowley and Saffi 2018: 644; cf. Hall et al. 2019). To put it in other words, it is a system designed to decide whether manuscripts are qualified for publication and to detect potential flaws that those manuscripts might still have (Kirman, Simon and Hays 2019). It exists to ensure the accuracy of information as well as the quality of presentation and writing (Horbach and Halfman 2018), and it shall identify fraud (Mulligan, Hall and Raphael 2013). Messy articles shall be cleaned up and reorganized, and «invalid articles» rejected (ibid.). As Lee et al. (2013) summarize, it is a tool for the world of science to take social responsibilities seriously and to make science a self-regulated and normatively driven field.

For systematic reviews, the inclusion of peer reviewed journal articles is a common selection rule because the publications have undergone a rigorous and external process of review and quality control, which increases the likelihood that the information is accurate and reliable (Horbach and Halfman 2018). Aside from quality reasons, excluding grey and non-peer reviewed literature from systematic reviews can also help to reduce the number of sources to be reviewed, making the process more feasible. Additionally, peer reviewed literature is often indexed in databases, making it easier to locate and to include it in a systematic review. However, excluding non-peer reviewed from a systematic review can have several disadvantages (McKenzie et al. 2022; Chow and Eckholm 2018). Firstly, grey and non-peer reviewed literature may contain important information that is not found in peer reviewed literature. For example, government reports, conference proceedings, and theses may contain data and findings that have not yet been published in a peer reviewed journal. Excluding this information can lead to a biased or incomplete understanding of a research question. Secondly, grey or non-peer reviewed literature may be published by smaller or less well-known organizations, which may be overlooked in a search process that focuses solely on peer reviewed literature. This can lead to the exclusion of valuable information that may be relevant for the research question. Thirdly, by only including peer reviewed journal articles, evidence selection bias – referring to the phenomenon where certain studies or data are selectively included or excluded from a review or meta-analysis based on their results or conclusions (Drucker, Fleming and Chan 2016) – might be increased. Researchers may miss important studies that have not yet been published, that were rejected by peer

reviewed journals but still contain valuable information for a systematic review. Studies might have been refused in the review process due to some kind of bias or «publication code» inside the field and then they may potentially be published without peer review as a second step (i.e. Bond 2018).

In order not to lose these insights, systematic reviews profit from a complete overview of the research field. Critics argue, inter alia, that possible inadequacies of reviews may be due to a lack of training which may potentially lead to unnecessarily harsh criticism or a rejection of innovative results (Ferguson 2020). Peer review is also controversially discussed with reference to racial bias (Hojat, Gonnella and Caellegh 2003), language bias and specialty bias (Armstrong 1997; Goldbeck-Wood 1999). Using peer review and bibliometrics (e.g. citation counts and impact factors) as quality indicators of research could threaten small and highly specialized research (Henderson, Shurville and Fernstrom 2009). Aside from this, according to the Matthew Effect, people give disproportionate credit to those researchers who are already well known and highly cited (Heesen and Bright 2021). There are several studies suggesting a gender bias favouring male researchers (e.g. Birukou et al. 2011), although there does not seem to be a consensus in the current state of research (Hojat, Gonnella and Caellegh 2003).

Grey literature may also be more accessible to specific communities, such as studies that were conducted in low-income countries, where peer reviewed journals may not be as prevalent (Salager-Meyer 2008). Excluding these sources can lead to a lack of representation and diversity in the results of the systematic review. Finally, the argument that valuable findings might be lost seems to be even more relevant in the humanities, because – as Johnson and Hermanowicz (2017) critically claim, «perceptions of truth vary more in the humanities than in the sciences, and thus it is more problematic to study peer review» (p. 516). For the field of education and ICT education studies, another problem arises: Not every piece of research is published as a journal article and even then, not all journals have established a peer review process for quality control yet. For example, Kuhberg-Lasson, Singleton and Sondergeld (2014) showed that between 1995 and 2008, over 50% of the outputs generated in education studies were published without peer review. The key position of non-peer reviewed articles in education studies was also confirmed in a survey (n=643) by Bambey (2016). Although peer reviewed journals were rated as important for academic work by 93% of respondents, a large proportion (60.51%) also attributed high relevance to non-peer reviewed journals. Taken together with the problem of publication bias, the question arises if peer reviewed white literature is a valid and appropriate inclusion criterion in systematic reviews to comprehensively capture the state of research in education studies.

Apart from this, an important decision must be made according to whether the systematic review is conducted in the field of media education or not. Historically, peer review is the most common instrument for quality control of scientific manuscripts in STEM³ disciplines and is also often considered as an important standard of good scientific practice in the SSH⁴ fields. It can be defined as a process in which an author's research is being assessed and evaluated by field-specific and methodological experts (peers). The reviewees are thereby encouraged to meet the recognised high standards of their discipline. The underlying paradigm is rather developmental (in the sense of improvement) than controlling and is meant to promote knowledge-transfer between experts. The STEM disciplines seem to have almost universal and consistent standards while the former seem to have varying evaluation procedures. Taken together with the problem of publication bias and publication culture, the question arises if peer reviewed white literature is a valid and appropriate inclusion criterion in systematic reviews to comprehensively capture the state of research in education studies. This will be part of the following analysis.

3. Comparison of two systematic reviews with/without the inclusion criterion «peer review journal article» on media education for pupils with additional or special educational needs in inclusive, integrative or segregated teaching settings

The above-mentioned blind spot of scientific research on the impact of including grey and non-peer reviewed literature in educational systematic reviews shall be enlightened by the study at hand – comparing two systematic reviews using an identical search string for the database search, the only difference being that one analysis focussed on empirical articles with peer review and the other one focussed on non-peer reviewed and grey studies. The theoretical framework shall be introduced before descriptively outlining and discussing the methodology and results of the systematic reviews with regard to research question, search strategy, selection rules and a comparison of findings.

The two systematic reviews at hand aggregate empirical research on the use of digital media in the context of teaching pupils with special educational needs (SEN) in inclusive, integrative or self-contained settings. From a normative point of view, the curricular basis for integrating digital media in German schools are two papers published by the Standing Conference of the Ministers of Education and Cultural Affairs (KMK 2017 and 2021) stating that pupils shall acquire media competence throughout their school career. The documents emphasize learning *about* media and the demand for integrating digital teaching concepts and materials in school

3 Science, technology, engineering, mathematics.

4 Social sciences and humanities.

(learning *with* media). The assumption is that media have positive effects and offer the chance of individualization (cf. Kamin and Schluchter 2018) on condition that technologies and educational goals fit well together (Irion and Scheiter 2018: 10). The inclusive use of digital media can provide social participation and empower pupils with SEN.

Participation shall be created on three levels (cf. Bosse et al. 2019):

- participation in media (representation of all groups in media),
- participation at media (free access to media content – e.g. via assistive technology or because the principles of the universal design for learning (www.cast.org) are respected) and
- participation through/via media (via media competence).

Considering the state of the art concerning teaching *with/about* media in schools, we find that the results of the scientific research are unclear (cf. Herzig 2014; Schaumburg 2019), especially for students with additional and/or special educational needs. There are some studies on the computer-based learning of students with «low to average» level of disability (Fitzgerald, Kourey, and Mitchem 2008). Some studies focus on computer-assisted instruction for autistic students (Pennington 2010). Some meta-analyses and systematic reviews do exist on the international level, as e.g. the review by Sánchez-Serrano et al. (2020), who examined the use of ICT for learners with disabilities in 96 studies from 2009 to 2019 or the meta-analysis by Aspiranti, Larwin, and Schade (2020). Yet, all in all, it must be said that the potential of digital media for inclusion has not been sufficiently dealt with in scientific research. The existing studies either focus on a limited spectrum of SEN-areas (e.g. autism) and/or on a hardware-oriented perspective. This led us to the conclusion that a review on the research on the studies published in German was due – both *with* and *without* peer review.

4. Methodology used in our systematic reviews

To compare how the inclusion criterion «peer reviewed journal articles» impacts the output generated in a systematic review in the context of media-related education studies, two systematic reviews on digital learning in teaching students in inclusive and special educational contexts were conducted based on the criteria of the PRISMA group (Preferred Reporting Items for Systematic reviews and Meta-Analyses) established by Moher et al. (2009) and guided by the Cochrane guidelines for conducting and evaluating systematic reviews (Higgins et al. 2019). The research

question, search string⁵, inclusion criteria and the research team were identical in both reviews, the only difference being the inclusion criterion «articles with peer review» (research period: 15.09.-25.11.2020) and «articles without peer review and grey literature» (research period: 07.12.-18.03.2021). The search string was developed for the FIS database and then adapted, with eight national and international databases relevant to education studies searched with a pre-defined search string. The databases were FIS, ERIC, PubPsych, Education Source, PsychINFO, PsychARTICLES, Psychology & Behavior, and the Web of Science. For the second review, databases that specialize in grey literature were added, namely Pedocs, BASE, Open Grey and OpenDissertations.

While the former databases are generally known, the databases for the second review shall be briefly described in terms of their range and thematic focus. PEDOCS⁶ is a database for educational literature operated by the Leibniz Institute for Educational Research and Educational Information. PEDOCS contains several thousand documents, mainly from the fields of educational sciences, educational psychology, social sciences and educational research (pedocs, n. d.). The documents include dissertations, theses, post-doctoral theses, conference papers and reports. PEDOCS has a thematic focus on Germany and Europe, but there are also international documents in the database. BASE⁷ (Bielefeld Academic Search Engine) is one of the largest and most widely used search engines for grey literature (University of Wolverhampton n. d.). It is operated by Bielefeld University and contains over 100 million documents from various sources such as repositories, libraries and publishers. BASE has a broad thematic focus and covers many subject areas, including science, technology, medicine, social sciences, humanities and more. Documents in BASE include various types of materials such as scholarly articles, books, conference papers, dissertations, reports and more. OpenGrey⁸ is a database of grey literature in Europe run by the European Commission's Joint Research Centre (JRC) (University of Wolverhampton n. d.). OpenDissertations⁹ is a database that provides

5 (DIGITAL* OR MEDIEN* OR TECHNOLO* OR VIRTUELLE OR COMPUTER* OR MULTIMEDIA* OR TABLET OR APP OR LERNPLATTFORM OR INTERNET* OR SOFTWARE OR MOBILE OR CYBER OR GAME) AND (SCHUELER OR UNTERRICHT OR LERN* OR LEHR* OR SCHUL*) AND (SONDERPAED* OR FOERDERPAED* OR SONDERSCHUL*OR BEEINTRAECHTIG* OR HANDICAP* OR BEHIND* OR «GEISTIGE BEHINDERUNG» OR KOERPERBEHIND* OR LERNBEHINDERUNG OR «EMOTIONALE ENTWICKLUNG» OR «SOZIALE ENTWICKLUNG» OR ESE OR ADHS OR ADS OR AUTIS* OR «GEISTIGE ENTWICKLUNG» OR «KOERPERLICHE ENTWICKLUNG» OR «MOTORISCHE ENTWICKLUNG» OR SCHWERHOERIG* OR GEHOERLOS* OR HOERSCHAED* OR TAUB* OR BLIND* OR INKLUSI* OR BARRIERE* OR HETEROGEN* OR TEILHABE OR PARTIZIPATION).

6 <http://pedocs.de>.

7 <https://www.base-search.net>.

8 <https://opengrey.eu>.

9 <https://biblioboard.com/opendissertations/>.

free access to electronic versions of dissertations and theses submitted to accredited universities in the US. It is operated by the non-profit organisation EBSCOhost (EBESCO 2023).

The studies were evaluated according to seven inclusion criteria, namely, (1) publication and data collection between 2010 and 2020, (2) German language, (3) original empirical studies with peer review published in journals (review I) and original empirical studies without peer review published in journals or as (part of) a doctoral thesis or in a conference manual (review II). (4) The considered population were students of a general or vocational school or of a school for special education needs and the (5) intervention was school education *with* and/or *about* digital media. (6) The outcome should refer to additional or special educational needs and (7) the type of study should be empirical studies (quantitative, qualitative or mixed methods).

Using the same search string, a total of 5.565 original studies were identified in review I. In review II 9.186 records were found (see Figure 1). After removal of duplicates and screening, 181 articles with peer review and 77 articles without peer review were assessed on the basis of full texts. Finally, 15 studies with peer review and 19 studies without peer review were selected for data extraction in the reviews. For each included publication, basic information was extracted with a coding strategy adapted to the research question (Mertens and Quenzer Alfred et al. 2022; Quenzer-Alfred and Mertens et al. 2023). A quality assessment was conducted for studies that had not been peer reviewed. Quantitative studies were assessed according to the quality criteria of the Cochrane Collaboration (Higgins et al. 2019) and qualitative studies were reviewed for research design, sampling, data collection, data analysis, reporting, ethics, documentation and reflexivity and neutrality, taking the Spencer et al. (2003) assessment framework for qualitative studies into account.

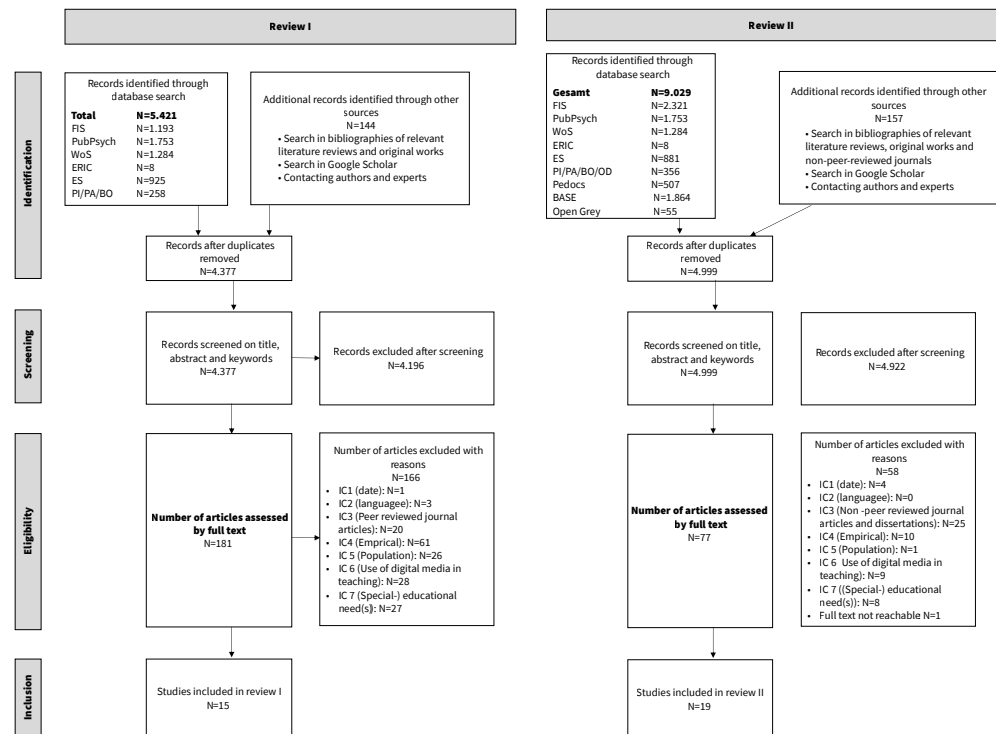


Fig. 1: PRISMA Flow-chart.

5. Results of the two systematic reviews

In total, 34 studies were found between the two systematic reviews. 31 studies were conducted in Germany, one in Austria and two in Switzerland. Between 2010 and 2020 there was no increase in publications. So, the first conclusion to be drawn is from a quantitative perspective: Not having taken the journal articles without peer review and dissertations into account, more than half of the findings would have been missed. Hence, to get a comprehensive view of the research field, screening studies with and without peer review as well as grey literature seems to be obligatory. Interestingly, most studies in both reviews used quantitative research methods, whereas just two studies used qualitative methods in review I and 5 studies in review II (see Figure 2).

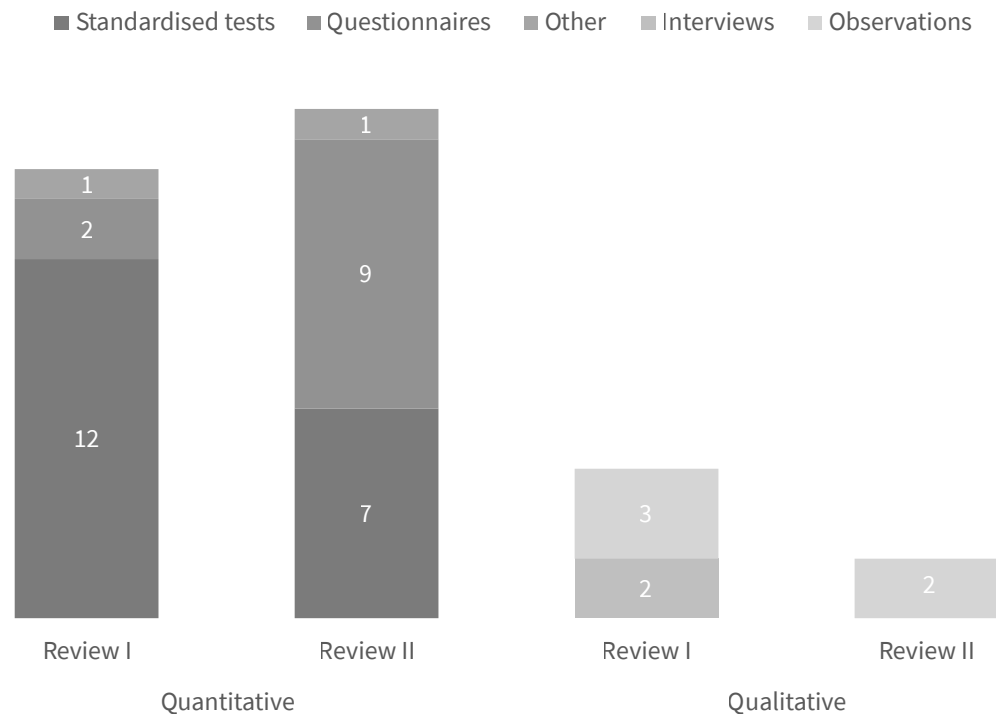


Fig. 2: Comparison of methods used in systematic review I and II.

The research methods which were used ranged from standardized tests, to questionnaires, interviews, observations and other research methods. In review I twelve studies were found using standardized instruments whereas in review II there were slightly more studies based on questionnaires (9), whereas only 7 studies used standardized instruments. In total, 27 quantitative studies, six qualitative studies and one mixed methods study were identified. However, generalisations cannot be made on such a narrow corpus basis.

As far as the type of schools is concerned, the whole range of school types was addressed except for vocational school (see Figure 3). Considerably more studies in special needs schools (11) were found in review II, in comparison to only four special needs school studies with peer review. It can be said that in both reviews, the emphasis was on individual learning to compensate for deficits, rather than on group learning with a common topic.

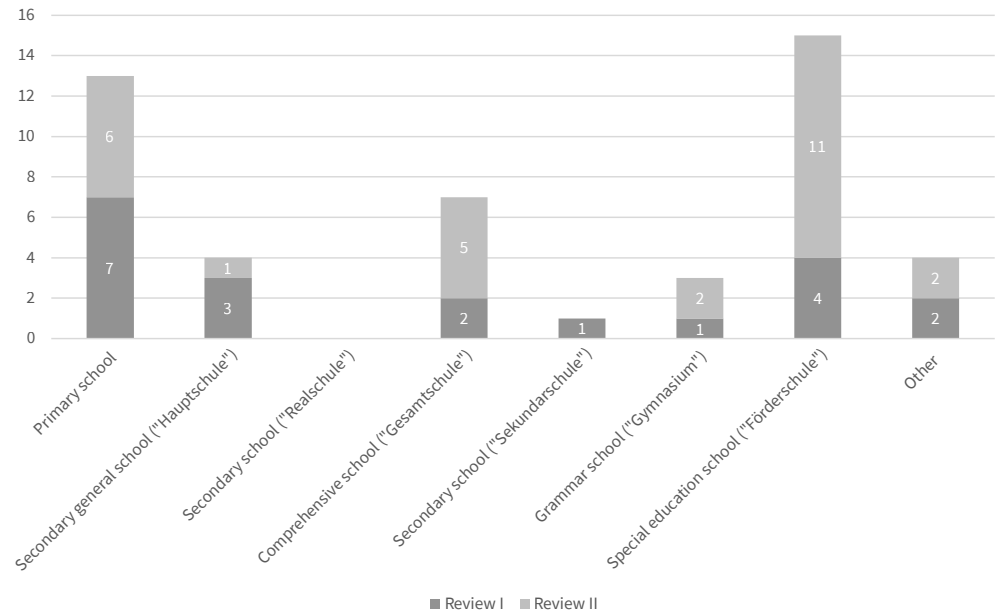


Fig. 3: Distribution of school types.

Screening the included studies according to whether the context of the study was inclusive, integrative or segregating it can be summarized that only nine studies took place in inclusive settings, whereas the context was integrative in thirteen studies and segregating in twelve studies (see Figure 4). Journal articles with peer review focussed on integrative settings, while articles without peer review and grey literature had more records for segregating but also inclusive settings.

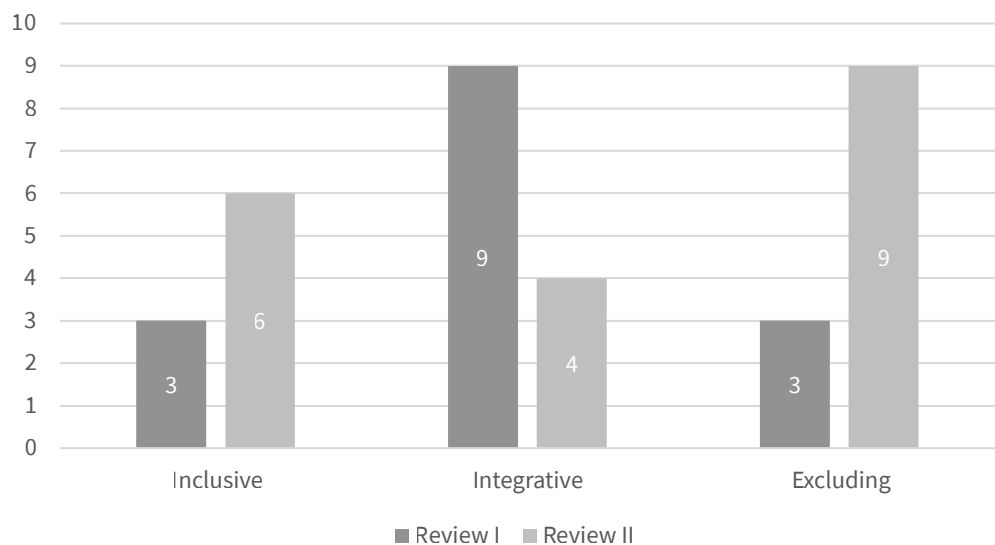


Fig. 4: Distribution of teaching settings.

A clear focus on evaluating training programs for the individual support of pupils can be found. The use of assistive technologies, as well as measures to improve accessibility, were considered very rarely, although more records were still found in articles without peer review and dissertations (see Figure 5).

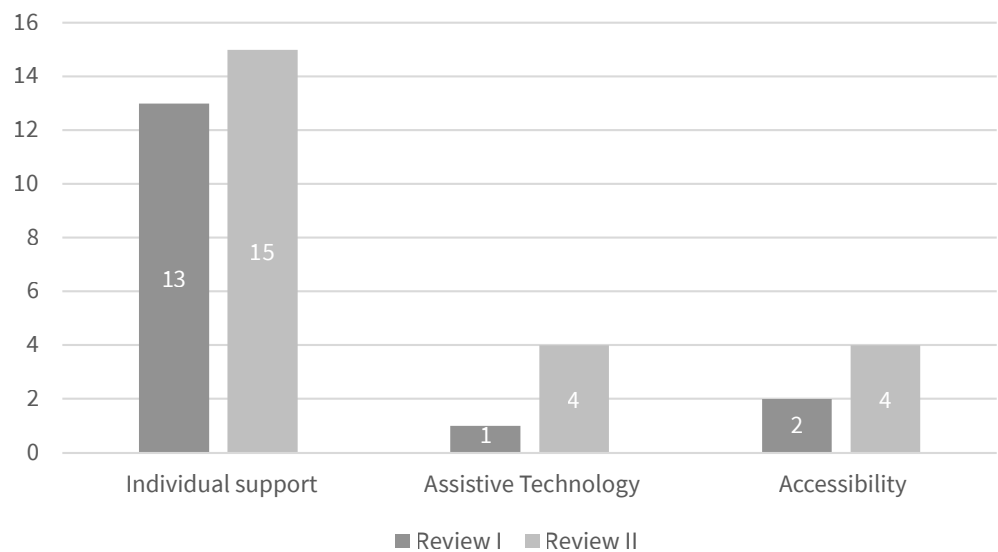


Fig. 5: Special educational function of digital media.

With regard to participation *in*, *at* and *through* media, the focus of empirical studies is on participation *through* media – regardless of publication status (figure 6).

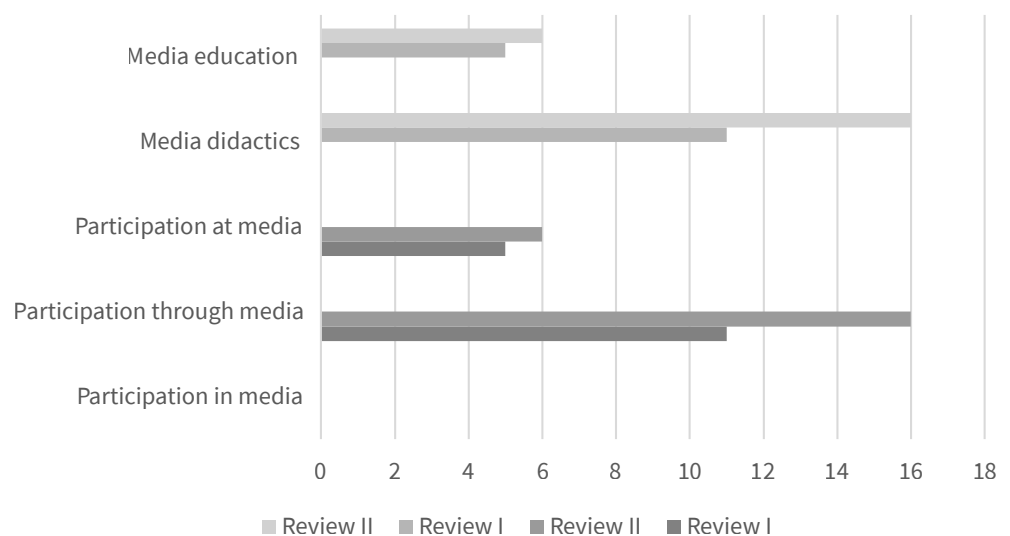


Fig. 6: Participation in, at and through media in the reviews.

Studies that examine the use of digital media for participation *at* media are underrepresented in both reviews. No study on participation *in* media could be identified at all. Hence, the focus of German-language research on the use of digital media for pupils with an additional or special educational focus is on the compensatory use of digital media. In other words, digital media are used to compensate for deficits in the treatment of special needs. In addition, it can be concluded for German-speaking research that there are only a few studies on learning *through/via* media and that, in consequence, there is a very clear research desideratum with regard to the acquisition of media competence in the context of special educational needs and inclusion (see Figure 6). A necessary interconnection between the didactic use of digital media and media competence acquisition is also missing in both reviews. Instead, the main research interest of the literature lies in learning *with* or *through* media, i.e. the use of digital media.

Looking at different learning and support areas, the majority of studies addressed performance deficits in reading, numeric skills and writing (n=14) (see Figure 7). Comparing the two reviews it can be seen that articles without peer review and dissertations (review II) were more likely to address the area of special need «learning», «communication» and «seeing» whereas for the other areas there were just as many articles with peer review as without. Few studies addressed topics independently of specific learning areas with broader research questions and goals, i.e. related to inclusive classroom environments (see Figure 7).

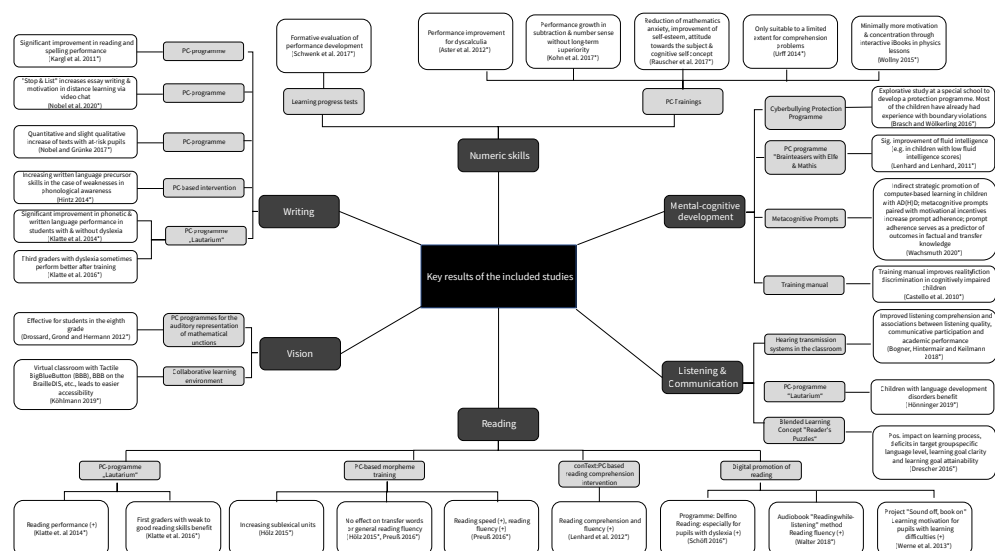


Fig. 7: Key findings with categorized by learning areas.

Author (year)	Key findings
Bosse (2018)*	Quality criteria for the design/use of inclusive teaching media; audiovisual media = high potential for connecting with pupils' interests; important: action and production-oriented tasks, cooperative & collaborative learning
Michna et al. (2016)*	Design of inclusive learning environments using the Universal Design for Learning significant knowledge gain in subject knowledge tests; materials handled equally well by all groups; learners in the inclusive classes rated the unit as more positive in contrast to the mainstream class
Mihajlovic (2012)*	Frequency of use and type of use of computers and the internet in teaching practice at special schools; use regularly and usually several times a week; frequency varies according to subject; computers in lessons mainly for word processing (22%) and internet use (21%); internet use by teachers varies in frequency; use within lessons primarily for information research (54%)
Wicki and Burkhardt (2018)* in Calabrese et al. (2018)	Use of assistive technologies (AT) to support learning and educational processes; mostly use of analogue AT; if electronic ones are used, then those already known from everyday life (smartphones, tablets, etc.); technologies from the field of vision hardly ever used
Müller and Fleischer (2013)*	Media literacy training; media literacy is more likely to be promoted if there is an assumption that pupils have or can acquire the necessary cognitive skills; practical relevance of media is decisive; increasing use in the higher classes; use of media in lessons depends mainly on teachers' own initiative
Schiedek and Menges (2018)*	Evaluation of the method of interactive video annotation in inclusive physical education; quality of the contributions of students with special needs corresponds to the average of the learning group; digitally supported communication: significant positive effect on professional contributions

Tab. 2: Overview of studies without a specific learning area.

6. Limitations

While this study provides valuable insights into the use of publication status as a common inclusion criterion in systematic reviews, it also has limitations that must be taken into account when interpreting the findings. First, the study has a limited scope as it only focuses on two systematic reviews that explore digital media use in specific teaching settings in the German speaking context, thus limiting generalizability to other areas of education and media-related education studies. Secondly, there is no standard for a comprehensive quality assessment of educational empirical studies with different methods, which makes a unified quality assessment challenging. Thirdly, some specialized sub-fields of additional needs could not be represented in the search string due to a limited number of characters accepted by the search engine. Apart from this, a couple of terms which can be considered to be highly relevant to the research question (like «inclusion») led to misleading

results in the database search. Fourth, despite every effort especially grey literature is difficult to find and might be that not every piece of relevant research was found increasing a publication bias.

7. Conclusion

Summarizing the results from the reviews – first of all with regard to the systematic reviews, *not* yet with regard to this article’s methodological research question – the findings of both reviews are similar. Regardless of the review type there is a focus on individualized learning; hardly any study focusing on «learning *about* media» and a strong emphasis on reading, writing and mathematics. Considering the study design, it can be concluded there are mostly intervention studies – which is surprising, as elaborate basic research would have been the logical precondition for elaborating concepts. The study designs are mostly quantitative and experimental with a low number of participants in the majority of the studies. More research from a qualitative perspective is needed, e.g. interviews in relation to experienced barriers to participation. High emphasis is given to participation in society *through* media whereas the aim of learning *about* media (acquisition of media competence) is hardly recognizable. Not any single study could be identified focusing on participation *in* media; and participation *at* media is also underrepresented. The special educational needs areas in Germany are not represented in a balanced way in the reviewed studies. All in all, there is a focus on cognitive areas for the acquisition of certain skills or assistive technologies (psychomotor level), whereas affective components (commitment, motivation) and interdisciplinary competencies (creative thinking, emotional experience) are considered insufficiently. Studies focus on the individual support of pupils, especially in training programs, and the use of assistive technology is hardly considered. This one-sided focus on the rehabilitative use of digital media can be criticized. Nevertheless, it must be said that the function (with regard to SEN) is not always easy to identify. Both reviews reveal a didactic focus (learning *with* media) and hardly any study deals with learning *about* media. It can be said that the interdependence of media and inclusion – which is discussed in theoretical concepts (cf. Liesen and Rummler 2016; Kamin and Schluchter 2018; Bosse 2017; Bosse and Schluchter 2019; Bosse et al. 2019) is currently not researched in an evidence-based way for school education. Only 15 studies in review I and 19 studies in review II met the inclusion criteria. Thus, there is a fundamental gap in the context of inclusive digital learning.

The methodological research question was, *if* and *if so, how* publication status as a selection criterion changes the overall result of systematic reviews in the field of media-related educational studies. This contribution has shown that incorporating grey literature into a systematic review can provide valuable insights

and perspectives, and through this, reduce evidence selection bias. By expanding the scope of the literature search to include grey literature in review II, access to a wider range of information was gained and uncovered relevant studies (n=19) that would have otherwise been missed to answer the original research question on evidence-based use of digital media for participation *in, at and through* media. In consequence, the number of relevant results was more than doubled. This can ultimately lead to a more comprehensive understanding of the topic being studied, and help to answer initial research questions more effectively. It can be summarized (in the context of media-related education studies) that a complete picture of the empirical state of the art has only been possible via considering journal articles without peer review and grey literature¹⁰. However, the comparison of the two reviews confirmed that there are also several challenges associated with finding grey literature, which include:

- Difficulties in identification: grey literature is often not indexed in standard bibliographic databases, making it difficult to locate using traditional search methods.
- Limited accessibility: grey literature was less often available online, and access was sometimes restricted.
- Time-consuming process: In comparison to review I, searching for grey literature was more time-consuming due to the large number of sources that needed to be searched and the lack of standardization in how grey literature is published and indexed.
- Variability in quality: white literature without peer review and grey literature is not subject to the same level of peer review and quality control as traditionally-published literature, which can lead to concerns about the reliability and validity of the information it contains.

To overcome these challenges, researchers should employ a comprehensive search strategy that includes multiple databases and sources, including those that specialize in grey literature. Additionally, using tools such as quality assessments may help to evaluate the quality of the literature found, and ensure that the literature included in the review is reliable. In our study, for review II, a quality assessment was added for studies with missing peer review and for dissertations – both for quantitative (Higgins et al. 2019) and qualitative (Spencer et al. 2003) studies.

In conclusion, the present study supports the argument that neglecting grey literature in systematic reviews in the field of media education studies may increase evidence selection bias. For the limited branch of media-related education

¹⁰ This is in contrast to the statistical fact that the overall number of peer reviewed journals has risen from 16,925 in 2002 to 23,973 in 2008 (Tenopir and King 2009) but goes in line with Kuhberg-Lasson, Singleton, and Sondergeld (2014) who showed that a large number of outputs generated in education studies are published without peer review.

sciences, much high-level research is still published in dissertations, monographs, mission-oriented research papers such as the prominent international comparative school performance study ICILS (International Computer and Information Literacy Study) (Eickelmann et al. 2019). All these findings would have been ignored as a basis for future research if the systematic review had included peer reviewed articles only. For some aspects, such as «media education» there would have been a complete blind spot – if not considering the non-peer reviewed and grey results, too. In particular, research results based on qualitative designs face the risk of not being acknowledged, if publication status had been a dominant exclusion criterion. In contrary to the STEM sciences, few journals with high impact factors do exist in this specialised branch. Action research and evaluative reports in particular do not seem to be published in the peer review format. These findings cannot be generalized, of course. Yet, the question should be critically raised in the scientific community of inclusive digital learning and special educational needs, whether peer review is a «must»-criterion for reading and publishing research findings. In our reviews, valuable insight could be gained via the non-peer reviewed articles and dissertations so that an open approach can be recommended for this research community.

Researchers are therefore advised to reflect very sensitively when conducting systematic reviews about whether publication status is a suitable inclusion criterion. Including grey literature can help to increase the understanding of the context and background of a topic, and the ability to identify gaps in existing research. Additionally, incorporating grey literature can help to enhance the comprehensiveness and credibility of a study. The systematic review method may need to be revised in such a way that published reviews should identify and summarize research thoroughly irrespective of publication status. This would not weaken the peer review seal of quality as an expression of good scientific practice, and, at the same time, important research work could be upgraded and given greater visibility.

References¹¹

- *Aster, Michael von, Tanja Käser, Karin Kucian, and Markus Gross. 2012. «Calcularis – Rechenschwäche mit dem Computer begegnen». *Schweizerische Zeitschrift für Heilpädagogik* (6): 32–36.
- *Bianchi, Michael. 2012. *Kommunikative Kompetenz und Teilhabe. Der Computer als Hilfsmittel zur Erweiterung dieser Kompetenz bei Menschen mit Handicap*. Hamburg: Diplomica Verlag.
- *Bogner, Barbara, Manfred Hintermair, and Annerose Keilmann. 2018. «Bedeutung von Hörqualität für die kommunikative Partizipation im Unterricht und den Schulerfolg bei hörgeschädigten Grundschulern an allgemeinen Schulen». *Zeitschrift für Audiologie (Audiologie Acoustics)* 57 (2): 50-60. <https://doi.org/10.4126/FRL01-006412909>.

¹¹ *= studies included in the reviews.

- *Bosse, Ingo. 2018. «Qualitätskriterien für audiovisuelle und digitale Medien für den inklusiven Unterricht. Eine Evaluationsstudie des digitalen Angebots „Planet Schule“». *Zeitschrift für Heilpädagogik* (6): 256–70.
- *Brasch, Christa, and Udo Wölkerling. 2016. «Deine Firewall gegen Cybermobbing. Projekttag an Förderschulen zum Schutz vor Cybermobbing und sexuellen Übergriffen für 11- bis 17-Jährige mit Lernschwierigkeiten». *Kindesmisshandlung und -vernachlässigung* 19 (2): 184–93. <https://doi.org/10.13109/kind.2016.19.2.184>.
- *Castello, Armin, Benjamin Fauth, Nadia Juga, and Martina Pohl. 2010. «Differenzierung fiktionaler und nicht-fiktionaler Fernsehhalte bei Kindern an Lernförderschulen. Evaluation eines Trainingsmanuals». *Empirische Sonderpädagogik* 2 (2010) 4: 50–61. <https://doi.org/10.25656/01:9354>.
- *Drescher, Madlen. 2016. *Förderung des Leseverständnisses hörgeschädigter SchülerInnen – Konzeption, Entwicklung und Evaluation der digitalen Lernumgebung LESERÄTSEL*. Braunschweig: Technische Universität Braunschweig.
- *Drossard, Trixi, Florian Grond, and Thomas Hermann. 2012. «Interaktive Sonifikation mathematischer Funktionen als Unterrichtsmethode für blinde und sehbehinderte Schülerinnen und Schüler». *Blind – sehbehindert* (1): 42–54.
- *Hintz, Anna-Maria. 2014. «Evaluation einer computerbasierten Förderung schriftsprachlicher Vorläuferkompetenzen in der Schuleingangsphase». *Empirische Sonderpädagogik* 6 (3): 260–76. <https://doi.org/10.25656/01:9934>.
- *Hölz, Andreas. 2015. «Kurzzeitige Effekte eines Trainings der Leseflüssigkeit von leseschwachen Kindern mit isolierten Silben und Morphemen: Short-term effects of training the reading fluency of dyslexic children with isolated syllables and morphemes». Pabst Science Publishers. <https://doi.org/10.25656/01:10868>.
- *Hönninger, Anita. 2019. «Evaluation des computerbasierten Trainingsprogramms Lautarium bei Kindern mit Sprachentwicklungsstörungen». *Forschung Sprache* 1 (7): 38–57.
- *Kargl, Reinhard, Christian Purgstaller, Claudia Mrazek, Karina Ertl, and Andreas Fink. 2011. «Förderung der Lese- und Rechtschreibkompetenz auf Basis des morphematischen Prinzips». *Zeitschrift für Heilpädagogik* 62 (2): 61–68.
- *Klatte, Maria, Claudia Steinbrink, Kirstin Bergström, and Thomas Lachmann. 2014. «Effekte des computerbasierten Trainingsprogramms „Lautarium“ auf die phonologische Verarbeitung und die Lese-Rechtschreibleistungen bei Grundschulkindern». In *Legasthenie und Dyskalkulie Neue Methoden zur Diagnostik und Förderung*, edited by Schulte-Körne, 127–44. Bochum: Winkler.
- *Klatte, Maria, Claudia Steinbrink, Kirstin Bergström, and Thomas Lachmann. 2016. «„Lautarium“ – Ein computerbasiertes Trainingsprogramm für Grundschulkinde mit Lese-Rechtschreibschwierigkeiten». In *Förderprogramme für Vor- und Grundschule. Tests und Trends*, edited by Marcus Hasselhorn, and Wolfgang Schneider. Vol. 1, 115–42. Göttingen: Hogrefe.
- *Köhlmann, Wiebke. 2016. *Zugänglichkeit virtueller Klassenzimmer für Blinde*. Berlin: Logos Verlag Berlin GmbH Comeniushof. <https://doi.org/10.5281/zenodo.155933>.

- *Kohn, Juliane, Larissa Rauscher, Tanja Käser, Karin Kucian, Ursina McCaskey, Anne Wyschkon, Günter Esser, and Michael von Aster. 2017. «Effekte des ‘Calcularis’-Trainings. Teil 1: Domänen-spezifische Veränderungen». *Lernen und Lernstörungen* 6 (2): 51–63. <https://doi.org/10.1024/2235-0977/a000166>.
- *Lenhard, Wolfgang, Alexandra Lenhard, Marcus Hasselhorn, and Wolfgang Schneider. 2016. „conText“ – Training des Leseverständnisses mithilfe semantischer Technologien. In *Förderprogramme für Vor- und Grundschule Tests und Trends – Jahrbuch der pädagogisch-psychologischen Diagnostik 14*, edited by Marcus Hasselhorn, and Wolfgang Schneider, 209–24. Göttingen: Hogrefe.
- *Lenhard, Wolfgang, Herbert Baier, Darius Endlich, Alexandra Lenhard, Wolfgang Schneider, and Joachim Hoffmann. 2012. «Computerunterstützte Leseverständnisförderung: Die Effekte automatisch generierter Rückmeldungen». *Zeitschrift für Pädagogische Psychologie* 26 (2): 135–48. <https://doi.org/10.1024/1010-0652/a000066>.
- *Lenhard, Alexandra, und Wolfgang Lenhard. 2011. «Computerbasierte Intelligenzförderung mit den ‚Denkspielen mit Elfe und Mathis‘ - Vorstellung und Evaluation eines Computerprogramms für Vor- und Grundschüler». *Empirische Sonderpädagogik* 3 (2): 105–20.
- *Michna, Dagmar, Insa Melle, and Franz B. Wember. 2016. «Gestaltung von Unterrichtsmaterialien auf Basis des Universal Design for Learning. Am Beispiel des Chemieanfangsunterrichts in der Sekundarstufe I». *Sonderpädagogische Förderung heute* 61 (3): 286–303. <https://doi.org/10.3262/SOF1603286>.
- *Mihajlovic, Christopher. 2012. «Die Nutzung von Computer und Internet an Förderschulen». *Medien + Erziehung* 56 (1): 25–31.
- *Müller, Stephanie, and Sandra Fleischer. 2013. «Medienkompetenz und geistige Behinderung. Einsatz von Medien im Schulalltag von geistig behinderten Heranwachsenden». *Medien + Erziehung* 57 (3): 55–59.
- *Nobel, Kerstin, and Matthias Grünke. 2017. «Über die Auswirkungen einer PC-gestützten Schreibförderung auf die Länge und Qualität von Aufsätzen von risikobelasteten Fünftklässlerinnen und Fünftklässlern». *Empirische Sonderpädagogik* 9 (4): 323–40.
- *Nobel, Kerstin, Anne Barwasser, Conny Melzer, and Matthias Grünke. 2020. «Webbasiertes Distanzlernen. Erfahrungen im Rahmen einer Pilotstudie während der Covid-19-Schulschließungen zur digitalen Förderung der Aufsatzfähigkeiten von Schülerinnen und Schülern mit dem Förderschwerpunkt Lernen in der Sekundarstufe I». *Zeitschrift für Heilpädagogik* 9 (71): 465–79.
- *Preuß, Isabell. 2016. *Computer-basierte Förderung der Lesefertigkeit von Kindern aus Diagnose- und Förderklassen*. Würzburg: Universität Würzburg. <https://doi.org/123932>.
- *Rauscher, Larissa, Juliane Kohn, Tanja Käser, Karin Kucian, Ursina McCaskey, Anne Wyschkon, Svenja Moraske, Günter Esser, and Michael von Aster. 2017. «Effekte des „Calcularis“-Trainings. Teil 2: Veränderungen psychosozialer Merkmale». *Lernen und Lernstörungen* 6 (2): 75–86. <https://doi.org/10.1024/2235-0977/a000168>.

- *Schiedek, Steffen, and Katharina Menges. 2018. *Stärkung der Schülerinnen- und Schülerkommunikation durch die Methode der Videoannotation im inklusiven Sportunterricht*. Hannover: Leibniz Universität Hannover.
- *Schöfl, Martin. 2016. «Delfino – Ein Online-Förderwerkzeug bei LRS für den Einsatz in der Klasse». *Lernen und Lernstörungen* 5 (2): 111–18.
- *Schwenk, Christin, Jörg-Tobias Kuhn, Philipp Doeblner, and Heinz Holling. 2017. «Auf Goldmünzenjagd: Psychometrische Kennwerte verschiedener Scoringansätze bei computergestützter Lernverlaufdiagnostik im Bereich Mathematik». *Empirische Sonderpädagogik* 9: 123–42. <https://doi.org/10.25656/01:15009>.
- *Urff, Christian. 2014. «Digitale Lernmedien zur Förderung grundlegender mathematischer Kompetenzen. Theoretische Analysen, empirische Fallstudien und praktische Umsetzung anhand der Entwicklung virtueller Arbeitsmittel». PhD diss., Pädagogische Hochschule Ludwigsburg.
- *Wachsmuth, Claudia. 2020. «Computerbasiertes Lernen mit Aufmerksamkeitsdefizit: Unterstützung des selbstregulierten Lernens durch metakognitive Prompts». PhD diss., TU Chemnitz.
- *Walter, Jürgen. 2018. «Zur Effektivität der Förderung der Leseflüssigkeit auf der Basis von Hörbüchern in Kombination mit wiederholtem Lesen: Weitere Evidenz». *Empirische Sonderpädagogik*, 10 (3): 248–72.
- *Werne, Birgit, Mareike Drinkhaus, Katharina Rank, and Katharina König. 2013. «Förderung der Lesemotivation und Lesekompetenz von Förderschülern durch ein multimediales Programm zur Leseförderung». [Improvement of reading motivation and reading competence among remedial education students using a multimedia remedial reading program]. *Zeitschrift für Heilpädagogik* 64 (3): 92–98.
- *Wicki, Monika T., and Susan C. A. Burkhardt. 2018. «Unterstützende Technologien im ersten Zyklus des Lehrplans 21». In *Sonderpädagogik in der digitalisierten Lernwelt. Beiträge der nationalen Tagung Netzwerk Forschung Sonderpädagogik*. La pédagogie spécialisée dans l'environnement numérique d'apprentissage. Actes des la journée d'étude du Réseau de recherche en pédagogie spécialisée, edited by Stefanie Calabrese, Barbara Egloff, Corinne Monney, Greta Pelgrims, Caroline Sahli Lozano, Diana Sahrai, Claudio Straccia, and Monika Wicki, 47-62. Bern: Edition SZH/CSPS.
- *Wollny, Christoph. 2015. *Digitale Medien in der Inklusion. Pilotstudie zum Einsatz eines interaktiven iBooks*. Köln.

- Armstrong, J. Scott. 1997. «Peer review for journals: Evidence on quality control, fairness, and innovation». *Science and engineering ethics* 3 (1): 63–84. <https://doi.org/10.1007/s11948-997-0017-3>.
- Aspiranti, Kathleen B., Karen H. Larwin, and Benjamin P. Schade. 2020. «IPads/tablets and Students with Autism: A Meta-Analysis of Academic Effects». *Assistive technology: the official journal of RESNA* 32 (1): 23–30. <https://doi.org/10.1080/10400435.2018.1463575>.
- Bambey, Doris. 2016. «Fachliche Publikationskulturen und Open Access. Fächerübergreifende Entwicklungstendenzen und Spezifika der Erziehungswissenschaft und Bildungsforschung». PhD diss., Technische Universität Darmstadt. <https://doi.org/10.25656/01:12331>.
- Birukou, Aliaksandr, Joseph Rushton Wakeling, Claudio Bartolini, Fabio Casati, Maurizio Marchese, Katsiaryna Mirylenka, Nardine Osman, Azzurra Ragone, Carles Sierra, and Aalam Wassef. 2011. «Alternatives to Peer Review: Novel Approaches for Research Evaluation». *Frontiers in computational neuroscience* 5: 1–12. <https://doi.org/10.3389/fncom.2011.00056>.
- Bond, Melissa. 2018. «Helping doctoral students crack the publication code: An evaluation and content analysis of the Australasian Journal of Educational Technology». *Australian Journal of Educational Technology* 34 (5): 167–81. <https://doi.org/10.14742/ajet.436>.
- Bosse, Ingo, and Jan-René Schluchter. 2019. «Berufsfeld Sekundarstufe I». In *Handbuch Inklusion und Medienbildung*, edited by Ingo Bosse, Jan-René Schluchter, and Isabel Zorn, 119–31. Weinheim: Beltz Juventa.
- Bosse, Ingo, Anna-Maria Kamin, and Jan-René Schluchter. 2019. «Medienbildung für alle: Inklusive Medienbildung – Zugehörigkeit und Teilhabe in gegenwärtigen Gesellschaften». In *Medienbildung für alle. Digitalisierung. Teilhabe. Vielfalt*, edited by Marion Brüggemann, Sabine Ede, and Angela Tillmann, 35–52. Schriften zur Medienpädagogik 55. München: kopaed.
- Bosse, Ingo. 2017. «Gestaltungsprinzipien für digitale Lernmittel im Gemeinsamen Unterricht. Eine explorative Studie am Beispiel der Lernplattform Planet Schule». In *Jahrbuch Medienpädagogik 13: Vernetzt und entgrenzt – Gestaltung von Lernumgebungen mit digitalen Medien*, edited by Kerstin Mayrberger, Johannes Fromme, Petra Grell, and Theo Hug, 133–49. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-16432-4_9.
- Brunton, Ginny, Claire Stansfield, Jenny Caird, and James Thomas. 2017. «Finding relevant studies». In *An introduction to systematic reviews*, edited by David Gough, Sandy Oliver, and James Thomas, 93–122. London: Sage.
- Cherry, M. Gemma, and Dickson, Rumona. 2014. «Defining my review question and identifying inclusion criteria». In *Doing a Systematic Review: a student's guide*, edited by Angela Boland, Gemma Cherry, and Rumona Dickson, 43–60. London: SAGE Publications.
- Chow, Jason C., and Eric Eckholm. 2018. «Do Published Studies Yield Larger Effect Sizes than Unpublished Studies in Education and Special Education? A Meta-review». *Educ Psychol Rev* 30: 727–44. <https://doi.org/10.1007/s10648-018-9437-7>.

- Gough, David, Thomas James, and Oliver Sandy. 2012. «Clarifying differences between review designs and methods». *Systematic Reviews* 1 (1): 28. <https://doi.org/10.1186/2046-4053-1-28>.
- Drucker, Aaron M., Patrick Fleming, and An-Wen Chan. 2016. «Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews». *Journal of Investigative Dermatology* 136 (11): e109–14. <https://doi.org/10.1016/j.jid.2016.08.021>.
- EBSCO. 2023. «EBSCO Open Dissertations». Accessed January 29, 2023. <https://www.ebsco.com/products/research-databases/ebsco-open-dissertations>.
- Eickelmann, Birgit, Wilfried Bos, Julia Gerick, Frank Goldhammer, Heike Schaumburg, Heike, Knut Schwippert, Martin Senkbeil, and Jan Vahrenhold. 2019. *ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking*. Münster; New York: Waxmann. <https://doi.org/10.25656/01:18166>.
- Ferguson, Christine L. 2020. «Open Peer Review». *Serials Review* 46 (4): 286–91. <https://doi.org/10.1080/00987913.2020.1850039>.
- Fitzgerald, Gail, Kevin Koury, and Katherine Mitchem. 2008. «Research on Computer-Mediated Instruction for Students with High Incidence Disabilities». *Journal of Educational Computing Research* 38 (2): 201–33. <https://doi.org/10.2190/EC.38.2.e>.
- Goldbeck-Wood, Sandra. 1999. «Evidence on Peer Review-Scientific Quality Control or Smokescreen?». *BMJ (Clinical research ed.)* 318 (7175): 44–45. <https://doi.org/10.1136/bmj.318.7175.44>.
- Haddaway, Neal, Woodcock, Paul, Macura, Biljana, and Collins, Alexandra. 2015. «Making literature reviews more reliable through application of lessons from systematic reviews». *Conservation Biology*. 29 (6): 1596–605. <https://doi.org/10.1111/cobi.12541>.
- Haddaway, Neal R., and Bayliss, Helen R. 2015. «Shades of grey: Two forms of grey literature important for reviews in conservation». *Biological Conservation* (191): 827–29.
- Hall, Jeremy L., William Hatcher, Bruce D. McDonald, Patricia Shields, and Jessica E. Sowa. 2019. «The art of peer reviewing: Toward an effective developmental process». *Journal of Public Affairs Education* 25 (3): 296–313. <https://doi.org/10.1080/15236803.2019.1616657>.
- Hartling Lisa, Robin Featherstone, Megan Nuspl, Kassi Shave, Donna M. Dryden, and Ben Vandermeer. 2017. «Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews». *BMC Med Res Methodol.* 17 (1): 64. <https://doi.org/10.1186/s12874-017-0347-z>.
- Heesen, Remco, and Liam Kofi Bright. 2021. «Is Peer Review a Good Idea?». *British Journal for the Philosophy of Science* 72 (3): 635–63. <https://doi.org/10.1093/bjps/axz029>.
- Henderson, Michael, Simon Shurville, and Ken Fernstrom. 2009. «The quantitative crunch». *Campus-Wide Information Systems* 26 (3): 149–67. <https://doi.org/10.1108/10650740910967348>.

- Herzig, Bardo. 2014. *Wie wirksam sind digitale Medien im Unterricht?* Gütersloh: Bertelsmann Stiftung. https://rsm-bst-live.bertelsmann-stiftung.de/fileadmin/files/bst/publikationen/grauepublikationen/studie_ib_wirksamkeit_digitale_medien_im_unterricht_2014.pdf.
- Higgins, Julian, and Sally Green. 2013. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0. The Cochrane Collaboration 2011.
- Higgins, Julian, James Thomas, Jacqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew J. Page, and Vivian Andrea Welch. 2019. *Cochrane Handbook for Systematic Reviews of Interventions*. Second edition. Cochrane book series. Hoboken, NJ, Chichester: Wiley Blackwell.
- Hojat, Mohammadreza, Joseph S. Gonnella, and Addeane S. Caelleigh. 2003. «Impartial Judgment by the „Gatekeepers“ of Science: Fallibility and Accountability in the Peer Review Process». *Advances in health sciences education: theory and practice* 8 (1): 75–96. <https://doi.org/10.1023/A:1022670432373>.
- Horbach, Serge, and Willem Halfman. 2018. «The Changing Forms and Expectations of Peer Review». *Research integrity and peer review* 3: 1–15. <https://doi.org/10.1186/s41073-018-0051-5>.
- Irion, Thomas, and Katharina Scheiter. 2018. «Didaktische Potenziale digitaler Medien. Der Einsatz digitaler Technologien aus grundschul- und mediendidaktischer Sicht». *Grundschule aktuell: Zeitschrift des Grundschulverbandes*, 142: 8–11. <https://doi.org/10.25656/01:16559>.
- Johnson, David R., and Joseph C. Hermanowicz. 2017. «Peer Review: From ‘Sacred Ideals’ to ‘Profane Realities’». In *Higher Education: Published Under the Sponsorship of the Association for Institutional Research (AIR) and the Association for the Study of Higher Education (ASHE)*. Bd. 32, edited by Michael B. Paulsen, 485–527. Higher Education v.32. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-48983-4_10.
- Jukola, Saana. 2017. «A Social Epistemological Inquiry into Biases in Journal Peer Review». *Perspectives on Science* 25 (1): 124–48. https://doi.org/10.1162/POSC_a_00237.
- Kamin, Anna-Maria, and Jan-René Schluchter. 2018. «Medienbildung und Inklusion – Perspektiven für Theorie und Praxis». In *Medienbildung und Inklusion – Perspektiven für Theorie und Praxis*, edited by Wolfgang Schill, and Renate Röllecke, 16–43. Düsseldorf: BZgA.
- Katzav, Joel, and Krist Vaesen. 2017. «Pluralism and peer review in philosophy». *Philosophers Imprint* 17: 1–20. <https://doi.org/10.1080/09608788.2016.1261794>.
- Kirman, Christopher R., Ted W. Simon, and Sean M. Hays. 2019. «Science Peer Review for the 21st Century: Assessing Scientific Consensus for Decision-Making While Managing Conflict of Interests, Reviewer and Process Bias». *Regulatory toxicology and pharmacology: RTP* 103: 73–85. <https://doi.org/10.1016/j.yrtph.2019.01.003>.
- KMK. 2017. *Strategie der Kultusministerkonferenz <Bildung in der digitalen Welt>. Beschluss der Kultusministerkonferenz vom 08.12.2016 in der Fassung vom 07.12.2017*, Berlin.

- KMK. 2021. Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland / Sekretariat. 2021. «*Lehren und Lernen in der digitalen Welt: Ergänzung zur Strategie der Kultusministerkonferenz Bildung in der digitalen Welt*». Beschluss der Kultusministerkonferenz vom 09.12.2021. Bonn [u.a.]: Sekretariat der Kultusministerkonferenz. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf.
- Kuhberg-Lasson, Veronika, Katja Singleton, and Ute Sondergeld. 2014. «Publikationscharakteristika im interdisziplinären Feld der Bildungsforschung». *Journal for educational research online* 6: 134–55. <https://doi.org/10.25656/01:9691>.
- La Paro, Karen, and Pianta, Robert C. 2000. «Predicting children's competence in the early school years: A meta-analytic review». *Review of Educational Research*, 70 (4): 443–84.
- Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. 2013. «Bias in peer review». *J Am Soc Inf Sci Tec* 64 (1): 2–17. <https://doi.org/10.1002/asi.22784>.
- Liesen, Christian, and Klaus Rummler. 2016. «Digitale Medien und Sonderpädagogik: Eine Auslegeordnung für die interdisziplinäre Verbindung von Medien und Sonderpädagogik». *Schweizer Zeitschrift für Heilpädagogik* 22 (4): 6–12. <https://doi.org/10.5281/zenodo.5882697>.
- McKenzie, Joanne E., Sue E. Brennan, Rebecca E. Ryan, Hilary J. Thomson, Renea V. Johnston, and James Thomas. 2022. «Chapter 3: Defining the Criteria for Including Studies and How They Will Be Grouped for the Synthesis» *Cochrane Training*. 2022. Accessed January 29, 2023. <https://training.cochrane.org/handbook/current/chapter-03>.
- Mertens, Claudia, Carolin Quenzer-Alfred, Anna-Maria Kamin, Tim Homrighausen, Tina Niermeier, and Daniel Mays. 2022. «Empirischer Forschungsstand zu digitalen Medien im Schulunterricht in inklusiven und sonderpädagogischen Kontexten – eine systematische Übersichtsarbeit», *Zeitschrift für Empirische Sonderpädagogik*, 26-46; https://www.psychologie-aktuell.com/fileadmin/Redaktion/Journale/esp-2022-1/ESP_2022-1_2.pdf.
- Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. «Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement». *Annals of internal medicine* 151 (4): 264–69. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>.
- Müller, Uwe Thomas. 2009. *Peer review-Verfahren zur Qualitätssicherung von Open-Access-Zeitschriften. Systematische Klassifikation und empirische Untersuchung*. Unter Mitarbeit von Humboldt-Universität zu Berlin. Berlin. <https://doi.org/10.18452/15885>.
- Mulligan, Adrian, Louise Hall, and Ellen Raphael. 2013. «Peer review in a changing world: An international study measuring the attitudes of researchers». *J Am Soc Inf Sci Tec* 64 (1): 132–61. <https://doi.org/10.1002/asi.22798>.
- Page, Matthew J, Julian P.T. Higgins, and Jonathan A.C. Sterne. 2022. «Chapter 13: Assessing Risk of Bias Due to Missing Results in a Synthesis». *Cochrane Training*. 2022. Accessed January 29, 2023. <https://training.cochrane.org/handbook/current/chapter-13>.

- pedocs. n. d. «pedocs allgemein – pedocs» peDOCS. https://www.pedocs.de/info_pedocs.php.
- Pennington, Robert C. 2010. «Computer-Assisted Instruction for Teaching Academic Skills to Students With Autism Spectrum Disorders: A Review of Literature». *Focus Autism Other Dev Disabl* 25 (4): 239–48. <https://doi.org/10.1177/1088357610378291>.
- Pollock, Michelle, Ricardo M. Fernandes, Lorne A. Becker, Dawid Pieper, and Lisa Hartling. 2022. «Chapter V: Overviews of Reviews». In *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022), edited by Higgins Julian, Thomas James, Chandler Jaqueline, Cumpston Miranda, Li Tianjing, Page Matthew, Welch Vivian. Cochrane, 2022. <https://www.training.cochrane.org/handbook>.
- Quenzer-Alfred, Carolin, Claudia Mertens, Tim Homrighausen, Anna-Maria Kamin, and Daniel Mays. 2023. «Systematisches Review des empirischen Forschungsstands zu digitalen Medien für SchülerInnen mit einem zusätzlichen oder einem sonderpädagogischen Förderbedarf unter Berücksichtigung inklusiver, integrativer und exkludierender Unterrichtsszenarien». In *Bildung für eine digitale Zukunft. Edition ZfE*, vol 15, edited by Scheiter, Katharina., Gogolin, Ingrid. Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-37895-0_6.
- Rowley, Jennifer, and Laura Sbaffi. 2018. «Academics' attitudes towards peer review in scholarly journals and the effect of role and discipline». *Journal of Information Science* 44 (5): 644–57. <https://doi.org/10.1177/0165551517740821>.
- Salager-Meyer, Françoise. 2008. «Scientific publishing in developing countries: Challenges for the future». *Journal of English for Academic Purposes* 7 (2): 121–32. <https://doi.org/10.1016/j.jeap.2008.03.009>.
- Sánchez-Serrano, Sarasola Jose Luis, Alicia Jaen-Martinez, Marta Montenegro-Rueda, and Jose Fernandez-Cerero. 2020. «Impact of the Information and Communication Technologies on Students with Disabilities. A Systematic Review 2009–2019». *Sustainability* 12 (20). <https://doi.org/10.3390/su12208603>.
- Schaumburg, Heike. 2019. «Chancen und Risiken digitaler Medien in der Schule. Medienpädagogische und -didaktische Perspektiven». In *Individuell fördern mit digitalen Medien: Chancen, Risiken, Erfolgsfaktoren*. Vol. 3, edited by Bertelsmann Stiftung, 19–94. Gütersloh: Bertelsmann Stiftung.
- Schulte, Jolika, and Zimmer, Karin (unpublished). «Haben nicht-wissenschaftliche Quellen einen Mehrwert? Plädoyer für eine umfassende Suchstrategie bei der Erstellung von Systematic Reviews in der Bildungsforschung». *Posterbeitrag auf der GEBF 2023*, Universität Duisburg-Essen.
- Shaw, Philip, Angus Phillips, and Maria Bajo Gutiérrez. 2022. «The Death of the Monograph?», Mai, 1–14. *Pub Res Q* 38, 382–95. <https://doi.org/10.1007/s12109-022-09885-2>.
- Spencer, Liz, Jane Ritchie, Jane Lewis, and Lucy Dillon. 2003. *Quality in Qualitative Evaluation: A framework for assessing research evidence. A Quality Framework*. <https://dera.ioe.ac.uk/21069/2/a-quality-framework-tcm6-38740.pdf>.

Stern, Cindy, Zoe Jordan, and Alexa McArthur. 2014. «Developing the review question and inclusion criteria: The first steps in conducting a systematic review». *American Journal of Nursing* 114 (4): 53–56.

Tenopir, Carol, and Donald W. King. 2009. «The growth of journals publishing». In *The future of the academic journal*, edited by Bill Cope, and Angus Phillips, 105–24. Oxford, UK: Chandos Publishing. <https://doi.org/10.1007/s10668-012-9337-9>.

University of Wolverhampton. n.d. «Guide to Searching Grey Literature» Accessed January 29, 2023. <https://www.wlv.ac.uk/lib/media/departments/lis/skills/study-guides/LS124-Searching-Grey-Literature.pdf>.

Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

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

Potenziale und Herausforderungen von Forschungssynthesen im Kontext der Digitalisierung im Bildungsbereich am Beispiel des Critical Reviews

Anna Heinemann¹ , Jan Koschorreck² , Annika Wilmers³ , Pia Sander¹  und
Jens Leber¹ 

¹ Universität Duisburg-Essen

² DIE | Deutsches Institut für Erwachsenenbildung

³ DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Zusammenfassung

Forschungssynthesen haben im Kontext der Digitalisierung im Bildungsbereich, u. a. im Hinblick auf die hohe Dynamik der Entwicklungen und Innovationen ein grosses Potenzial, systematische Überblicke über den aktuellen Stand zu geben. Sie können damit sowohl für Forschende als auch für Mitglieder der Praxiscommunity von grossem Wert sein. Critical Reviews stellen eine spezifische Form der Forschungssynthese dar, da zusätzlich zu einem strukturierten Überblick die eingeschlossene Literatur kritisch analysiert und eingeordnet wird. Gegebenenfalls schliessen sie dabei an bestehende Konzepte oder Modelle an und können damit ergänzende Perspektiven auf das Forschungsfeld aufzeigen oder zur Modellbildung beitragen. Das Vorgehen orientiert sich dabei an umfassend ausgearbeiteten Systematic Reviews, ermöglicht durch die in den vorgestellten Fallbeispielen verkürzte Bearbeitungszeit aber eine vergleichsweise zeitnahe Reaktion auf aktuelle Bildungsthemen. Die in diesem Beitrag vorgestellten Critical Reviews sind in den Bildungsbereichen Schule und Erwachsenenbildung entstanden. Die Beschreibung der beiden Fallbeispiele sowie die anschliessende Diskussion der methodischen Herausforderungen und Potenziale dienen der Reflexion und Einordnung von Forschungssynthesen, hier am Beispiel des Critical Reviews, im Kontext der Digitalisierung im Bildungsbereich. Zusätzlich werden Implikationen auf der Arbeitsebene reflektiert.

Potentials and Challenges of Research Syntheses in the Context of Digitisation in Education by Means of the Critical Review

Abstract

In the context of digitisation in education, research syntheses have great potential to provide systematic overviews of the state of art, especially regarding the dynamic of developments and innovations. Thus, they can be of great value for researchers as well as for members of the practice community. Critical reviews represent a specific type of research synthesis where in addition to giving a structured overview, the included literature is critically analysed and classified. Where appropriate, they connect to existing concepts or models and thus can reveal complementary perspectives on the research field or they contribute to model building. The procedure is oriented towards comprehensively elaborated systematic reviews, but the abbreviated procedure used in the following two cases enables stakeholders to respond quickly to educational hot topics. The critical reviews presented in this contribution refer to the education sectors of school and adult and continuing education. The description of the two case studies and the subsequent discussion of the methodological challenges and potentials serve to reflect on and classify research syntheses, in this case the critical review, in the context of digitisation in education. In addition, implications at the operative level are discussed.

1. Forschungssynthesen in der Bildungsforschung

Bildung in einer zunehmend digitalisierten Welt repräsentiert dynamische gesellschaftliche Entwicklungen (Kerres 2018; Kultusministerkonferenz 2021) und damit verbunden ein sich schnell entwickelndes Forschungsfeld, in dem fortlaufend neue innovative Konzepte, Technologien und Strategien der Transformation erprobt werden (Weingart et al. 2022). Für Wissenschaftlerinnen und Wissenschaftler ist es daher eine Herausforderung, den Überblick über neueste Erkenntnisse und laufende Erprobungen zu behalten sowie aktuelle Forschungsdesiderate zu identifizieren. Ebenso stellt dies eine Schwierigkeit für die Bildungspraxis dar. Denn Handlungspraktiken aus der Vielzahl der Forschungsergebnisse zu extrahieren und für sich nutzbar zu machen, ist anspruchsvoll. Forschungssynthesen, z. B. in Form von Critical Reviews, können einen wichtigen Beitrag leisten, den Forschungsstand zu einem abgegrenzten Thema pointiert und strukturiert zusammenzufassen und somit für Wissenschaft und Praxis nutzbar zu machen (Zawacki-Richter et al. 2020).

Der Mehrwert von Forschungssynthesen wurde ursprünglich in der Medizin erkannt, in der man zu einer Fragestellung aus der vorhandenen wissenschaftlichen Literatur evidenzbasierte Erkenntnisse generierte. In den Sozialwissenschaften finden Forschungssynthesen vor allem seit dem ausgehenden 20. Jahrhundert

Anwendung (vgl. Petticrew und Roberts 2006 sowie für die Bildungsforschung Gough und Oliver 2017) und werden seitdem methodisch weiterentwickelt. Sutton et al. (2019) identifizieren vor dem Hintergrund der Gesundheitswissenschaften, aber ohne Festlegung auf eine bestimmte Fachrichtung, 48 Reviewtypen, die sie in sieben übergeordneten Kategorien gruppieren, wobei sie Critical Reviews gemeinsam mit *Narrative Reviews*, *Review Summaries*, *Integrative Reviews* und *State of the Art Reviews* der Familie der traditionellen Reviewformate zuordnen.

Das methodische Vorgehen von Critical Reviews orientiert sich grundsätzlich an umfassend ausgearbeiteten Systematic Reviews, weist aber «an klar definierten Schnittstellen eine [...] methodisch abgeschwächte Vorgehensweise» auf (Wilmers et al. 2020a, 16). Dabei fokussieren sie vor allem eine konzeptionelle Erfassung des Forschungsfeldes und können somit einen Beitrag zur weiteren Theorie- oder Modellbildung im Feld liefern (Grant und Booth 2009, 94). Schwerpunkte bilden eine ausführliche und systematische Suche von bestehender Literatur (Wilmers et al. 2020a), deren narrative Einordnung in den Forschungskontext und eine qualitative Einschätzung der Aussagekraft der Literatur hinsichtlich ihres Beitrags zur Beantwortung der jeweiligen Forschungsfrage sowie die Identifikation von Forschungsdesideraten. Eine ausführliche Qualitätsprüfung der ausgewählten Literatur findet hingegen nicht in dem Umfang statt, der für ein Systematic Review Standard wäre. Vielmehr steht die kritische Einordnung der Literaturbefunde in den Forschungskontext im Vordergrund. Charakteristisch für Critical Reviews sind entsprechend eher weit gefasste Fragestellungen – im Gegensatz zu z. B. Fragestellungen, die Wirkungen einzelner Interventionsarten erarbeiten (Wetterich und Plänitz 2021, 25–28).

In dem vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Metavorhaben «Digitalisierung im Bildungsbereich» werden seit 2019 jährlich fünf Critical Reviews erstellt und als Reviewbände veröffentlicht (Wilmers et al. 2020b; Wilmers, Achenbach, und Keller 2021, 2022). Die Critical Reviews synthetisieren und strukturieren aktuelles Forschungswissen zu einer abgegrenzten Fragestellung der digitalen Bildung und bearbeiten diese bildungssektorspezifisch. Somit entstehen Forschungssynthesen zu Bildungssektoren (Kindheit, Jugend & Familie, Schule, Beruf, Erwachsenenalter und in der Lehrer:innenausbildung), die wesentliche Erkenntnisse zusammenfassen und offene Forschungsfragen identifizieren. Von dieser sektorspezifischen Bearbeitung profitieren nicht nur Forscherinnen und Forscher, sondern auch die unterschiedlichen Akteure der Bildungspraxis – ein erklärtes Ziel der Förderpolitik in Deutschland (Kerres, Sander, und Waffner 2022).

Am Beispiel von zwei Critical Reviews aus dem genannten Forschungsprojekt identifiziert der vorliegende Beitrag Potenziale und Herausforderungen von Forschungssynthesen im Kontext der Digitalisierung. Durch die transparente Auseinandersetzung mit bestimmten und wiederkehrenden Aspekten der Reviewerstellung soll ein Beitrag zur anhaltenden methodischen Diskussion über

Forschungssynthesen allgemein, beziehungsweise zu einzelnen Subtypen hiervon in der Bildungsforschung geleistet werden (Cooper, Hedges, und Valentine 2019; Piggott et al. 2021; Zawacki-Richter et al. 2020).

2. Entstehungskontext und methodisches Vorgehen

Im Metavorhaben «Digitalisierung im Bildungsbereich» werden im Zeitraum von fünf Jahren insgesamt 20 Critical Reviews ausgearbeitet, die wesentliche Bereiche der digitalen Bildung abdecken. Dabei lassen sich die einzelnen Reviews nicht nur zur Beantwortung von Forschungsfragen in den jeweiligen Bildungssektoren heranziehen, sondern können auch über die Einzeltexte hinaus miteinander in Beziehung gesetzt werden, beispielsweise indem ein Vergleich der Reviews inhaltliche und methodische Gemeinsamkeiten oder Auffälligkeiten aufdeckt (z. B. Wilmers et al. 2023).

Die Aufteilung im Projektteam, das insgesamt für die einzelnen Durchläufe aus bis zu 13 Personen besteht, setzt sich aus zentralen Elementen (Koordination, Reflexion der Methodik, Recherche, Transferaufgaben) und dezentralen Elementen (Screening und Auswertung der Literatur, sektorenspezifischer Transfer) der Zusammenarbeit zusammen, wobei das Projektteam durchgehend im engen Austausch zu allen Arbeitsschritten steht. Dieses Vorgehen soll gleichermassen den hohen methodischen Standards in der Reviewerstellung als auch den inhaltlichen Herausforderungen im Themenfeld Digitalisierung und Bildung gerecht werden.

Das Projektteam wählt für jeden Reviewband mit jeweils fünf Reviews ein übergeordnetes Thema, für das anschliessend in jedem Bildungssektor eine spezifische Forschungsfrage entwickelt wird. Für die Literatursuche legt das Team gemeinsam einige grundlegende Parameter fest: Hierzu zählen in erster Linie die Auswahl der Sprachen Deutsch und Englisch, die Eingrenzung des Suchzeitraums auf jeweils fünf Jahre, die Eingrenzung des Recherchezeitraums auf durchschnittlich drei Monate sowie die Auswahl der Datenbanken FIS Bildung im Fachportal Pädagogik und ERIC (Education Resources Information Center) als Basis der Recherche. Auf dieser Grundlage werden im Verlauf der Recherche parallel zur Entwicklung der Suchstrings die Suchkriterien weiter angepasst und nach Bedarf spezifiziert, z. B. durch die Hinzunahme spezialisierter Datenbanken und erweiterten Suchtaktiken. Spezifische für den jeweiligen Sektor und dessen Forschungslage relevante Ein- und Ausschlusskriterien werden zusätzlich definiert. Anhand eines Kodierschemas beschreiben die Autor:innen der jeweiligen Bildungssektoren die recherchierte Literatur, erfassen die für das Review wesentlichen inhaltlichen Kernaussagen und schätzen deren Qualität ein. Letzteres beinhaltet aufgrund des Reviewformats zwar keine ausführliche Qualitätsprüfung, dennoch wird – zumindest für quantitative Studien – die

Aussagekraft vor dem Hintergrund gängiger Kriterien wie Validität oder Reliabilität betrachtet, damit über die Aufnahme der jeweiligen Studie ins Review entschieden werden kann (Booth, Sutton, und Papaioannou 2016; Grant und Booth 2009).

Um die Herausforderungen und Potenziale von Critical Reviews als eine Form von Forschungssynthesen in der Bildungsforschung herauszuarbeiten, zieht dieser Beitrag zwei Fallbeispiele heran, die im Kontext des Metavorhabens «Digitalisierung im Bildungsbereich» entstanden sind. Sie sind im Bildungsbereich der schulischen Bildung sowie der Erwachsenenbildung anzusiedeln und basieren auf aktuellen Publikationen aus dem Themenfeld «Lehren und Lernen mit digitalen Medien». In den Fallbeispielen wird das jeweilige Vorgehen zunächst skizziert¹, um sich danach den jeweils identifizierten Herausforderungen und Potenzialen der Methodik zu widmen. Die Beschreibung der beiden Critical Reviews (Kap. 3 & 4) sowie die anschließende Diskussion (Kap. 5) dienen der Reflexion und Einordnung von Forschungssynthesen im Kontext der Digitalisierung im Bildungsbereich.

Die Auswahl der beiden Critical Reviews für diesen Beitrag erfolgte auf Grundlage zweier Kriterien: zum einen waren die Autor:innen des Beitrags an der Erstellung der Critical Reviews massgeblich beteiligt. Zum anderen bieten die unterschiedlichen Ausgangslagen der ausgewählten Bildungsbereiche eine breite Abdeckung von sowohl unterschiedlichen, z. T. bereichsspezifischen als auch generischen Herausforderungen.

3. Fallbeispiel 1: Überschulische Curriculumsentwicklung im Kontext der Digitalisierung

3.1 Methodik des Critical Reviews

Das erste hier vorgestellte Critical Review fokussiert die Curriculumsentwicklung im allgemeinbildenden Schulsektor im Kontext der Digitalisierung. Ziel war die Darstellung und Einordnung davon, wie Digitalisierung international in den schulischen Bildungskanon aufgenommen und gewichtet wird und welche gesellschaftlichen Begründungsmuster und Ziele damit einhergehen. Die qualitative Einordnung der Primärforschung erfolgte in Anlehnung an das Modell zur vergleichenden Analyse politischer Massnahmen zur Digitalisierung in der Schulbildung nach Kozma (2008).

Die Literaturrecherche umfasste wissenschaftliche Publikationen in dem Zeitraum von Januar 2016 bis Juni 2021 und wurde über die Datenbanken FIS Bildung sowie ERIC, Web of Science – Social Citation Index, ERC (Education Research

¹ Für die ausführlichen Beschreibungen des Vorgehens siehe die zugrundeliegenden Critical Reviews aus dem Sektor der allgemeinbildenden Schule (Heinemann et al. 2022) und aus der Erwachsenenbildung (Koschorreck und Gundermann 2022).

Complete) und LearnTechLib (The Learning and Technology Library) durchgeführt. Dabei wurden die Suchbegriffe in drei Blöcken aus den Kategorien allgemeinbildende Schule (z. B. *Schule/school*), Curriculumsentwicklung (z. B. *Lehrplan/educational curriculum*) und Digitalisierung (z. B. *Medienintegration/media integration*) sowohl in deutscher als auch in englischer Sprache definiert. Unter der Kategorie der Curriculumsentwicklung wurden Suchbegriffe im Hinblick auf Bildungspläne und den überschulischen Bildungskanon definiert, um die politische Dimension der Curriculumsentwicklung aufzugreifen.

Forschungsfrage	Wie wird Curriculumsentwicklung im Schulsektor auf veränderte gesellschaftliche Entwicklungen im Kontext der Digitalisierung angepasst?
Bezugsmodell	Rahmenmodell für die vergleichende Analyse politischer Massnahmen zur Digitalisierung in der Schulbildung nach Kozma (2008)
Datenbanken	ERC, ERIC, FIS Bildung, LearnTechLib, Web of Science
Dokumententyp	Artikel in wissenschaftlichen Journals
Erscheinungsdatum	Januar 2016 – Juni 2021
Sprachen	Englisch & Deutsch

Tab. 1: Literaturrecherche Fallbeispiel 1.

Die Literaturrecherche ergab nach Eliminierung von Duplikaten 2.444 Titel. Für die Aufnahme einer Publikation wurden im anschließenden Screening-Prozess inhaltliche, theoretische, empirische und methodische Kriterien berücksichtigt. In einem ersten Screening von Titeln und Abstracts wurden Publikationen ausgeschlossen, die thematisch nicht zur Fragestellung oder zum Schulbildungsbereich passten, sowie solche ohne wissenschaftlichen Anspruch. Publikationen, die sich auf die Entwicklung der Covid-19-Pandemie oder auf fachspezifische Überlegungen konzentrierten, wurden ebenfalls ausgeschlossen, da die allgemeine und politisch geplante Lehrplanentwicklung im Fokus des Critical Reviews stand. Darüber hinaus wurden nur Veröffentlichungen in wissenschaftlichen Zeitschriften berücksichtigt.

405 Publikationen wurden nach dem Screening der Titel und Abstracts für die Volltextlektüre ausgewählt. Im Rahmen der inhaltsorientierten Kodierung der Volltexte wurde der Inhalt einer Publikation in Bezug auf die Forschungsfrage des Reviews untersucht. In einem nächsten Schritt unterzogen die Autor:innen die Publikationen einer Plausibilitätsprüfung und schätzten die Publikationen qualitativ ein. In diesem dritten Schritt wurden insbesondere die empirischen Beiträge hinsichtlich ihrer Wissenschaftlichkeit eingeordnet und überprüft, ob das methodische Vorgehen nachvollziehbar dargestellt wurde. Abbildung 1 zeigt diesen Prozess in Form eines Flowdiagramms.

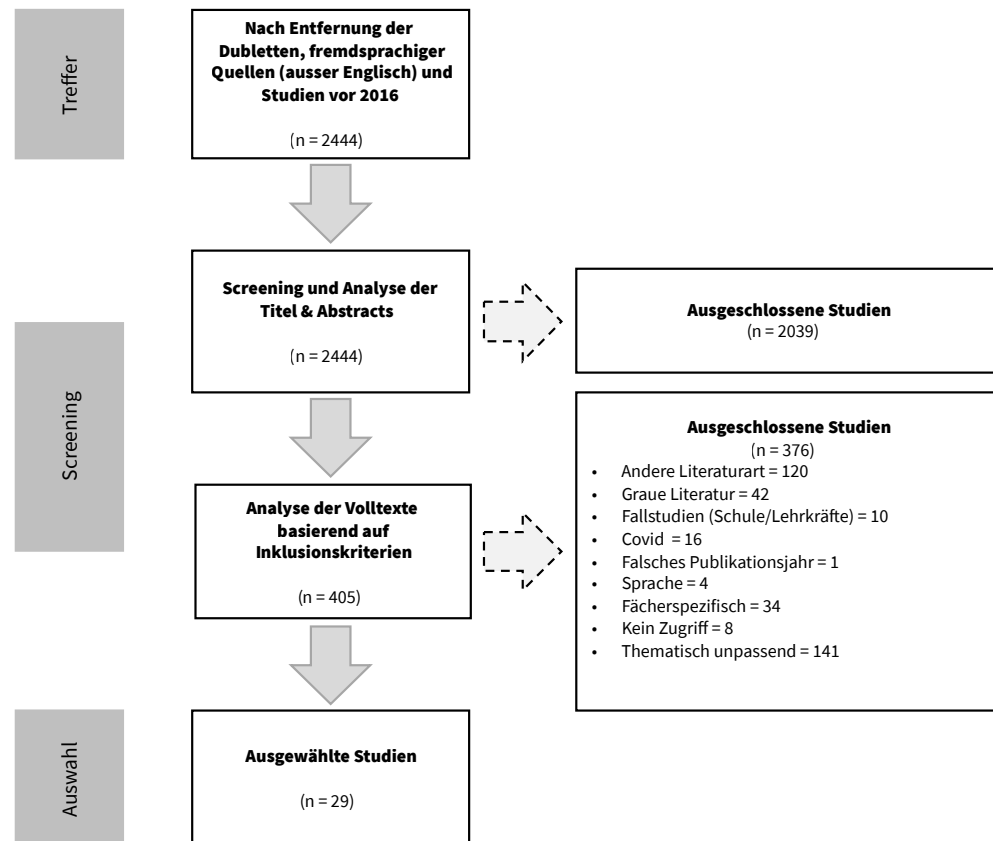


Abb. 1: Reviewprozess im Fallbeispiel 1 in Anlehnung an Page et al. (2021).

Schliesslich wurden die eingeschlossenen 29 Publikationen hinsichtlich des Bezugsmodells analysiert und deren Inhalte im Critical Review kategorisch vorgestellt sowie diskutiert.

3.2 Methodische Herausforderungen und Potenziale

Der hier exemplarisch vorgestellte Such- und Auswahlprozess hatte zum Ziel, einen Überblick über den aktuellen Forschungsstand zur Curriculumentwicklung in der digitalisierten Bildungswelt für den Schulsektor zu geben sowie die Befunde kritisch einzuordnen. Für die Literaturrecherche bedeutete dies, zum einen Suchwörter zu verwenden, die eher weit gefasst sind (z. B. «Digitalisierung»; «Schule»), zum anderen aber auch für die Forschungsfrage spezifische Suchwörter einzusetzen (z. B. «Bildungskanon»), um politisch initiierte Curriculumveränderungen aufzugreifen zu können. Bedeutsam für alle Arbeitsschritte war, dass der Begriff der Curriculumentwicklung international keinem einheitlichen Konzept folgt.

3.2.1 Literaturrecherche

Mehrsprachigkeit und Bedeutungen der Forschungsbegriffe

Eine Herausforderung bestand demnach in der sprachlichen Formulierung der Suchwörter und -strings, da die Begriffe im Deutschen und Englischen teilweise nicht einheitlich verwendet werden, in der anderen Sprache so nicht existieren oder es verschiedene sprachliche Bias im Forschungsfeld gibt (Petticrew und Roberts 2006; Marín et al. 2023). So ist das Wort «Curriculum» im deutschsprachigen Kontext v. a. mit der Ebene einer Schule, eines Fachs oder politisch konnotiert, während im Englischen auch auf die Vorbereitung und Durchführung von Unterricht referenziert werden kann. Das deutsche Suchwort «Bildungskanon» und die damit einhergehende Bedeutung ist hingegen im Englischen nicht verbreitet und mussten somit im englischen Suchstring umschrieben werden. Dementsprechend wurde für die Formulierung des Suchstrings Zeit eingeplant, um eine Balance zwischen möglichst offenen und möglichst präzisen Termini sowohl auf der Ebene der Zweisprachigkeit als auch auf der Ebene der unklaren Begrifflichkeit der Curriculumsentwicklung zu finden.

Hohe Quantität an Forschungsarbeiten

Der Schulsektor wird in hohem Mass beforscht, wodurch sich der Vorteil ergab, dass eine Vielzahl von möglichen Studien im internationalen Schulforschungssektor gefunden und somit auf eine differenzierte Grundlage für die Identifikation geeigneter Publikationen für das Review zurückgegriffen werden konnte. Dies zeigte sich bereits in den beiden Critical Reviews, die im Zuge des Projekts in den Jahren 2020 und 2021 im Bereich der Schule durchgeführt wurden (Waffner 2020; Waffner 2021). Eine Begründung hierfür kann u. a. darin liegen, dass ein Schulsystem weltweit in vielen Ländern existiert und in den meisten betrachteten Ländern ein Grossteil der Bevölkerung einer bestimmten Alterskohorte erfasst wird. Hierdurch kann eine Übertragbarkeit oder zumindest ein Vergleich mit dem deutschen Schulsystem begünstigt werden.

3.2.2 Kodierung & Synthese der Publikationen

Bezugsmodell als Orientierung im grossen und heterogenen Literaturkorpus

Durch die grosse Trefferzahl in der Literaturrecherche bestand eine Herausforderung darin, die eigene Forschungsfrage sowie Analysekatoren zu präzisieren. Parallel zur hohen Trefferzahl zeigte sich eine grosse Heterogenität sowohl in der Methodik als auch in den inhaltlichen Schwerpunkten. Hilfreich war an dieser Stelle die Integration des Bezugsmodells, um die heterogene, u. a. qualitativ durchgeführte und von offenen Forschungsfragen geprägte Literatur zu synthetisieren. Die Entscheidung zum Bezugsmodell erfolgte hier also induktiv, da es im Korpus mehrfach

zitiert wurde und einen passenden Rahmen für die Forschungsfrage im Critical Review bot. Nach der generellen inhaltlichen und qualitativen Einschätzung wurden schliesslich Publikationen aufgenommen, die zumindest ein Kriterium des Rahmenmodells aufgriffen.

Diversität der fachlichen Hintergründe

Bei der inhaltlichen Kategorisierung zeigte sich zudem die Herausforderung, dass im Schulsektor viele Veröffentlichungen fachspezifischer Natur sind. Entsprechend erschien eine Synthese in vielen Fällen nicht praktikabel, da in den Publikationen unterschiedliche inhaltliche Gesichtspunkte von curricularen Entwicklungen und damit assoziierte Kompetenzen der Schüler:innen thematisiert wurden. Auch zeigt sich in diesem Punkt eine grosse Anzahl an involvierten Fachkulturen, Bezugstheorien und Definitionen des Bildungskanons und curricularer Entwicklungen im politischen Kontext, die zum einen in ihrer Vielfalt den medienpädagogischen Diskurs bereichern. Zum anderen wird ein Vergleich und die Synthese hierdurch aber deutlich erschwert, da ähnliche Konzepte in den Fachdisziplinen anders benannt und nuanciert werden.

Thematische Herausforderung: Aktualität der Bezugsdokumente (Curricula)

Eine Besonderheit lag darin, dass viele Publikationen u.a. Dokumentenanalysen von politisch initiierten Curricula beinhalteten. So kann eine Publikation zwar in dem passenden Suchzeitraum zwischen 2016 und 2021 entstanden sein, die Bezugsdokumente, also die beforschten Curricula, können sich aber möglicherweise auf einen früheren Zeitpunkt beziehen. Diese Herausforderung ist v.a. hinsichtlich der Schnelllebigkeit und hohen Dynamik im Kontext der Digitalisierung ein wichtiger Aspekt, der in der Erstellung von Reviews generell eine Limitation darstellen kann. Insbesondere, wenn nicht deutlich wird, ob das beforschte Curriculum noch das aktuelle politische Bezugsdokument ist. Graue Literatur könnte hier zwar in manchen Fällen Inhalte für aktuelleres Geschehen in der Politik liefern, wurde aber dennoch ausgeschlossen, um den Literaturkorpus vergleichbar und synthetisierbar zu halten (vgl. Wilmers, Keller, und Anda 2021).

3.2.3 Zwischenfazit

Viele dieser Herausforderungen könnten so auch in der Durchführung anderer Arten von Reviews, z. B. dem Systematic Review auftreten. Die Besonderheit im Format des Critical Reviews macht aber zusätzlich deutlich, dass die Suche nach besonders relevanter Literatur für die Theorie- und Modellbildung erschwert werden kann, wenn die beforschten Begriffe nicht einheitlich verwendet werden und die Verschlagwortung von relevanten Publikationen sehr uneinheitlich ist (s. z. B. Wilmers et al. 2020a). Ein grosses Potenzial besteht aber auch darin, dass in einem Review

relevante Literatur aus verschiedenen Fachbereichen zusammengetragen werden kann. Somit bietet das hier durchgeführte Critical Review die Chance, verschiedene Forschungscommunities miteinander zusammenzuführen, ihre aktuellen zentralen Erkenntnisse und Diskussionen vernetzt aufzugreifen und zu reflektieren. Von diesem Überblick können sowohl Forschende als auch praktizierende Lehrkräfte profitieren.

Generell bietet das Format des Critical Reviews für die Frage danach, welchen Raum verschiedene Länder und Regionen der Bildung im digitalen Wandel in den politisch initiierten Curricula geben, grosses Potenzial, relevante Studien und Ergebnisse sowie sichtbare Tendenzen narrativ darzustellen und zu diskutieren. Durch die Vernetzung relevanter Literatur aus dem internationalen Raum sowie verschiedener Fachdisziplinen bietet es zudem die Möglichkeit, Wissen zu generieren und dementsprechend modellerweiternd zu wirken. Durch die grundlegende Orientierung an einem Bezugsmodell konnte das narrative Vorgehen unterstützt werden. Des Weiteren konnten zusätzliche Kategorien und Forschungsdesiderate durch das Critical Review in der gesichteten Literatur identifiziert und diskutiert werden. Das Format des Critical Reviews erforderte aber in vielen Einzelschritten auch die Einplanung von genügend Zeit und Ressourcen, z. B. für die Erstellung geeigneter Suchparameter oder die sprachliche und kulturelle Auseinandersetzung mit den im internationalen Kontext behandelten Inhalten der Curriculumsentwicklung.

4. Fallbeispiel 2: Gestaltung von Blended Learning-Umgebungen in der Erwachsenen- und Weiterbildung

4.1 Methodik des Critical Reviews

Das zweite Critical Review, das im Rahmen dieses Beitrags vergleichend betrachtet wird, bezieht sich auf den Bereich der Erwachsenen- und Weiterbildung. Es trägt empirische Ergebnisse zur *Gestaltung von Blended Learning (BL)-Umgebungen* zusammen. Das Thema war vom Diskurs zu digitalen Lernformaten in der Praxis der Erwachsenenbildung/Weiterbildung inspiriert. Ziel war es, effektive Variationen in der Gestaltung zu identifizieren (bspw. in der Sequenzierung von On- und Offlineanteilen oder der Lerninhalte), die sich förderlich auf Lernmotivation und -ergebnis auswirken. Entsprechend lag der Fokus auf dazu passenden empirischen Primärstudien. Die Einordnung der ausgewählten Studien erfolgte in Anlehnung an das deskriptive Orchestrierungsmodell nach Weinberger (2018). Das Modell wurde deduktiv eingesetzt; es diente der Formulierung untergeordneter Forschungsfragen sowie als Referenz für die Konzeption der Suche und die Gestaltung der Matrix zur Analyse und Synthese der Literatur.

Die Recherche in den Datenbanken FIS Bildung, ERIC, ERC, WoS, LearnTechLib und BASE erfasste den Zeitraum 2016 bis 2021. Die Suchstrings umfassten Begriffe mit thematischem Bezug aus den Kategorien Digitalisierung, Blended Learning, Erwachsenenbildung, Lernmotivation, Lernerfolg, Kursgestaltung sowie mobile Endgeräte.

Forschungsfrage	Wie können BL-Arrangements für Erwachsene förderlich für Lernmotivation und Lernerfolg orchestriert werden?
Bezugsmodell	Deskriptives Orchestrierungsmodell nach Weinberger (2018)
Datenbanken	ERIC, ERIC, FIS Bildung, LearnTechLib, Web of Science
Dokumententyp	Artikel in wissenschaftlichen Journals, Beiträge in Sammelbänden, Dissertationen
Erscheinungsdatum	Januar 2016 – November 2021
Sprachen	Englisch & Deutsch

Tab. 2: Literaturrecherche Fallbeispiel 2.

Die Suchergebnisse wurden auch hier zunächst nach Titel, Abstract und Schlagworten gesichtet und in Bezug auf den Bildungsbereich oder das Thema gefiltert. Nicht-empirische Arbeiten sowie Reviews wurden in jeweils einer eigenen Kategorie erfasst und ggf. in der Erarbeitung des theoretischen Hintergrunds und/oder zur Diskussion der Ergebnisse aus den Studien referenziert. Im Rahmen des Reviews wurde BL als Format definiert, das Online- und Präsenzanteile kombiniert. Aus diesem Grund wurden empirische Studien zu reinen Online-Formaten nicht berücksichtigt. Darunter fielen auch solche zu Online-Formaten unter Pandemiebedingungen.

Alle Publikationen die im Laufe des Screenings als initial passend zu Thema bzw. Bildungsbereich identifiziert oder im ersten Screening noch nicht eindeutig bewertet werden konnten, wurden anschliessend im Volltext geprüft. Dabei wurden sie nach formalen Kriterien und anhand einer eigens dafür erstellten Analysematrix auf ihren Gehalt zur Fragestellung hin untersucht. Zudem wurden eine Plausibilitätsprüfung und eine Einschätzung der Qualität entlang üblicher Kriterien (Reliabilität, Validität, Generalisierbarkeit) durchgeführt. Alle Schritte wurden von beiden Autor:innen im Review vorgenommen und ggf. kommunikativ validiert. Abbildung 2 dokumentiert den Ablauf des Screenings.

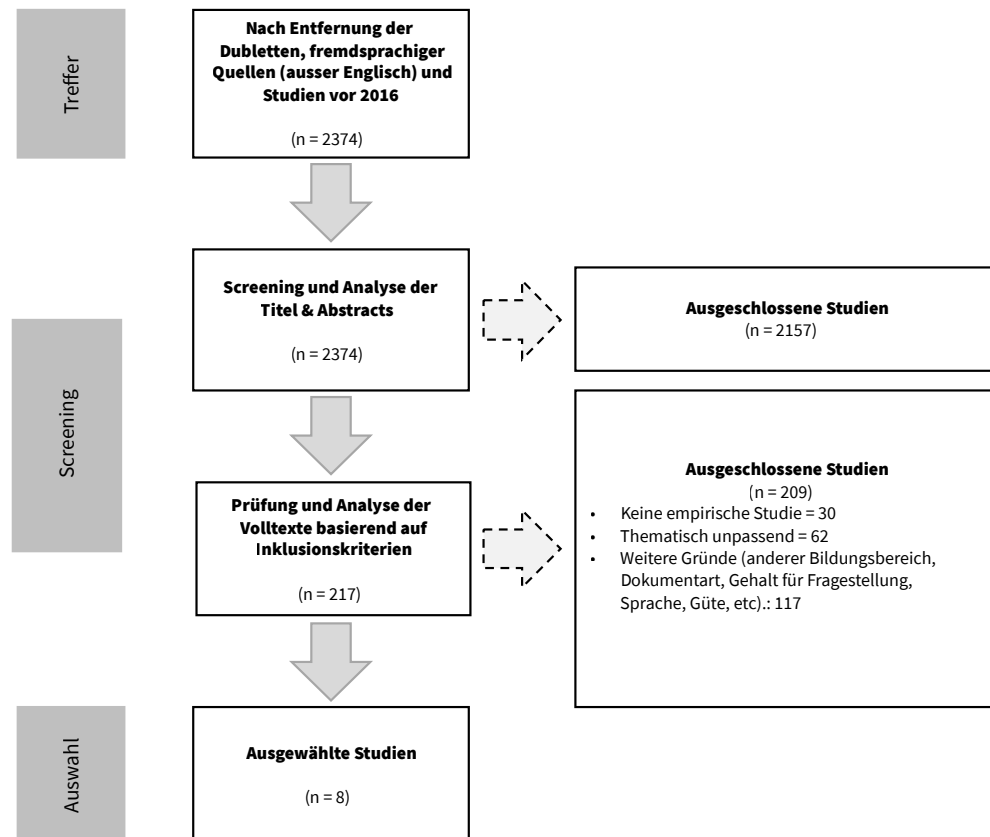


Abb. 2: Reviewprozess im Fallbeispiel 2 in Anlehnung an Page et al. (2021).

Die ausgewählten Studien wurden anschliessend im Rahmen des Critical Reviews mit Blick auf das ausgewählte Bezugsmodell in ihren relevanten Ergebnissen beschrieben und diskutiert.

4.2 Methodische Herausforderungen und Potenziale

Die Operationalisierung der Forschungsfrage bzw. untergeordneten Forschungsfragen in Suchstrings und die nachfolgenden Analyseschritte waren von der Diffusität des Begriffs BL bzw. der begrifflichen Vielfalt in diesem Kontext geprägt. So wurden z. B. andere Begriffe wie «hybrides Lernen» synonym verwendet. Auch wurde unter BL eine Mischung aus unterschiedlichen Arbeitsformen bzw. Anwendungen verstanden. Dies machte eine iterative Reflexion sowohl in der konzeptionellen Erarbeitung als auch in der späteren theoretischen Auseinandersetzung notwendig.

4.2.1 Literaturrecherche

Nationale und internationale Spezifika des Bildungsbereichs

Bei der Sichtung der Forschung im Bereich Erwachsenenbildung/Weiterbildung in den vorangegangenen Reviews (Koschorreck und Gundermann 2020, 2021) zeigte sich, dass der Bildungsbereich im Vergleich zu anderen Bildungssektoren wie Schule wenig beforscht ist. Dieses Muster wiederholte sich auch in diesem Fall und schlug sich in einer geringen Zahl deutschsprachiger Studien im Korpus nieder (7,2%), weshalb das Screening über Artikel in wissenschaftlichen Journals hinaus auf Beiträge in Sammelbänden und Dissertationen ausgeweitet wurde. In Bezug auf die internationale Forschung zeigte sich ausserdem, dass die Studien mehrheitlich zu Fragen der «Erwachsenenbildung» («Adult Education») im Kontext Hochschule durchgeführt wurden und werden: Die hierzulande übliche Trennung der Bildungsbereiche Hochschule und Erwachsenenbildung/Weiterbildung ist der historischen Entwicklung des Bereichs EB/WB und seiner besonderen Strukturiertheit in Deutschland geschuldet (Schrader und Martin 2021). Dies führt bei einer internationalen Recherche zu einer hohen Zahl von Treffern und macht die Identifikation von internationalen Studien, die sich genuin der Erwachsenenbildung bzw. Weiterbildung widmen, zu einer Herausforderung. Darin liegt aber auch die Chance, das eigene Verständnis des Gegenstands zu reflektieren und zu schärfen, wenn bereichsfremde Arbeiten mangels passender Verschlagwortung und/oder Informationen in die Prüfung der Volltexte gelangen.

Bezugsmodell als Werkzeug und Herausforderung

Der Bezug auf ein passendes Modell zum Thema erleichterte zunächst die Konzeption des Reviews deutlich, z. B. hinsichtlich der Formulierung der leitenden Forschungsfragen oder der Festlegung der passenden Studienarten. Gleichwohl war wie in Fallbeispiel 1 die Auswahl passender Suchbegriffe und -strings eine Herausforderung, wie sich nach der ersten Testrecherche zeigte. Beispielsweise wurde der Begriff der Orchestrierung in keinem Beitrag verwendet, auch wenn Fragen der Gestaltung der BL-Umgebung thematisiert wurden. Dementsprechend mussten die Ebenen des deskriptiven Modells im Abgleich mit der tatsächlich verwendeten Verschlagwortung der Beiträge in sinnvolle Suchstrings operationalisiert werden. Dabei musste wiederholt zwischen Offenheit und Präzision abgewogen werden (z. B. Verwendung von «Adaptivität» versus Teilmengen mit und ohne Trunkierungen wie «Feedback*»). Diese Entscheidungen können jedoch die Anzahl der nicht verwertbaren Treffer deutlich erhöhen, was sich im Vorfeld kaum abschätzen lässt. Darüber hinaus erschwerte der Modellbezug das Screening bzw. die Auswahl der Studien, wenn sich wie in diesem Fall die Studienlage als dünn und methodisch heterogen herausstellte: Eine sehr strenge Orientierung am Modell führte dazu, dass

tendenziell alle Titel im Volltext geprüft werden mussten, deren Anschlussfähigkeit an das Modell implizit war, aber nicht durch das Screening von Titel, Abstract und Schlagworten klar bestimmt werden konnte.

4.2.2 Kodierung und Synthese

Erwartete und tatsächliche Studienlage hoch different

Auch die spätere Auswahl der im Review analysierten Studien war stark davon abhängig, wie enggeführt oder weit der Bezug der Ergebnisse zum Modell durch die Erarbeitenden ausgelegt wurde. Ursprünglich sollten vor allem Ergebnisse aus Interventionsstudien bzw. experimentellen Studien mit rigider Kontrolle der Bedingungen im Mittelpunkt der Synthese stehen. Die gefundenen Studien hatten jedoch nahezu durchweg eher evaluativen Charakter. Interventionsstudien bzw. experimentelle Studien konnten entweder nicht identifiziert werden oder wurden aus verschiedenen Gründen ausgeschlossen (anderer Bildungsbereich, reine Online-Angebote etc.). Dementsprechend waren die Effekte z. B. nicht eindeutig auf das jeweilige Arrangement oder bestimmte Aktivitäten zurückführbar. Die Anlage der analysierten Beiträge erschwerte die Bearbeitung der Fragestellungen zusätzlich, denn ihr Informationsgehalt war diesbezüglich fast durchgehend lückenhaft, was ihre Analyse einschränkte. Dadurch war die Aggregation relevanter Daten auf höherer Abstraktionsebene nicht möglich.

Erschwert wurde die Identifikation passender Studien durch den Umstand, dass die Autor:innen des Reviews nicht selbst im Forschungsfeld BL aktiv sind. Darin liegt einerseits der Vorteil einer grösseren Unvoreingenommenheit in der späteren Studienausswahl, andererseits war deshalb eine intensive Einarbeitung notwendig, die aber in der Praxis durch den Projektzeitplan limitiert war.

4.2.3 Zwischenfazit

Mit Blick auf das Thema bietet das Format des Critical Reviews durch seinen eher narrativen Charakter sowohl eine gute Schablone, um relevante Einzelergebnisse intensiv zu diskutieren und systematisch zu reflektieren als auch den Raum, in der Erarbeitung auf die vorgefundene Studienlage zu reagieren. Die im Vergleich zu Systematic Reviews in Bezug auf Theorie und Design der rezipierten Studien eklektischer angelegte Studienausswahl und -analyse ermöglicht dabei, sowohl multifacetierte Fragestellungen zu bearbeiten, die direkt an Anforderungen der Bildungspraxis anknüpfen, als auch die theoretische und empirische Sättigung des gewählten Forschungsgegenstandes zumindest grob einschätzen zu können. Die Aufdeckung dieser Aspekte im Rahmen des Critical Reviews half auch dabei, methodische Polaritäten zu identifizieren und Herausforderungen in der Beforschung des Gegenstandes zu skizzieren. Nach dem Verständnis der Autor:innen dieses Beitrags liegt der

Charakter eines Critical Reviews auch in der Diskussion der Ergebnisse aus Sicht der Bildungspraxis. Allerdings war der diesbezügliche Orientierungswert der Ergebnisse im vorliegenden Fall durch die geringe Zahl an dazu vorhandenen Studien deutlich eingeschränkt. Dementsprechend wurden vor allem mögliche Forschungsperspektiven hergeleitet und Potenziale zur Verbesserung der Vergleichbarkeit von Studien im Themenfeld reflektiert. Konkret wurde eine Schablone skizziert, in der Eigenschaften der untersuchten BL-Angebote standardisiert berichtet werden könnten. Dies kann einen Beitrag zu methodischer Vielfalt leisten, aber auch die Rezipierbarkeit des Gegenstands im Sinne von Good-Practice-Beispielen im Bildungsbereich fördern.

5. Diskussion

Dieser Artikel illustriert in verschiedenen Facetten divergierende, aber auch übereinstimmende Herausforderungen und Potenziale von Forschungssynthesen. Diese beziehen sich auf methodische und arbeitspraktische Aspekte, die sich in zwei Critical Reviews im Bereich *Bildung im digitalen Wandel* in verschiedenen Bildungssektoren gezeigt haben. Im Folgenden werden ausgewählte Erkenntnisse der Fallbeispiele reflektiert und anschliessend methodologisch eingeordnet. Darüber hinaus werden Strategien beschrieben, wie man möglichen Herausforderungen begegnen kann. Die aufgeführten Herausforderungen und Potenziale beziehen sich nicht ausschliesslich auf die Durchführung eines Critical Reviews, sondern können in ähnlicher Form auch in anderen Reviewformaten auftreten (Bedenlier et al. 2020).

5.1 Methodische Erkenntnisse und Implikationen zu Forschungssynthesen in der Bildungsforschung

Critical Reviews basieren auf einer umfassenden Literatursuche, die als Grundlage für die systematische Zusammenfassung von Forschungsergebnissen dient. Eine Auslotung der Suchanfragen zwischen Präzision und Offenheit (Bedenlier et al. 2020; Higgins 2020) ist essenziell für die Durchführung von Forschungssynthesen. Zudem steigt die Diffusität, wenn Forschungsarbeiten auf nicht einheitlich definierten Termini beruhen und in der internationalen Forschungscommunity kein Konsens über die zu beforschenden Konzepte und Konstrukte existiert (Buntins, Kerres, und Heinemann 2021). Des Weiteren ist in der Bildungswissenschaft die Komplexität von Lern- und Bildungssituationen zu beachten, welche klare (kritische) Einordnungen und Bewertungen der Forschungslage erschwert (Hammersley 2020). Im Critical Review bekommt diese Komplexität umso mehr Gewicht, da hier häufig eine eher offene Forschungsfrage fokussiert wird und die teilweise vorhandene qualitative Forschung der zu synthetisierenden Literatur die Komplexität des

Forschungsgegenstands im Review unterstreicht (Higgins 2020; Petticrew 2015). Forschungssynthesen haben in den Bildungswissenschaften keine lange Tradition wie beispielsweise in der Medizin (Borrego, Foster, und Froyd 2014; Sutton et al. 2019; Torgerson, Hall, und Lewis-Light 2021) und müssen hier dementsprechend kohärent dem breiten Forschungskontext (Borrego, Foster, und Froyd 2014) angepasst werden, um der eigenen Komplexität von Lehr- und Lernsituationen gerecht zu werden.

Grundsätzlich scheint es nicht möglich zu sein, von der Trefferzahl auf die empirische Ergiebigkeit des durch die Recherche(n) entstehenden Korpus zu schließen. Dabei kann je nach Fragestellung und Thema der Informationsgehalt von Studien selbst dann unbefriedigend sein, wenn Autor:innen die üblichen wissenschaftlichen Gütekriterien einhalten und darlegen. Eine Herausforderung für Reviews in der Bildungsforschung besteht also darin, dass sich aufgrund der fachspezifischen Gegebenheiten weder einfach abschätzen lässt, wie viel Literatur in einer Recherche gefunden wird, noch wie gut diese zur Beantwortung einer Reviewfrage herangezogen werden kann. Dementsprechend gestaltet sich die Studienlage in der Bildungsforschung sehr heterogen (Kröner, Christ, und Penthin 2019; Reich-Claassen und Tippelt 2010). Heterogenität bezog sich im Literaturkorpus der beiden Fallbeispiele sowohl auf die Vielzahl der verwendeten Studiendesigns, von vorhandenen Publikationsformaten und von inhaltlichen und interdisziplinären Ausrichtungen mit jeweils unterschiedlichen und häufig fachspezifischen Ausprägungen. In der Folge liefern auch sehr aufwendige Suchen oft zu wenige oder zu viele Ergebnisse und erfordern im weiteren Verlauf eine Nachjustierung der Suchparameter, zusätzliche Testsuchen und generell flexiblere Recherchestrategien.

Letztendlich kann durch diese Faktoren sowohl ein Mangel (Fallbeispiel 2) als auch ein Überfluss (Fallbeispiel 1) an Studien entstehen. Ein Mangel stellt aus Sicht der Autor:innen vor allem mit Blick auf mögliche Forschungsdesiderate ein relevantes Ergebnis dar und bietet die Chance, die Ergebnisse der ausgewählten Studien detaillierter und intensiver zu reflektieren. Zudem kann z. B. auf methodischer Ebene die Entscheidung getroffen werden, die einzuschließenden Publikationstypen auszuweiten. Einem Überfluss kann mit entsprechender Reflexion der eigenen Analysestrategie und ggf. einer passenden (Re-)Fokussierung begegnet werden.

Eine Stärke von Forschungssynthesen insgesamt liegt in der Möglichkeit, internationale Forschungsliteratur einzubeziehen und damit unterschiedliche Regionen weltweit zu berücksichtigen. Wird ein Review auf Grundlage von Forschungsliteratur in verschiedenen Sprachen durchgeführt, so bedeutet die Formulierung des Suchstrings weitaus mehr als eine reine Übersetzung. Denn Begriffe und damit assoziierte Konzepte existieren eventuell nicht in anderen Ländern und Sprachen oder es werden gänzlich andere Begriffe für ähnliche Ideen genutzt. Eine gründliche Recherche ist demnach Voraussetzung für die adäquate Erstellung eines mehrsprachigen Suchstrings. Nicht selten bezeichnen entweder zwei völlig verschiedene

Begriffe aus unterschiedlichen Kontexten auf Deutsch und Englisch dieselbe Konstellation oder aber zwei vermeintlich ähnliche Begriffe zwei ganz unterschiedliche Situationen. In beiden Fällen besteht eine erhöhte Gefahr, Verzerrungen im Review zu produzieren. Diesen Verzerrungen kann in einem Critical Review zumindest damit begegnet werden, dass Unterschieden in den Definitionen durch die qualitative Einordnung und das beschreibende Format im Review Raum gegeben werden kann.

Critical Reviews bieten sich aufgrund ihres reflexiven Charakters insbesondere an, Bezugsmodelle aufzunehmen, wie in beiden Fallbeispielen geschehen. Dies kann bereits in der Konzeption eines Critical Reviews oder in der späteren Synthese der Ergebnisse Anwendung finden und so die Erarbeitung strukturieren (Newman und Gough 2020). Sollten auch modellbildende Studien in die Analyse aufgenommen werden, müssen diese vor dem Hintergrund des Bezugsmodells u.U. besonders reflektiert werden. Die Entscheidung, ob ein Bezugsmodell verwendet werden sollte, und die Wahl der jeweiligen Strategie der Modellintegration muss deshalb in Aufwand und Nutzen mit Blick auf das Thema und das eigene Vorwissen zum Forschungsfeld sorgfältig abgewogen werden.

Die kritische Ebene eines Critical Reviews ist, wie diese Diskussion zeigt, wesentlicher Bestandteil dieser Form einer Forschungssynthese. Um alle als relevant identifizierten Publikationen kritisch einordnen und analysieren zu können, bedarf es viel Zeit und Expertise, die sich nicht nur auf den Bereich der Bildung im digitalen Wandel begrenzt. In einem dynamischen Themen- und Forschungsfeld wie demjenigen der Digitalisierung kann eine entsprechende Analyse durch Faktoren in ihrer Aussagekraft eingeschränkt werden, die nicht in der Kontrolle der Erarbeitenden liegen, beispielsweise durch neue Trends mit Schnittmengen zum Thema oder Ereignisse wie die Covid-19 Pandemie, deren Auswirkungen nicht abschätzbar sind. Generell fällt es bei Themenfeldern mit dynamischer Entwicklung im Reviewprozess schwer, die Literaturrecherche zu einem bestimmten Zeitpunkt abzuschliessen, während aufgrund der Aktualität der Thematik fortlaufend und zügig weitere Literatur publiziert wird. Ebenso wirkt sich auch der eigentlich notwendige und normale Zeitverzug zwischen der Veröffentlichung einer Publikation und der Einspeisung dieser Publikation in ein Datenbanksystem ungünstig aus, da auch diese dazu führt, dass besonders aktuelle Literatur in einem Review möglicherweise nicht mehr berücksichtigt werden kann. Ein iteratives Vorgehen in der Recherche kann insofern unterstützend wirken, solange die grundsätzliche Systematik in der Bearbeitung beibehalten wird.

Das narrative Format eines Critical Reviews kann insbesondere auch im Themenfeld Lehren und Lernen mit digitalen Medien, das zwar einerseits global angelegt ist, aber durch die jeweiligen länderspezifischen Bildungssysteme mit ihren kulturellen und politischen Kontexten geprägt wird, einen Überblick über den Forschungsdiskurs leisten. Critical Reviews können einen Schwerpunkt in der Beschreibung und

Diskussion relevanter Aussagen setzen, ohne dass eine Vergleichbarkeit mit identischen Kriterien in der Studienauswahl gegeben sein muss. Auch unabhängig vom Spannungsfeld zwischen internationalen und regionalen Kontexten erleichtert ein solches Vorgehen grundsätzlich die Integration sehr heterogener Forschung. Mithilfe einer eher offenen Kodierung (Sutcliffe, Oliver, und Richardson 2017) lässt sich demnach eine uneinheitliche Forschungslage abbilden und trotzdem die Grundlage für eine Analyse bilden, die über eine einfache Kartierung der Forschungslandschaft hinausgeht.

5.2 Implikationen auf der Arbeitsebene

Hinsichtlich der Erarbeitung von Forschungssynthesen allgemein lassen sich einige Aspekte auf der Arbeitsebene festhalten, die zusammenfassend hilfreich für eine erfolgreiche Durchführung erscheinen.

Der Faktor Zeit stellt mit Blick auf die hier vorgestellten Fallbeispiele in mehrfacher Hinsicht eine Limitation dar. Zunächst steht Digitalisierung nicht für sich, sondern wird immer in Verbindung mit anderen bildungsrelevanten Themen diskutiert und beforscht. Je nach (Projekt-)Kontext ergibt sich, dass Reviews im Gegensatz zu anderen wissenschaftlichen Beiträgen ggf. von nicht im Themenfeld aktiven Forschenden erstellt werden. In solchen Fällen besteht Bedarf nach zusätzlicher Expertise und/oder genügend Zeit für die Mitarbeitenden im Projekt, sich vor der Literaturrecherche in das Bezugsthema einzuarbeiten. Ist dies aufgrund begrenzter Ressourcen nur eingeschränkt möglich, ist der Austausch mit Expert:innen zum Thema eine Möglichkeit, vor allem gegenüber der Alternative alleiniger Erarbeitung, die entsprechenden Schritte, Entscheidungen und Analyseergebnisse kommunikativ zu validieren und damit abzusichern. Darüber hinaus können so ggf. weitere Aspekte des Themas durch die Diskussion aufgedeckt werden. In einem Review, das sich in der Regel in einem genau eingegrenzten Bearbeitungszeitraum bewegt und mit der Auswertung an einen bestimmten Suchzeitraum anschließt, ist es essenziell, für diese ressourcenintensiven Arbeitsschritte über die reguläre Bearbeitungszeit hinaus ausreichend und flexibel Zeit einzuplanen.

Für den Umgang mit Mehrsprachigkeit der eingeschlossenen Studien sollten genügend Zeit und Ressourcen eingeplant werden, um passende Suchen zu definieren. Die einzelnen Arbeitsschritte der systematischen Recherche und Analyse sowie des Schreibprozesses sollten im Vorhinein realistisch geplant werden, denn hier kann es passieren, dass wenige Mitarbeitende einen Datensatz von mehreren Tausend Publikationen auf ihre inhaltliche Passung zum Thema prüfen müssen. Hinzu kommt, dass Einträge in den Datenbanken teilweise unvollständig sind, was sich z. B. an fehlenden Abstracts oder mangelhafter oder gar irreführender Verschlagwortung zeigt.

Dies führt dazu, dass tendenziell eine grössere Zahl von Studien im Volltext beschafft und geprüft werden muss. Die bisherigen Erfahrungen haben gezeigt, dass diese Zahl selten unter zehn Prozent des Gesamtumfangs des Literaturkorpus liegt.

Ein Potenzial im Arbeitsprozess von Reviews im Allgemeinen zeigte sich im vorliegenden Projekt durch eine Teamgestaltung, die methodisches und inhaltliches Fachwissen zusammenbringt. Indes bleibt anzumerken, dass aber auch mit einem solchen Vorgehen aufgrund der Schnelllebigkeit im Themenfeld Digitalisierung und der andauernden methodischen Diskussion um Reviews in der Bildungsforschung eine durchgehende intensive und kritische Auseinandersetzung mit den Fragestellungen im Themenfeld und den Arbeitsprozessen nötig bleibt.

6. Fazit

Critical Reviews können in der Bildungsforschung im Kontext der Digitalisierung von besonderem Nutzen sein: Aufgrund der gründlichen Recherche können sie den Forschungsstand zunächst relativ umfassend und systematisch erfassen, während die im Vergleich zu anderen Verfahren der Forschungssynthesen etwas weniger aufwendige Analyse eine relativ schnelle Auswertung und Darstellung der Ergebnisse ermöglicht. Damit können Critical Reviews vergleichsweise zeitnah auf Forschungsfragen reagieren und aktuelle Forschungsdebatten untermauern (Wright und Michailova 2022). So kann ein Critical Review in einem dynamischen Forschungsfeld mit vielen offenen Fragen eine Möglichkeit darstellen, um theoretische, empirische oder forschungsmethodische Potenziale aufzudecken und die Planung künftiger und kumulativ angelegter Forschungsaktivitäten zu unterstützen (Terhart 2021, 157f.). In der Regel betrachten Critical Reviews ein Thema ausserdem eher aus einer «Bigger Picture» Perspektive. Dadurch werden sie inhaltlich auch für die Praxis und politische Akteure interessant und begünstigen – eine entsprechende Aufbereitung in weiteren Transferderivaten (z. B. als Podcasts oder Blogbeiträge) vorausgesetzt – deren Zugang zu den rezipierten Forschungsergebnissen. Critical Reviews können so bei gleichzeitig realistischer Einordnung ihrer Aussagekraft einen relevanten Mehrwert im Themenfeld der (digitalen) Bildung bieten. Grundsätzlich bleibt indes festzuhalten, dass eine weitere methodische Schärfung des Reviewformats in der Bildungsforschung wünschenswert wäre und seine Anwendung zukünftig erleichtern würde.

Literatur

- 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: Methodology, Perspectives and Application*, herausgegeben von Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond, und Katja Buntins, 111–27. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-27602-7_7.
- Booth, Andrew, Anthea Sutton, und Diana Papaioannou. 2016. *Systematic Approaches to a Successful Literature Review*. Second edition. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.
- Borrego, Maura, Margaret J. Foster, und Jeffrey E. Froyd. 2014. «Systematic Literature Reviews in Engineering Education and Other Developing Interdisciplinary Fields». *J. Eng. Educ.* 103 (1): 45–76. <https://doi.org/10.1002/jee.20038>.
- Buntins, Katja, Michael Kerres, und Anna Heinemann. 2021. «A scoping review of research instruments for measuring student engagement: In need for convergence». *International Journal of Educational Research Open* 2:100099. <https://doi.org/10.1016/j.ijedro.2021.100099>.
- Cooper, Harris M., Larry V. Hedges, und Jeff C. Valentine, Hrsg. 2019. *The handbook of research synthesis and meta-analysis*. Unter Mitarbeit von Harris Cooper, Larry V. Hedges und Jeffrey C. Valentine. 3rd edition. New York: Russell Sage Foundation. <https://doi.org/10.7758/9781610448864>.
- Gough, David, und Sandy Oliver, Hrsg. 2017. *An Introduction to Systematic Reviews*. Second edition. Los Angeles: SAGE.
- 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>.
- Hammersley, Martyn. 2020. «Reflections on the Methodological Approach of Systematic Reviews». In *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*, herausgegeben von Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond, und Katja Buntins, 23–39. Wiesbaden: Springer VS. https://link.springer.com/chapter/10.1007/978-3-658-27602-7_2. https://doi.org/10.1007/978-3-658-27602-7_2.
- Heinemann, Anna, Jens Leber, Pia Sander, und Chirine Ahmad. 2022. «Strategien und Maßnahmen für die schulische Curriculumsentwicklung im Kontext der Digitalisierung. Ein Critical Review zur vergleichenden Analyse politischer Initiativen». In *Bildung im digitalen Wandel: Lehren und Lernen mit digitalen Medien*, herausgegeben von Annika Wilmers, Michaela Achenbach, und Carolin Keller. 1. Auflage, 69–102. Digitalisierung in der Bildung 3. Münster: Waxmann. <https://doi.org/10.31244/9783830996224.03>.

- Higgins, Julian P. T. 2020. *Cochrane Handbook for Systematic Reviews of Interventions*. Second edition. Wiley Cochrane. Hoboken, NJ: Wiley-Blackwell. <https://www.training.cochrane.org/handbook>.
- Kerres, Michael. 2018. *Mediendidaktik: Konzeption und Entwicklung digitaler Lernangebote: De Gruyter*. <https://doi.org/10.1515/9783110456837>.
- Kerres, Michael, Pia Sander, und Bettina Waffner. 2022. «Zum Zusammenwirken von Bildungsforschung und Bildungspraxis: Gestaltungsorientierte Bildungsforschung als Ko-Konstruktion». *Bildungsforschung*, Nr. 2: Bildungsforschung: Gemeinsam mit Bildungspraxis? Wege, Dynamiken, Klärungen. <https://doi.org/10.25539/BILDUNGSFORSCHUN.V012.935>.
- Koschorreck, Jan, und Angelika Gundermann. 2020. «Die Implikationen der Digitalisierung für das Lehrpersonal in der Erwachsenen- und Weiterbildung: Ein Review ausgewählter empirischer Ergebnisse und weiterer theoriebildender Literatur». In *Bildung im digitalen Wandel. Die Bedeutung für das pädagogische Personal und für die Aus- und Fortbildung*, herausgegeben von Annika Wilmers, Carolin Anda, Carolin Keller, und Marc Rittberger, 159–93. Münster: Waxmann.
- Koschorreck, Jan, und Angelika Gundermann. 2021. «Die Bedeutung der Digitalisierung für das Management von Weiterbildungsorganisationen». In *Bildung im digitalen Wandel. Organisationsentwicklung in Bildungseinrichtungen.*, herausgegeben von Annika Wilmers, Michaela Achenbach, und Carolin Keller, 161–92. Münster: Waxmann. <https://doi.org/10.31244/9783830994558.06>.
- Koschorreck, Jan, und Angelika Gundermann. 2022. «Orchestrierung von Blended Learning in der Erwachsenen- und Weiterbildung: Ein Critical Review». In *Bildung im digitalen Wandel: Lehren und Lernen mit digitalen Medien*, herausgegeben von Annika Wilmers, Michaela Achenbach, und Carolin Keller. 1. Auflage, 161–86. Digitalisierung in der Bildung 3. Münster: Waxmann. <https://doi.org/10.31244/9783830996224.06>.
- Kozma, Robert B. 2008. «Comparative Analysis of Policies for ICT in Education». In *International Handbook of Information Technology in Primary and Secondary Education*. Bd. 20, herausgegeben von Joke Voogt, und Gerald Knezek, 1083–96. Boston, MA: Scholars Portal. https://doi.org/10.1007/978-0-387-73315-9_68.
- Kröner, Stephan, Alexander Christ, und Marcus Penthin. 2019. «DiKuBi-Meta [TP2]: Forschungssynthesen zur Digitalisierung in der kulturellen Bildung». In *Forschung zur Digitalisierung in der kulturellen Bildung*, herausgegeben von Benjamin Jörissen, Stephan Kröner, und Lisa Unterberg, 215–28. Kulturelle Bildung und Digitalität 1. München: kopaed.
- Kultusministerkonferenz. 2021. «Lehren und Lernen in der digitalen Welt: Ergänzung zur Strategie der Kultusministerkonferenz <Bildung in der digitalen Welt>». https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf.

- Marín, Victoria I., Katja Buntins, Svenja Bedenlier, und Melissa Bond. 2023. «Invisible Borders in Educational Technology Research? A Comparative Analysis». *Educational technology research and development: ETR & D*, 1–22. <https://doi.org/10.1007/s11423-023-10195-3>.
- Newman, Mark, und David Gough. 2020. «Systematic Reviews in Educational Research: Methodology, Perspectives and Application». In *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*, herausgegeben von Olaf Zawacki-Richter, Michael Kerres, Svenja Bedenlier, Melissa Bond, und Katja Buntins, 3–22. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-27602-7_1.
- Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer et al. 2021. «The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews». *BMJ (Clinical research ed.)* 372:n71. <https://doi.org/10.1136/bmj.n71>.
- Petticrew, Mark. 2015. «Time to Rethink the Systematic Review Catechism? Moving from «What Works» to «What Happens»». *Systematic reviews* 4:36. <https://doi.org/10.1186/s13643-015-0027-1>.
- Petticrew, Mark, und Helen Roberts. 2006. *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Wiley Interscience. <https://doi.org/10.1002/9780470754887>.
- Pigott, Terri D., Charles Tocci, Ann Marie Ryan, und Aaron Galliher. 2021. «Quality of Research Evidence in Education: How Do We Know?». *Review of Research in Education* 45 (1): vii–xii. <https://doi.org/10.3102/0091732X211001824>.
- Reich-Claassen, Jutta, und Rudolf Tippelt. 2010. «Stichwort: Evidenzbasierung». *DIE Zeitschrift für Erwachsenenbildung* (4): 22–23. <https://www.die-bonn.de/id/9077/about/html>.
- Schrader, Josef, und Andreas Martin. 2021. «Weiterbildungsanbieter in Deutschland: Befunde aus dem DIE-Weiterbildungskataster». *ZfW* 44 (3): 333–60. <https://doi.org/10.1007/s40955-021-00198-z>.
- Sutcliffe, Katy, Sandy Oliver, und Michelle Richardson. 2017. «Describing and Analysing Studies». In *An Introduction to Systematic Reviews*, herausgegeben von David Gough, und Sandy Oliver. Second edition, 123–44. Los Angeles: SAGE.
- Sutton, Anthea, Mark Clowes, Louise Preston, und Andrew Booth. 2019. «Meeting the Review Family: Exploring Review Types and Associated Information Retrieval Requirements». *Health information and libraries journal* 36 (3): 202–22. <https://doi.org/10.1111/hir.12276>.
- Terhart, Ewald. 2021. «Zukunftsthemen der Schulforschung». In *Schule und Schulpolitik während der Corona-Pandemie: Nichts gelernt?* herausgegeben von Detlef Fickermann, Benjamin Edelstein, Julia Gerick, und Kathrin Racherbäumer. 1. Auflage, 147–75. Die Deutsche Schule. Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis - Beiheft 18. Münster: Waxmann. <https://doi.org/10.31244/9783830994589.07>.
- Torgerson, Carole, Jill Hall, und Kate Lewis-Light. 2021. «Systematic Reviews». In *Research methods and methodologies in education*, herausgegeben von Rob Coe, Michael Waring, Larry V. Hedges, und Laura D. Ashley. 3rd edition, 166–79. Los Angeles: SAGE.

- Waffner, Bettina. 2020. «Unterrichtspraktiken, Erfahrungen und Einstellungen von Lehrpersonen zu digitalen Medien in der Schule». In *Bildung im digitalen Wandel. Die Bedeutung für das pädagogische Personal und für die Aus- und Fortbildung*, herausgegeben von Annika Wilmers, Carolin Anda, Carolin Keller, und Marc Rittberger, 57–102. Münster: Waxmann. <https://doi.org/10.25656/01:20766>.
- Waffner, Bettina. 2021. «Schulentwicklung in der digital geprägten Welt: Strategien, Rahmenbedingungen und Implikationen für Schulleitungshandeln». In *Bildung im digitalen Wandel. Organisationsentwicklung in Bildungseinrichtungen.*, herausgegeben von Annika Wilmers, Michaela Achenbach, und Carolin Keller, 67–103. Münster: Waxmann. <https://doi.org/10.25656/01:23605>.
- Weinberger, Armin. 2018. «Orchestrierungsmodelle und -szenarien technologie-unterstützten Lernens». In *Digitalisierung und Bildung*, herausgegeben von Silke Ladel, 117–39. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-18333-2_7.
- Weingart, Peter, Holger Wormer, Thomas Schildhauer, Birte Fähnrich, Otfried Jarren, Christoph Neuberger, Jan-Hendrik Passoth, und Gert G. Wagner. 2022. «Gute Wissenschaftskommunikation in der digitalen Welt: Politische, ökonomische, technische und regulatorische Rahmenbedingungen ihrer Qualitätssicherung». *Wissenschaftspolitik im Dialog* (19). https://www.bbaw.de/files-bbaw/publikationen/wissenschaftspolitik_im_dialog/BBAW_Wissenschaftspolitik-Dialog_19_2022.pdf.
- Wetterich, Cita, und Erik Plänitz. 2021. *Systematische Literaturanalysen in den Sozialwissenschaften: Eine praxisorientierte Einführung*. Opladen, Berlin, Toronto: Verlag Barbara Budrich.
- Wilmers, Annika, Michaela Achenbach, und Carolin Keller, Hrsg. 2021. *Bildung im digitalen Wandel. Organisationsentwicklung in Bildungseinrichtungen*. Münster: Waxmann. <https://doi.org/10.31244/9783830994558>.
- Wilmers, Annika, Michaela Achenbach, und Carolin Keller, Hrsg. 2022. *Bildung im digitalen Wandel: Lehren und Lernen mit digitalen Medien*. 1. Auflage. Digitalisierung in der Bildung 3. Münster: Waxmann. <https://doi.org/10.31244/9783830996224>.
- Wilmers, Annika, Carolin Anda, Carolin Keller, Michael Kerres, und Barbara Getto. 2020a. «Reviews zur Bildung im digitalen Wandel: Eine Einführung in Kontext und Methodik». In *Bildung im digitalen Wandel. Die Bedeutung für das pädagogische Personal und für die Aus- und Fortbildung*, herausgegeben von Annika Wilmers, Carolin Anda, Carolin Keller, und Marc Rittberger, 7–29. Münster: Waxmann. <https://doi.org/10.31244/9783830991991.01>.
- Wilmers, Annika, Carolin Anda, Carolin Keller, und Marc Rittberger, Hrsg. 2020b. *Bildung im digitalen Wandel. Die Bedeutung für das pädagogische Personal und für die Aus- und Fortbildung*. Münster: Waxmann. <https://doi.org/10.31244/9783830991991>.

- Wilmers, Annika, Angelika Gundermann, Katharina Hähn, Gabriele Irle, Jan Koschorreck, Iris Nieding, Monique Ratermann-Busse, Bettina Waffner, Carolin Anda, und Carolin Keller. 2023. «Kompetenzen des pädagogischen Personals in der digitalen Welt. Eine vergleichende Übersicht zum Forschungsstand in verschiedenen Sektoren des Bildungssystems». In *Bildung für eine digitale Zukunft*, herausgegeben von Katharina Scheiter, und Ingrid Gogolin. Edition ZfE 15. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-37895-0_12.
- Wilmers, Annika, Carolin Keller, und Carolin Anda. 2021. «Reviews zur Organisationsentwicklung in Bildungseinrichtungen. Zur Methodik und Diskussion von Bias in Reviewverfahren der Bildungsforschung». In *Bildung im digitalen Wandel. Organisationsentwicklung in Bildungseinrichtungen.*, herausgegeben von Annika Wilmers, Michaela Achenbach, und Carolin Keller, 7–32. Münster: Waxmann.
- Wright, Alex, und Snejjina Michailova. 2022. «Critical literature reviews: A critique and actionable advice». *Management Learning* 54 (2): 177-97. <https://doi.org/10.1177/13505076211073961>.
- Zawacki-Richter, Olaf, Michael Kerres, Svenja Bedenlier, Melissa Bond, und Katja Buntins, Hrsg. 2020. *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-27602-7>.

Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

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

Cognitive Biases in Screening Processes – Search Strategies in Educational Technology Research

A Systematic Review on Learning with Virtual Reality

Katja Buntins¹ , Miriam Mulders¹ , and Nadine Schröder² 

¹ Universität Duisburg-Essen

² Hochschule Bielefeld

Abstract

The aim of this article is to empirically investigate the advantages and disadvantages of different search strategies in synthesizing research papers that use Virtual Reality (VR) educational technology. The aim is to identify cognitive biases on the part of the reviewers through different concrete searches. Using two search strategies, the study identifies the extremes between an AND search that finds as few irrelevant studies as possible but overlooks relevant ones, and an OR search that searches as broadly as possible but picks up many irrelevant studies. The article aims to show how systematic searches in educational research should be designed to adequately address the typical challenges of systematic analyses (e.g., recall-precision problem, cognitive load). The search strategies were developed based on a Google Scholar search for existing systematic reviews on VR. Here, the two search strategies differed only in terms of their linkage between a technological (VR) and a pedagogical search term. The two elements were linked with either an AND or an OR. The search items were screened in a two-person cross design and evaluated on different measures of precision and recall. There was no evidence that the more comprehensive search (OR) is superior to the narrower search (AND), but slight evidence of cognitive biases in the screening or search process in the more comprehensive search (OR). These results should be further evaluated, investigated, and, above all, replicated in further studies.



Kognitive Prozesse in Screening-Prozessen – Suchstrategien in der Bildungstechnologieforschung. Systematische Übersichtsarbeiten zum Lernen mit virtueller Realität

Zusammenfassung

Ziel dieser Studie ist es, die Vor- und Nachteile verschiedener Suchstrategien bei der Synthese von Forschungsarbeiten, die die Bildungstechnologie der virtuellen Realität (VR) nutzen, empirisch zu untersuchen. Hierbei ist das Ziel, kognitive Verzerrungen seitens der Reviewer:innen durch verschieden konkrete Suchen zu identifizieren. Mittels zweier Suchstrategien sollen die Extrema zwischen einer Suche (AND), die möglichst wenig irrelevante Studien findet, aber dafür auch relevante übersieht und einer Suche (OR), die möglichst breit sucht, aber hierbei viele irrelevante aufnimmt, dargestellt werden. Die Studie will aufzeigen, wie systematische Suchen in der Bildungsforschung gestaltet sein sollten, um die typischen Herausforderungen systematischer Analysen (z.B. Recall-Precision-Problem, kognitive Belastung) adäquat zu berücksichtigen. Die Suchstrategien wurden auf der Grundlage einer vorangegangenen Google Scholar-Suche nach bereits durchgeführten systematischen Übersichten zur VR entwickelt. Hierbei unterschieden sich die zwei verschiedenen Suchstrategien nur in Bezug auf ihre Verknüpfung zwischen einem technologischem (VR) und einem pädagogischen Suchterm. Die beiden Elemente wurden entweder mit einer AND oder einer OR Verbindung verknüpft. Die Suchbeiträge wurden in einem Kreuzdesign von zwei Personen gescreent und in Bezug auf verschiedene Präzisions- und Recallmaße evaluiert. Es fanden sich keine Hinweise dafür, dass die umfangreichere Suche (OR) der engeren Suche (AND) überlegen ist und jedoch leichte Hinweise auf kognitive Verzerrungen im Screening bzw. Suchprozess bei der umfangreicheren Suche (OR). Diese sollten in weiteren Studien weiter evaluiert, untersucht und vor allem repliziert werden.

1. Introduction

Research syntheses aim to comprehensively aggregate results on a specific research question (Gough, Oliver, and Thomas 2017) and then, depending on the type of research review, to describe, summarize, and quantify findings or derive new models or theories from the current state of research (Grant and Booth 2009). When creating such syntheses, researchers need an appropriate and efficient search strategy to find the most relevant research results on the question at hand. Ideally, a search strategy should find all relevant research and exclude all irrelevant research (Sampson and McGowan 2006). However, researchers often fail to plan and think through their search strategies sufficiently, to orient themselves toward previously successful syntheses of research, to involve information scientists, or even to test their search strategies in advance. Accordingly, many research syntheses are

carried out with inadequate scientific support under high time pressure to produce and publish results. Researchers rarely consider that such decisions affect their research and that different search strategies lead to different research results. What is more, they often do not consider the potentially far-reaching implications of different search strategies. The output, that is, the number of research articles, may vary greatly depending on the search strategy. This has further implications for the time needed to screen the papers, the cognitive load for screeners, and the likelihood of overlooking relevant criteria when there are too many criteria. There are few evidence-based contributions that examine how different search strategies affect further activities in syntheses of research (Geersing et al. 2012; Rogers, Bethel, and Boddy 2017).

Despite the lack of attention to how search strategies may affect research syntheses, researchers in many disciplines are quick to summarize the existing body of research on their topic (Andrews and Farris 1972; Lasser et al. 2020). The time pressure researchers are under can affect the psychological processes of those synthesizing research, impairing their ability to concentrate and increasing their cognitive load. Using the educational technology of Virtual Reality (VR) as an example, we aim to investigate the advantages and disadvantages of two different search strategies in the context of a specific synthesis of research.

In fast-growing research fields with a large body of literature (Larsen and Ins 2010), research syntheses play an important role (Eden 2002). This is particularly true of fields like educational technology that are developing quickly and must be translated quickly into practice because of their high practical relevance (Radianti et al. 2020; Wu, Yu, and Gu 2020). It is important that existing knowledge in the field, best practice examples, effects, mediators, and moderators are made available in a readable and condensed form, which is the aim of meta-analyses and other types of systematic reviews (Liu et al. 2017; Makransky and Petersen 2021).

This article is divided into the following parts. Part 2 presents the theoretical background. Part 3 discusses challenges of developing adequate search strategies in research syntheses. The focus here is on research about learning with VR, a topic in the field of educational technology that is currently attracting substantial research interest. Part 4 outlines the research questions and methodological approach, including the sample, procedure, and data analysis. Part 5 discusses the results, and Part 6 concludes by exploring the implications and limitations of the work.

2. Theoretical background

Depending on the type of research synthesis, researchers proceed systematically through a series of steps (Grant and Booth 2009). These include:

1. developing a research question,
2. defining inclusion and exclusion criteria,
3. searching,
4. screening,
5. quality checking,
6. coding,
7. extracting the results,
8. and synthesizing.

The methodological issues in research syntheses can be simplified by considering the relevant studies as a population, like the populations that are the subject of primary research. By looking at them in this way, approaches from primary research can be applied to secondary research. The population of the searched studies is defined by the research question, the operationalization of the population in the search (search query), and the specification of the inclusion and exclusion criteria. The population here refers to all studies about which the researcher wants to make statements, regardless of whether they are found. It can also be considered a true population. The literature search is a very important part of any research synthesis. In addition to deciding on the search terms, the literature search also includes the selection of databases and other search strategies, such as snowball searching or searching for all literature published on a topic to date (Gehanno, Rollin, and Darmoni 2013; Gusenbauer and Haddaway 2020; Shaffril et al. 2021).

3. Challenges of literature search in the context of research syntheses

3.1 *The recall-precision problem*

The proportion of studies found by a search is called **coverage** (Martín-Martín et al. 2021). If the found part of the population of studies does not fit the searched population, this is called a **coverage error** (Alvarez and VanBeselaere 2005). A coverage error is therefore a non-sampling error. Two different types of coverage errors may occur (see Figure 1). On the one hand, the search results may contain studies that are not relevant according to the criteria used in the synthesis: This is called **over-coverage**. On the other hand, the results may contain studies that are not covered by the search strategy: This is called **under-coverage**.

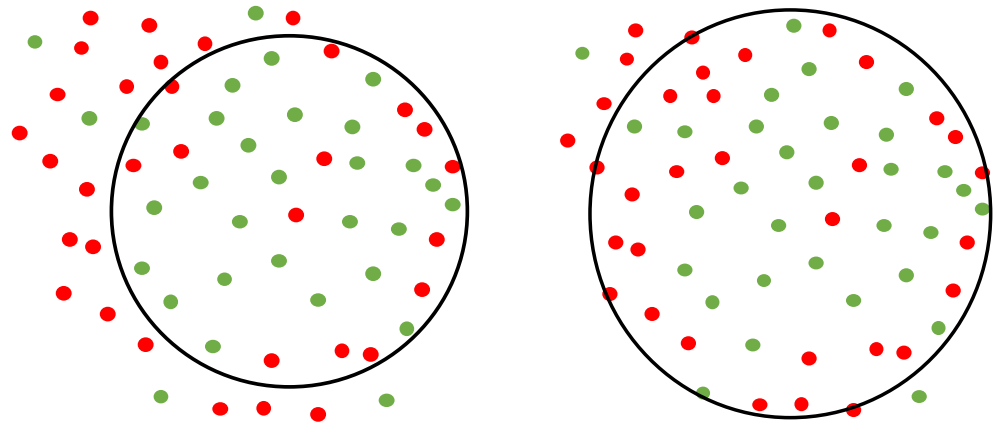


Fig. 1: Under- and over-coverage biases (red dots = not part of the population of relevant studies; green dots = part of the population of relevant studies, circle = found results).

In most literature syntheses, both forms of biases occur. This is described as the **recall-precision problem** (Bramer, Giustini, and Kramer 2016). Although there are a number of different formal definitions of recall, recall is defined here as the studies included in the search (n_{inc}) relative to all available and relevant studies (N) (Eysenbach, Tuische, and Diepgen 2001). Formally, then:

$$Recall = \frac{n_{inc}}{N}$$

This means that the closer the **recall** value is to one, the greater the proportion of relevant literature found in the search and the lower the under-coverage bias. However, if one tries to calculate the recall empirically, one faces the problem that the true population cannot be determined. Otherwise there would be no research on the advantages and disadvantages of different search strategies (Evans 2002). For this reason, recall must be estimated. Various comparison variables can be used for this purpose, such as the results of a database search. This produces a relative estimator for the quality of a subpopulation search (Straube et al. 2021). If one assumes that the included studies contain only correctly screened studies (i.e., positive examples), then the recall can also be described as sensitivity (Cooper and Varley-Campbell et al. 2018). **Precision**, on the other hand, describes the proportion of included studies in the review out of all studies (n_{inc}) that were found in the search (s_{all}). Formally, then:

$$Precision = \frac{n_{inc}}{s_{all}}$$

Again, of course, it should be noted that there are various formal definitions of precision. However, all these definitions agree that the more accurately the search string captures the true population, the closer the precision value is to one. The lower the precision value, the greater the over-coverage bias and the greater the number of irrelevant studies. A way to deal with the over-coverage bias is to screen the abstracts based on inclusion and exclusion criteria (Lee et al. 2012; Sampson, Tetzlaff, and Urquhart 2011). The aim of every research synthesis is to achieve a perfect balance between sensitivity and precision.

The recall-precision problem occurs when one factor is increased and the other simultaneously decreases (Stock and Stock 2013). Thus, if the recall is to be increased to obtain a more complete set of relevant documents, the precision decreases, so that more irrelevant documents are found. Conversely, if precision is to be increased to produce less work, this will decrease the number of relevant documents. To address the trade-off between work load (Bedenlier et al. 2020) and completeness, it is important to consider the details of the search strategy. These include the match between the research question and the search strategy as well as the quality of the search strategy.

There are several empirical research approaches and studies in medicine and other natural sciences that have dealt with the question of what changes in search strategy have what effect on various result parameters (e.g., number of hits, recall, precision). For example, there are a large number of studies dealing with the recall and precision of various databases. Bramer, Giustini and Kramer (2016) looked at the quality of Google Scholar compared to other databases in the context of liver cancer, referring to previous studies that had asked the same question before in other biomedical and medical contexts (Bramer et al. 2013; Gehanno, Rollin, and Darmoni 2013). Other studies looked at whether a non-standardized data-based search leads to more results than a standardized database search. Cooper and Lovell et al. (2018) found in the context of a study with very few matching studies that a non-standardized search leads to more matching hits. Furlan, Irvin and Bombardier (2006) asked to what extent a standardized database search can help with very specialized questions or lead to not finding relevant literature. They found that a standardized search was possible for their research objective, which was to identify nonrandomized studies. These findings indicate that the search strategy has an important influence on recall and precision. For this reason, search strategy standards are being developed. For example, Cooke, Smith and Booth (2012) explored whether the **PICO** (Population, Intervention, Comparison, Outcome) or **SPIDER** (Sample, Phenomenon of Interest, Design, Evaluation, Research Type) frameworks lead to more efficient exploitation of data. These frameworks help to conduct a search that is as sensitive as possible but still prescient (Methley et al. 2014). An addition to the PICO framework was “study type” or “study design”, so the acronym became

PICOS (Akers, Aguiar-Ibáñez, and Baba-Akbari 2009). In some studies, the S in PICOS stands for setting (Robinson, Saldanha, and Mckoy 2011). Studies evaluating these approaches have come to different conclusions. This inconsistency in the findings suggests that the choice of the right search framework depends on the research topic (Cooke, Smith, and Booth 2012; Methley et al. 2014; Rehman 2021).

The quality of systematic reviews may be affected at various points in the screening process. The search is an important step that affects all of the steps that follow. In this study, we use the example of VR to address the question of the extent to which the (im)precision of the search affects cognitive processes. The following section deals with researchers' cognitive processing when screening articles.

3.2 Cognitive load during screening processes

In the literature cited above, it is often assumed that a complex and nearly complete literature search is usually not possible, primarily due to the time involved, which would significantly limit the timeliness of the study. In addition, the retrieval of larger quantities of literature through more imprecise search strings leads to more literature that must be screened. However, Sampson, Tetzlaff and Urquhart (2011) showed that researcher errors already occur in the development of search strings. Shepperd, Bowes and Hall (2014) examined the extent of researcher bias and concluded that researchers are by far the greatest source of error in literature reviews. Biases are associated with several concepts, including prior knowledge, statistical and data processing skills, interests, and opportunities. Additionally, Süß and Schmiedek (2000) showed that in psychometric studies of cognitive performance that last several hours, fatigue and loss of motivation lead to a drop in performance, whereas effects of practice only help to improve the results slightly. König, Buhner and Murling (2005) revealed that working memory and cognitive processing were the most important predictors of multitasking performance in addition to attention and fluid intelligence. Yet there appears to be little research on cognitive resources of researchers during the review processes.

A simple, common theory for describing the limited cognitive processing capacity is that of cognitive load (Sweller 1988; Sweller, van Merriënboer, and Paas 2019). This empirically validated theory assumes that working memory is limited in its resources. When these limited resources are exceeded, no further input can be processed. Originally developed for teaching and learning settings (Buchner, Buntins, and Kerres 2021), it now has a variety of applications (Engström et al. 2017; Zhang et al. 2017). In cognitive load theory, three types of cognitive load are distinguished (Schnotz and Kürschner 2007; Sweller 2003): intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. Intrinsic cognitive load means that the cognitive load increases with complex and novel tasks. Extraneous cognitive load

refers to the form of the material, with material that contains a large amount of irrelevant information increasing cognitive load. Germane cognitive load refers to the load of the learning itself.

If this is applied to the screening of articles for research syntheses, the following assumptions can be derived:

1. The better one knows the research question and the inclusion and exclusion criteria, the lower the cognitive load. In other words, the more prior knowledge about the research topic is available and the better the criteria are remembered, the less the extraneous load and the better the performance in screening and in including and excluding articles correctly.
2. The more irrelevant articles there are in a search query, the higher the cognitive load. This is because as the number of articles increases, processing time increases, and fatigue and errors are more likely. The more irrelevant articles there are, or the more articles are generally included in a search query, the higher the cognitive load. This is because as the number of articles increases, processing time increases, and fatigue and errors are more likely.
3. The less specific the abstracts are, the higher the cognitive load. This is because if the concepts mentioned in the abstract remain unclear and are not sufficiently described, this often requires the reviewer to read the abstract more than once, to combine information, or even to look at the full text. Overall, the processing time per article and the cognitive load increases.
4. The vaguer the constructs in the research question are, the higher the cognitive load. This is because the less specifically the constructs were defined in advance, the more uncertain the reviewer becomes about when to include or exclude articles. The missing information creates uncertainty and increases the cognitive load. In addition, nonspecific constructs in the research question increase the likelihood that different reviewers will have different understandings of the concepts and potentially decrease interrater reliability. These assumptions are probably not exhaustive and, more importantly, they are not empirically tested. Rather, they are meant to indicate that it is important to pay more attention to cognitive resources and the resulting researcher biases.

3.3 Search strategies in the field of Virtual Reality

Research syntheses in the medical field have been studied extensively with regard to the choice of the search string, and there are large numbers of studies dealing with how a search string should be constructed (Salvador-Oliván, Marco-Cuenca, and Arquero-Avilés 2019). However, little research of this kind has been done in the

field of educational technology (Bedenlier et al. 2020). The aim of this article is to use the example of VR to look at which search string best considers the needed time and cognitive resources.

With educational technologies such as VR that allow a new form of access to learning content, it is important to summarize existing study findings in a readable and condensed form. Two examples are the syntheses of Radianti et al. (2020) and Wu, Yu and Gu (2020), which dealt with the effectiveness of VR in various educational settings and forms of technology and tried to draw conclusions regarding VR in educational scenarios in general. Systematic analyses that address more specific questions concerning VR are difficult to find. With the rapid advances in VR technology, more and more studies are being published, but most focus on the technology itself rather than on instructional parameters (Allcoat and Mühlénen 2018; Mulders, Buchner, and Kerres 2020). There is a lack of meta-studies that aggregate relevant primary studies to answer questions for research and practice.

In the field of educational technology, learning in VR is gaining importance. As a result of increasingly cost-effective software and hardware, more open-source programs, and educational institutions' improved technical equipment, teachers and facilities managers face the question of how they can use technologies such as VR in teaching and learning scenarios in a meaningful and value-generating way. Moreover, there are various VR visualization technologies. For example (1) ears-wrapped VR Head Mounted Displays (HMD) that cover the eyes or (2) VR with three-dimensional content displayed on two-dimensional smartphone or tablet screens. Therefore, not only general research syntheses are needed, but also more precise syntheses that are targeted to specific VR technologies and groups of learners. These kinds of studies are rare compared to studies in which the use of VR is compared with analogue formats (Buchner 2022).

Methodologically, however, this article aims to investigate and highlight the differences between two search strategies. It deals with the challenge of choosing a search strategy that is specific enough to find as many relevant studies as possible, but that does not return too many irrelevant hits. This means identifying all articles that investigate and compare the use of HMD- and desktop-based VR.

This article explores the advantages and disadvantages of two different search strategies:

1. a search strategy that is as specific as possible with the result of a manageable number of studies, but with the risk of some relevant studies not being found (AND search)
2. a wider search strategy that results in many studies being found, with some of them being irrelevant (OR search).

The two search strings differ only in how the strings were connected. Whereas in the AND search, all search strings of (1) HMD- and (2) desktop-based VR were combined with an AND, in the OR search, these strings were connected with an OR. As a result, the AND search finds only those studies that name both VR technologies in the abstract and/or in the title or keywords. The OR search finds all studies that name at least one of the two technologies. It is worth noting that the aim is not to compare the quality of different search terms but to see what dangers arise from the broader (OR) and narrower (AND) search strings.

4. Methods

4.1 Research questions

The main research question investigated in this article is: “What is the impact of two different search strategies on the screening process in a research review?”. Here, the aim is to place a special focus on the cognitive-psychological factors in screening. As described above, there is little research on fatigue processes during the screening process in reviews. By choosing an AND or an OR connection, we can address the following research questions:

1. How much larger is the data set in the OR search?
2. How many studies are not found in the AND search?
3. Is there evidence of cognitive biases in the two searches?

The two search designs were examined in the context of the content question, “Does desktop-based VR differ from HMD-based VR in terms of student learning and outcomes?”.

The search strategies were developed based on a prior Google Scholar search for previously conducted systematic reviews using the search terms “virtual reality”, “systematic review”, “education”, “technology”, “learning”, which identified the following meta-analyses, on which our own search strategies were based: Kavanagh et al. (2017), Radianti et al. (2020), Jensen and Konradsen (2018), Merchant et al. (2014).

From these search strings and the related research, as well as our prior knowledge regarding VR in educational settings, two search strings were created that included the technologies (HMD vs. desktop) to be compared (see Table 1). These were combined with two different educational context search strings, one related to the institutional context and one to the person. In addition, after screening some of the hits, we excluded some words using a NOT term. The search string was subsequently reflected upon, revised, and improved in dialogue with information scientists.

	Topic	Search String
Virtual reality	HMD-based VR	“head mounted display*” OR “head-mounted-display*” OR “HMD” OR “Oculus” OR “Samsung Gear” OR “Samsung Odyssey” OR “Google Cardboard” OR “Pimax” OR “Playstation VR” OR “Google Daydream” OR “HTC” OR “Pico” OR “Vive” OR “HP Reverb” OR “Valve Index” OR “Lenovo Mirage Solo” OR “immersive VR” OR “I-VR” OR “IVR” OR “immersive virtual reality*” OR “degrees of freedom VR” OR “degrees of freedom virtual reality*” OR “Kokoda VR” OR “immersive”
	AND/OR	
	Desktop-based VR	“desktop virtual reality*” OR “desktop VR” OR “desktop-VR” OR “desktop-3D” OR “DVR” OR “D-VR” OR “desktop 3D” OR “360 degrees video*” OR “360 degrees-video*” OR “360°-video*” OR “360° video*” OR “mobile VR” OR “laptop VR” OR “laptop 3D” OR “non-immersive” OR “low immersive” OR “less immersive”
	AND	
Education context	Institution	“school*” OR “gymnasium*” OR “youth organization*” OR “youth organisation*” OR “youth center*” OR “youth centre*” OR “leisure facility*” OR “higher education” OR “kindergarten” OR “university*” OR “colleg*” OR “apprenticeship*” OR “education*” OR “training*” OR “academic institute*” OR “academic context*” OR “learning institute*” OR “learning context*” OR “K-12*” OR “K12” OR “P-12” OR “P12” OR “museum*” OR “gallery*” OR “librar*” OR “academ*” OR “tutor*” OR “class” OR “classes” OR “learning center*” OR “learning centre*”
	OR	
	Person	learner* OR student* OR trainee* OR graduate* OR teacher*
	NOT	
	Exclusion	“dynamic voltage restorer” OR “hard-to-cook” OR “hydrothermal carbonization” OR “heat transfer coefficients” OR “maximal inspiratory pressure” OR “high-temperature combustion” OR “Pico hydro-power” OR “systematic review” OR “meta analysis” OR “heat transfer coefficient” OR “HIV testing and counseling” OR “desktop 3D printer” OR “integrated voltage regulators”

Tab. 1: Search strings.

The searches were then carried out on 16 October 2022 in *Scopus*, *Web of Science*, *PubMed*, and *ERIC*. These databases were selected because, according to studies on subject coverage, they offered broad coverage of this research topic (Gusenbauer and Haddaway 2020; Köstler 2023). Articles and conference proceedings were included in the search.

To answer the research question of which search strategy is more efficient, two searches were carried out. As presented in Table 1, we varied whether the search string had to contain (1) at least one HMD-based VR search term AND one desktop-based VR search term each or (2) at least one of the HMD-based VR search terms OR one of the desktop-based VR search terms. The two searches were cleaned of duplicates in their respective search queries. In theory, all search results from the AND search should also be found with the OR search.

4.2 Search results and research processes

As described above, two data sets were deliberately created that differ in the number of papers found. Consequently, the AND search results in 251 studies, while the OR Search results in 6403 studies. To ensure that the results are comparable, the analysis was carried out with 251 articles. For the AND linkage, all 251 were screened. The size of the sample led to a confidence interval of 95% and a margin of error of 6% (Kupper and Hafner 1989).

To evaluate both searches in terms of coverage, effort, and cognitive bias processes, both authors screened both searches. One author started with the OR link search and the other with the AND link search. Afterwards, they switched. This was done to distribute sequence effects and cognitive fatigue evenly between the two searches.

4.3 Inclusion and exclusion criteria

Inclusion and exclusion criteria were defined for the description of the population of studies, that is, the number of studies about which one wants to make statements (see Table 2). These include, above all, that there is a comparison of desktop-based and HMD-based VR. However, there are also other exclusion criteria, such as no educational context or no original data being collected. The exclusion criteria were applied hierarchically. This means that “1. no English language” was assigned if the study was in a language other than English, regardless of what other criteria would fit the study. Similarly, the exclusion criterion “6. secondary research” was only assigned if none of the previous criteria fit. It is therefore enough if one exclusion criterion is fulfilled for an article to be excluded. But all inclusion criteria must be fulfilled for the articles to be included.

Inclusion criteria	Exclusion criteria (in hierarchical order)
<ul style="list-style-type: none"> • English language • published between 2019 and 2022 • original empirical data • articles & conference proceedings • comparison of desktop-based and HMD-based VR • educational context • primary search 	<ol style="list-style-type: none"> 1. no English language 2. published before 2019 3. no original empirical data 4. other publication type than articles and conference proceedings 5. no comparison of desktop-based and HMD-based VR 6. no educational context 7. secondary research

Tab. 2: Inclusion and exclusion criteria.

4.4 Evaluation criteria

Once again, as a reminder, the study was conducted on 251 articles. The articles in the OR link search were randomly selected. These 2.251 studies were analyzed using different evaluation criteria. These are shown in Table 3.

There were three different goals here. The evaluation criteria “total number of studies” and “precision in the search” are used to determine the effort. The evaluation criteria “estimation of recall” and “studies not found” were used to determine coverage. The two inter-rater reliabilities (IRR) were used to represent biases due to cognitive processes. These differ, as described in the table, in that the inter-rater reliability (criteria compliance) is about the hierarchy of the exclusion criteria. In other words, here we check whether the same reason for exclusion was mentioned—which is an indicator of accuracy.

Evaluation criteria	Description	Formalization	Interpretation
Total number of studies	Total number of studies found in each search	$n_{fA \vee O}$	The more studies there are, the more work is required.
Estimation of recall	Proportion of included articles after screening from one search in all included articles after screening from both (To avoid biases, the included number of articles in the OR-link sample is estimated. The estimation takes place by scaling up.)	$R_{A \vee O} = \frac{n_{incA \vee O_E}}{(n_{incA} + n_{incE})}$	The higher the recall value, the more studies in total are found by the search strategy.
Precision in the search	Proportion of included articles after screening in the total number of studies	$P_{A \vee O} = \frac{n_{incA \vee O}}{n_{fA \vee O_S}}$	The higher the precision value, the more precisely the search string describes what is being searched for.
Inter-rater reliability (IRR)	Proportion of matches regarding inclusion and exclusion decisions after screening on the search	$IRR_{IE} = \frac{M_{inc \wedge excA \vee O_S}}{n_{fA \vee O_S}}$	The higher the IRR is, the more accurate reviewer's work.
Inter-rater reliability (criteria compliance)	Proportion of the search query matched with respect to the inclusion and exclusion criteria	$IRR_{IE} = \frac{M_{criA \vee O_S}}{n_{fA \vee O_S}}$	The higher the IRR is, the more accurate reviewer's work.
Studies not found	Number of missing studies in the AND-Link search found in the OR-link search		The higher the value, the more studies are overlooked in the narrower search string.
Included studies (unique)	Number of studies found in the one search only.	$n_A \neq \{O\} \vee n_O \neq \{A\}$	Study was only found with one search. The OR search refers to the entire sample of 6543 studies.

Legend:

<i>n</i> : number of studies	<i>inc</i> : included studies	<i>M</i> : matching	<i>A</i> : AND-link search	<i>E</i> : estimated
<i>f</i> : found studies	<i>exc</i> : excluded studies	<i>cri</i> : criterion	<i>O</i> : OR-link search	<i>s</i> : sampled studies

Tab. 3: Evaluation criteria.

5. Results

The aim of this study was to determine which of the search strategies (AND vs. OR) is best suited to generate a search query that contains as many relevant studies as possible, excludes as many irrelevant studies as possible, and overlooks as few relevant studies as possible in the screening process. Altogether, using the example of the educational technology VR, we have tried to find out what suitable search strategies include that adequately address typical challenges of systematic analyses (e.g., recall-precision problem, cognitive load). For this purpose, two search strategies were evaluated according to the different evaluation criteria described above. The results are shown in Table 4. Here it can be seen that the OR search leads to 6152 more studies than the AND search. It took each of the two scientific researchers about 2.5 hours to screen the 251 articles. The articles in the OR search were randomly sampled. One of us started with the 251 articles in the AND-link Search. The other one started with the OR search. Taken together, the additional effort screening all articles found in the OR search can therefore be estimated at about 60 hours per person (2.5 per 251 articles). Here, the percentage given in precision is significantly higher in the AND search. The recall estimates are about the same but are somewhat lower in the AND search. Looking at the interrater reliability, there are no relevant differences in either measure. There is no difference when looking at IRR inclusion and exclusion decisions. In the interrater reliability (criteria compliance) (see Table 2), there is slightly higher reliability for the OR search.

Evaluation criteria	AND search	OR search
Total number of studies	251	6403
Included numbers	49 of 251	2 of 251
Estimated numbers of included studies	-	51 (2*(6403/251))
Included studies	2 of 251	1 of 251
Recall estimation in the search	0.92	0.96
Precision in the search	0.195	0.008
Inter-rater reliability	0.964	0.964
Inter-rater reliability (criteria compliance)	0.773	0.800

Tab. 4: Results of the AND and OR search.

6. Discussion

The aim of this study was to empirically investigate the advantages and disadvantages of different search strategies using the example of a systematic literature search in the context of VR. The peculiarities of the field are, on the one hand, the fast-growing pace of the research field and, on the other, the multitude of studies comparing VR with other technologies or analogue formats instead of two VR realizations.

To this end, a broader and a narrower search were conducted and examined in terms of various recall, precision, and accuracy parameters. In the searches, there were no relevant differences in the accuracy of the screening (described using the two IRR evaluation criteria and none regarding a relevant low recall). On the contrary, more unique studies were found in the narrower (AND) search: studies that did not appear in the data set of the broader (OR) search. Since this is technically impossible, we suspect that human errors in removing duplicates occurred during this search. These were identified using the Eppi-Reviewer application and then automated for more than 98%, the rest manually. Subsequently, duplicates were searched for manually. This can lead to both technical and human errors. Another source of human error is the download limitation in *Scopus* and *ERIC*. In these databases, only a certain number of articles can be downloaded at the same time, so several files must be downloaded due to further filtering algorithms. These must be determined by means of logical closure.

There are very clear differences in precision – for example, there are 25 times more hits than in the broad OR search, and the precision value for the broader search is 0.008 compared to 0.195 for the narrower search (AND search). Hence, the workload is much higher for the broader search.

It should be noted that the inclusion and exclusion criteria in this search are manifest variables, meaning that they are directly observable. If we take a closer look at the missing matches, the cases where the same exclusion criteria were not selected, we see that they are all due to a lack of definition of the educational context. On the one hand, this could be due to a lack of concentration. On the other hand, however, there is a blurred distinction between psychoeducation and learning within the framework of rehabilitation measures. For example, the authors did not agree on whether a study on driving under the influence of alcohol among college students is education or still psychoeducation (Madigan and Romano 2020). To understand the processes here precisely, qualitative studies would be needed that apply the think aloud method, for example.

In the current search, we could not find any evidence of a greater cognitive load in the broader search (OR search). This means that the IRR is not lower here and is even higher in the exact code coding (criteria compliance). This can be interpreted as an indication that the cognitive load does not differ significantly. However, we believe that more studies are needed in this area. These vary further in terms of the fuzziness of the criteria, the experience of the reviewers, and the quantity of hits. It should also be noted that there is still a great deal of variance in the sample of 251 studies. In a further step, it would be interesting to see whether the values from the OR search can be confirmed in this way in other samples.

7. Conclusion

Our analysis showed that a narrow search string (AND search) does not lead to less relevant results, such as significantly fewer studies, than the broader search string (OR search). We will therefore continue to work with this in our upcoming meta-analysis. By doing so, we hope to make our work more efficient by systematically reducing the number of irrelevant articles.

Our finding is in line with Cooke's (2012) findings. However, it should be noted that the search string must indeed be as precise as possible. As mentioned above, cooperation with librarians and information scientists is crucial here. Studies have shown that they are not consulted in most review processes, although they would be willing to help (Grossetta Nardini et al. 2019). They can, for instance, help to refine a search string and point out missing elements in a search string.

Furthermore, an additional manual search should be attempted if necessary. On the one hand, this offers the possibility of finding literature outside of large databases and also grey literature or pre-print documents (Haddaway et al. 2015; Haddaway and Bayliss 2015).

7.1 Further research

In the field of VR, where there are already many studies with more being added all the time, it is crucial to design research syntheses efficiently. In this field, to systematically aggregate current technological trends, a systematic review is a state-of-the-art method providing profound results. Nevertheless, more research is needed on the question of how researchers can find relevant literature, also on other topics and research questions dealing with cognitive load and other cognitive variables.

A special focus should be placed on questions with fuzzy constructs, such as motivation, performance, or student engagement (Bond et al. 2020). There is hardly any research in this area to date. In the future, the influence of latent and thus fuzzy constructs on fatigue during screening and coding should be investigated.

Another question that has not been addressed adequately in the literature is how long articles can be screened before concentration decreases and fatigue appears, and to what extent this depends on the research question and the quality of the search string.

References

- Akers, Jo, R. Aguiar-Ibáñez, and A. Baba-Akbari. 2009. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. York, UK: Centre for Reviews and Dissemination, University of York.
- Allcoat, Devon, and Adrian von Mühlennen. 2018. "Learning in Virtual Reality: Effects on Performance, Emotion and Engagement". *Research in Learning Technology* 26. <https://doi.org/10.25304/rlt.v26.2140>.
- Alvarez, R. Michael, and Carla VanBeselaere. 2005. "Web-Based Survey". In *Encyclopedia of Social Measurement*. Edited by Kimberly Kempf-Leonard, 955–62. New York: Elsevier. <https://doi.org/10.1016/B0-12-369398-5/00390-X>.
- Andrews, Frank M., and George F. Farris. 1972. "Time Pressure and Performance of Scientists and Engineers: A Five-Year Panel Study". *Organizational Behavior and Human Performance* 8 (2): 185–200. [https://doi.org/10.1016/0030-5073\(72\)90045-1](https://doi.org/10.1016/0030-5073(72)90045-1).
- Bedenlier, Svenja, Melissa Bond, Katja Buntins, Olaf Zawacki-Richter, and Michael Kerres. 2020. "Learning by Doing? Reflections on Conducting a Systematic Review in the Field of Educational Technology". In *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*, edited by Olaf Zawacki-Richter et al., 111–27. Wiesbaden: Springer Fachmedien.
- Bond, Melissa, Katja Buntins, Svenja Bedenlier, Olaf Zawacki-Richter, and Michael Kerres. 2020. "Mapping Research in Student Engagement and Educational Technology in Higher Education: A Systematic Evidence Map". *International Journal of Educational Technology in Higher Education* 17 (1): 2. <https://doi.org/10.1186/s41239-019-0176-8>.
- Bramer, Wichor M., Dean Giustini, Bianca Kramer, and P. F. Anderson. 2013. "The Comparative Recall of Google Scholar Versus PubMed in Identical Searches for Biomedical Systematic Reviews: A Review of Searches Used in Systematic Reviews". *Systematic reviews* 2 (1): 115. <https://doi.org/10.1186/2046-4053-2-115>.
- Bramer, Wichor M., Dean Giustini, and Bianca M. R. Kramer. 2016. "Comparing the Coverage, Recall, and Precision of Searches for 120 Systematic Reviews in Embase, MEDLINE, and Google Scholar: A Prospective Study". *Systematic reviews* 5 (1): 39. <https://doi.org/10.1186/s13643-016-0215-7>.
- Buchner, Josef. 2022. "Systematic Reviews Als Analyseinstrument Der Forschungspraxis in Educational Technology Studien". Presentation at DGfE-Kongress 2022. <https://doi.org/10.13140/RG.2.2.19953.35687>.
- Buchner, Josef, Katja Buntins, and Michael Kerres. 2021. "A Systematic Map of Research Characteristics in Studies on Augmented Reality and Cognitive Load: A Systematic Map of Research Characteristics". *Computers and Education Open* 2: 100036. <https://doi.org/10.1016/j.caeo.2021.100036>.
- Cooke, Alison, Debbie Smith, and Andrew Booth. 2012. "Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis". *Qualitative Health Research* 22 (10): 1435–43. <https://doi.org/10.1177/1049732312452938>.

- Cooper, Chris, Rebecca Lovell, Kerryn Husk, Andrew Booth, and Ruth Garside. 2018. "Supplementary Search Methods Were More Effective and Offered Better Value Than Bibliographic Database Searching: A Case Study from Public Health and Environmental Enhancement". *Research synthesis methods* 9 (2): 195–223. <https://doi.org/10.1002/jrsm.1286>.
- Cooper, Chris, Joanna Varley-Campbell, Andrew Booth, Nicky Britten, and Ruth Garside. 2018. "Systematic Review Identifies Six Metrics and One Method for Assessing Literature Search Effectiveness but No Consensus on Appropriate Use". *Journal of Clinical Epidemiology* 99: 53–63. <https://doi.org/10.1016/j.jclinepi.2018.02.025>.
- Eden, Dov. 2002. "From the Editors". *Academy of Management Journal* 45 (5): 841–46. <https://doi.org/10.5465/amj.2002.7718946>.
- Engström, Johan, Gustav Markkula, Trent Victor, and Natasha Merat. 2017. "Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis". *Human Factors* 59 (5): 734–64. <https://doi.org/10.1177/0018720817690639>.
- Evans, David. 2002. "Database Searches for Qualitative Research". *Journal of the Medical Library Association JMLA* 90 (3): 290–93. <https://pubmed.ncbi.nlm.nih.gov/12113512>.
- Eysenbach, G., J. Tuische, and T. L. Diepgen. 2001. "Evaluation of the Usefulness of Internet Searches to Identify Unpublished Clinical Trials for Systematic Reviews". *Medical informatics and the Internet in medicine* 26 (3): 203–18. <https://doi.org/10.1080/14639230110075459>.
- Furlan, Andrea D., Emma Irvin, and Claire Bombardier. 2006. "Limited Search Strategies Were Effective in Finding Relevant Nonrandomized Studies". *Journal of Clinical Epidemiology* 59 (12): 1303–11. <https://doi.org/10.1016/j.jclinepi.2006.03.004>.
- Geersing, Geert-Jan, Walter Bouwmeester, Peter Zuithoff, Rene Spijker, Mariska Leeflang, and Karel Moons. 2012. "Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews". *PLoS one* 7 (2): e32844. <https://doi.org/10.1371/journal.pone.0032844>.
- Gehanno, Jean-François, Laetitia Rollin, and Stefan Darmoni. 2013. "Is the Coverage of Google Scholar Enough to Be Used Alone for Systematic Reviews". *BMC Medical Informatics and Decision Making* 13 (1): 7. <https://doi.org/10.1186/1472-6947-13-7>.
- Gough, David, Sandy Oliver, and James Thomas. 2017. *An Introduction to Systematic Reviews*. 2nd revised edition. Los Angeles: SAGE.
- Grant, Maria J., and 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>.
- Grossetta Nardini, Holly K., Janene Batten, Melissa C. Funaro, Rolando Garcia-Milian, Kate Nyhan, Judy M. Spak, Lei Wang, and Janis G. Glover. 2019. "Librarians as Methodological Peer Reviewers for Systematic Reviews: Results of an Online Survey". *Research Integrity and Peer Review* 4 (1): 23. <https://doi.org/10.1186/s41073-019-0083-5>.
- Gusenbauer, Michael, and Neal R. Haddaway. 2020. "Which Academic Search Systems Are Suitable for Systematic Reviews or Meta-Analyses? Evaluating Retrieval Qualities of Google Scholar, PubMed, and 26 Other Resources". *Research synthesis methods* 11 (2): 181–217. <https://doi.org/10.1002/jrsm.1378>.

- Haddaway, Neal R., and Helen R. Bayliss. 2015. "Shades of Grey: Two Forms of Grey Literature Important for Reviews in Conservation". *Biological Conservation* 191: 827–29. <https://doi.org/10.1016/j.biocon.2015.08.018>.
- Haddaway, Neal R., Alexandra Mary Collins, Deborah Coughlin, and Stuart Kirk. 2015. "The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching". *PLoS one* 10 (9): e0138237. <https://doi.org/10.1371/journal.pone.0138237>.
- Jensen, Lasse, and Flemming Konradsen. 2018. "A Review of the Use of Virtual Reality Head-Mounted Displays in Education and Training". *Education and Information Technologies* 23 (4): 1515–29. <https://doi.org/10.1007/s10639-017-9676-0>.
- Kavanagh, Sam, Andrew Luxton-Reilly, Burkhard Wuensche, and Beryl Plimmer. 2017. "A Systematic Review of Virtual Reality in Education". *Themes in Science and Technology Education* 10 (2): 85–119. <https://www.learntechlib.org/p/182115/>.
- König, Cornelius J., Markus Buhner, and Gesine Murling. 2005. "Working Memory, Fluid Intelligence, and Attention Are Predictors of Multitasking Performance, but Polychronicity and Extraversion Are Not". *Human Performance* 18 (3): 243–66. https://doi.org/10.1207/s15327043hup1803_3.
- Köstler, Verena. 2023. "Zwischen Präzision und Sensitivität: Generierung eines Studienkorpus am Beispiel einer Fragestellung zu Künstlicher Intelligenz (KI) in Bildungsprozessen". *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 54 (Research Syntheses): 1–27. <https://doi.org/10.21240/mpaed/54/2023.08.10.X>.
- Kupper, Lawrence L., and Kerry B. Hafner. 1989. "How Appropriate Are Popular Sample Size Formulas?". *The American Statistician* 43 (2): 101. <https://doi.org/10.2307/2684511>.
- Larsen, Peder Olesen, and Markus von Ins. 2010. "The Rate of Growth in Scientific Publication and the Decline in Coverage Provided by Science Citation Index". *Scientometrics* 84 (3): 575–603. <https://doi.org/10.1007/s11192-010-0202-z>.
- Lasser, Jana, Verena Ahne, Georg Heiler, Peter Klimek, Hannah Metzler, Tobias Reisch, Martin Sprenger, Stefan Thurner, and Johannes Sorger. 2020. "Complexity, transparency and time pressure: practical insights into science communication in times of crisis". *Journal of Science Communication* 19 (5): N01. <https://doi.org/10.22323/2.19050801>.
- Lee, Edwin, Maureen Dobbins, Kara DeCorby, Lyndsey McRae, Daiva Tirilis, and Heather Husson. 2012. "An Optimal Search Filter for Retrieving Systematic Reviews and Meta-Analyses". *BMC medical research methodology* 12 (1): 51. <https://doi.org/10.1186/1471-2288-12-51>.
- Liu, Dejian, Kaushal Kumar Bhagat, Yuan Gao, Ting-Wen Chang, and Ronghuai Huang. 2017. "The Potentials and Trends of Virtual Reality in Education". In *Virtual, Augmented, and Mixed Realities in Education*, 105–30. Springer, Singapore. https://doi.org/10.1007/978-981-10-5490-7_7.
- Madigan, Ruth, and Richard Romano. 2020. "Does the Use of a Head Mounted Display Increase the Success of Risk Awareness and Perception Training (RAPT) For Drivers?" *Applied ergonomics* 85: 103076. <https://doi.org/10.1016/j.apergo.2020.103076>.

- Makransky, Guido, and Gustav B. Petersen. 2021. "The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality". *Educational Psychology Review* 33 (3): 937–58. <https://doi.org/10.1007/s10648-020-09586-2>.
- Martín-Martín, Alberto, Mike Thelwall, Enrique Orduna-Malea, and Emilio Delgado López-Cózar. 2021. "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A Multidisciplinary Comparison of Coverage via Citations". *Scientometrics* 126 (1): 871–906. <https://doi.org/10.1007/s11192-020-03690-4>.
- Merchant, Zahira, Ernest T. Goetz, Lauren Cifuentes, Wendy Keeney-Kennicutt, and Trina J. Davis. 2014. "Effectiveness of Virtual Reality-Based Instruction on Students' Learning Outcomes in K-12 and Higher Education: A Meta-Analysis". *Computers & Education* 70: 29–40. <https://doi.org/10.1016/j.compedu.2013.07.033>.
- Methley, Abigail M., Stephen Campbell, Carolyn Chew-Graham, Rosalind McNally, and Sudeh Cheraghi-Sohi. 2014. "PICO, PICOS and SPIDER: A Comparison Study of Specificity and Sensitivity in Three Search Tools for Qualitative Systematic Reviews". *BMC Health Services Research* 14 (1): 579. <https://doi.org/10.1186/s12913-014-0579-0>.
- Mulders, Miriam, Josef Buchner, and Michael Kerres. 2020. "A Framework for the Use of Immersive Virtual Reality in Learning Environments". *International Journal of Emerging Technologies in Learning (iJET)* 15 (24): 208–24. <https://doi.org/10.3991/ijet.v15i24.16615>.
- Radianti, Jaziar, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. 2020. "A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda". *Computers & Education* 147: 103778. <https://doi.org/10.1016/j.compedu.2019.103778>.
- Rehman, Yasir. 2021. "Difference Between Quantitative and Qualitative Research Question-PICO Vs. SPIDER". *American Academic Scientific Research Journal for Engineering, Technology, and Sciences* 77 (1): 188–99. https://asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/6730.
- Robinson, Karen A., Ian J. Saldanha, and Naomi A. Mckoy. 2011. "Development of a Framework to Identify Research Gaps from Systematic Reviews". *Journal of Clinical Epidemiology* 64 (12): 1325–30. <https://doi.org/10.1016/j.jclinepi.2011.06.009>.
- Rogers, Morwenna, Alison Bethel, and Kate Boddy. 2017. "Development and Testing of a Medline Search Filter for Identifying Patient and Public Involvement in Health Research". *Health Information & Libraries Journal* 34 (2): 125–33. <https://doi.org/10.1111/hir.12157>.
- Salvador-Oliván, José Antonio, Gonzalo Marco-Cuenca, and Rosario Arquero-Avilés. 2019. "Errors in Search Strategies Used in Systematic Reviews and Their Effects on Information Retrieval". *Journal of the Medical Library Association JMLA* 107 (2): 210–21. <https://doi.org/10.5195/jmla.2019.567>.
- Sampson, Margaret, and Jessie McGowan. 2006. "Errors in Search Strategies Were Identified by Type and Frequency". *Journal of Clinical Epidemiology* 59 (10): 1057.e1-1057.e9. <https://doi.org/10.1016/j.jclinepi.2006.01.007>.

- Sampson, Margaret, Jennifer Tetzlaff, and Christine Urquhart. 2011. "Precision of Healthcare Systematic Review Searches in a Cross-Sectional Sample". *Research synthesis methods* 2 (2): 119–25. <https://doi.org/10.1002/jrsm.42>.
- Schnotz, Wolfgang, and Christian Kürschner. 2007. "A Reconsideration of Cognitive Load Theory". *Educational Psychology Review* 19 (4): 469–508. <https://doi.org/10.1007/s10648-007-9053-4>.
- Shaffril, Mohamed, Hayrol Azril, Samsul Farid Samsuddin, and Asnarulkhadi Abu Samah. 2021. "The ABC of Systematic Literature Review: The Basic Methodological Guidance for Beginners". *Quality & Quantity* 55 (4): 1319–46. <https://doi.org/10.1007/s11135-020-01059-6>.
- Shepperd, Martin, David Bowes, and Tracy Hall. 2014. "Researcher Bias: The Use of Machine Learning in Software Defect Prediction". *IEEE Transactions on Software Engineering* 40 (6): 603–16. <https://doi.org/10.1109/TSE.2014.2322358>.
- Stock, Wolfgang G., and Mechtild Stock. 2013. *Handbook of Information Science*. Berlin, Boston: De Gruyter Saur. <https://doi.org/10.1515/9783110235005>.
- Straube, S., J. Heinz, P. Landsvogt, and T. Friede. 2021. "Recall, Precision, and Coverage of Literature Searches in Systematic Reviews in Occupational Medicine: An Overview of Cochrane Reviews Recall, Precision Und Coverage Von Literatursuchen in Systematischen Reviews Aus Dem Bereich Arbeitsmedizin: Ein Überblick Über Cochrane Reviews". *GMS Medizinische Informatik, Biometrie und Epidemiologie* 17 (1). <https://doi.org/10.3205/mibe000216>.
- Süß, Heinz-Martin, and Florian Schmiedek. 2000. "Ermüdungs- Und Übungseffekte Bei Mehrstündiger Kognitiver Beanspruchung". *Experimental Psychology* 47 (3): 162–79. <https://doi.org/10.1026//0949-3964.47.3.162>.
- Sweller, John. 1988. "Cognitive Load During Problem Solving: Effects on Learning". *Cognitive Science* 12 (2): 257–85. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7).
- Sweller, John. 2003. "Evolution of Human Cognitive Architecture". *Psychology of learning and motivation* 43: 216–66.
- Sweller, John, Jeroen J. G. van Merriënboer, and Fred Paas. 2019. "Cognitive Architecture and Instructional Design: 20 Years Later". *Educational Psychology Review* 31 (2): 261–92. <https://doi.org/10.1007/s10648-019-09465-5>.
- Wu, Bian, Xiaoxue Yu, and Xiaoqing Gu. 2020. "Effectiveness of Immersive Virtual Reality Using Head-mounted Displays on Learning Performance: A Meta-analysis". *British Journal of Educational Technology* 51 (6): 1991–2005. <https://doi.org/10.1111/bjet.13023>.
- Zhang, Fan, Shamila Haddad, Bahareh Nakisa, Mohammad Naim Rastgoo, Christhina Candido, Dian Tjondronegoro, and Richard de Dear. 2017. "The Effects of Higher Temperature Setpoints During Summer on Office Workers' Cognitive Load and Thermal Comfort". *Building and Environment* 123: 176–88. <https://doi.org/10.1016/j.buildenv.2017.06.048>.

Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

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

Was meinst du eigentlich, wenn du von «Feedback» sprichst?

Chancen und Grenzen qualitativ-inhaltsanalytischer Scope-Reviews zur Herausarbeitung von Taxonomien zur Beschreibung didaktischer Szenarien am Beispiel «Feedback»

Annika Brück-Hübner¹  und Jennifer Schluer² 

¹ Justus-Liebig-Universität Giessen

² Technische Universität Chemnitz

Zusammenfassung

Ein Blick in die (medien-)pädagogische Forschungsliteratur verdeutlicht, wie vielfältig zentrale Begriffe wie z. B. der des «Feedbacks» verstanden werden. Grund hierfür sind nicht nur divergierende theoretische Annahmen und forschungsmethodische Herangehensweisen, sondern auch unterschiedliche lehr-/lernmethodische Auffassungen sowie die damit verbundene konkrete didaktische Umsetzung in der Lehre. Der Pluralität didaktischer Szenarien wurde in bisherigen Feedback-Reviews jedoch kaum Aufmerksamkeit geschenkt, obwohl didaktische Designs Auswirkungen auf den Lernerfolg und somit auch die Studienergebnisse haben können. Ursache hierfür ist eine oftmals unzureichende Beschreibung der Feedbackszenarien in den Primärstudien, welche wiederum die Herausarbeitung von Schlussfolgerungen und Effektstärken in Meta-Analysen erschwert. In diesem Beitrag wird ein qualitativ-inhaltsanalytischer Scope-Review der didaktischen Szenarien in Feedbackstudien vorgestellt, welcher das Ziel verfolgt, die Bandbreite an Implementierungen zu erfassen und eine Grundlage zur transparenten Beschreibung von Feedbackdesigns für zukünftige (Meta-)Studien zu schaffen. Hierzu wird eine Taxonomie entwickelt, die auf möglichst umfassende Weise unterschiedliche Umsetzungsvarianten abbilden soll. Daran anschließend werden im Beitrag die Chancen und Grenzen eines qualitativ-inhaltsanalytischen Scope-Reviews zur Herausarbeitung von Taxonomien zur Beschreibung didaktischer Designs diskutiert.

What Do You Actually Mean When You Talk about ‹Feedback›? Chances and Limitations of Qualitative Content-Analytical Scoping Reviews for Developing Pedagogical Taxonomies Using the Example of ‹Feedback› Implementations

Abstract

A close look at the empirical literature in the field of (media) pedagogy reveals that numerous key terms, such as the term ‹feedback›, have been utilized in many different ways. The reasons for this are not only divergent conceptual understandings and empirical approaches, but also varying teaching and learning methods, which result in a plethora of concrete didactic implementations in the classroom. However, this plurality of pedagogical scenarios has received little attention in previous feedback reviews, although variations in pedagogical design can affect the learning success and thus also the results of empirical studies. Often, the feedback scenarios have been described insufficiently in the primary studies, which in turn complicates the calculation and interpretation of effect sizes in meta-analyses. The authors therefore conducted a qualitative content-analytical scoping review of pedagogical scenarios in feedback research to capture the wide range of possible implementations. For this purpose, a taxonomy has been developed that charts different pedagogical variants in a comprehensive manner, which could serve as a foundation for the transparent description of feedback designs in future studies and meta-analyses. On that basis, the paper discusses the chances and limitations of qualitative content-analytical scoping reviews for the creation of such pedagogical taxonomies.

1. Einleitung

Die (medien-)pädagogische Forschung ist sehr vielfältig, was nicht nur aus einer Pluralität an theoretischen Annahmen und forschungsmethodischen Herangehensweisen, sondern auch aus der Diversität didaktischer Szenarien und der ihr zugrundeliegenden lehr-/lerntheoretischen Auffassungen resultiert. Ein prominentes Beispiel ist der Begriff ‹Feedback› mitsamt seinen unterschiedlichen Implementierungen in Primärstudien. Hattie (2009) stellte in seiner Meta-Analyse das Thema ‹Feedback› als grundlegendes Element guter Lehre heraus. Kurz nach ihrem Erscheinen hat diese Studie die Aufmerksamkeit der Fachwelt auf sich gezogen und sie wurde bis heute laut ResearchGate bereits 10.254mal zitiert.¹ Zugleich wurde auch Kritik an dem methodischen Vorgehen laut. Was genau ist mit ‹Feedback› gemeint? Wie sollte es in der Lehre eingesetzt werden, um lernwirksam zu sein? Es zeigte sich, dass die Einzelergebnisse der Primärstudien keineswegs konsistent sind, sondern eine grosse Varianz in den Effektstärken aufweisen. Als Ursache für die Differenzen wird u. a. angeführt,

¹ https://www.researchgate.net/publication/270585193_Visible_Learning_A_Synthesis_of_Over_800_Meta-Analyses_Relating_to_Achievement, Stand 04.04.23, 11:22 Uhr.

dass die *unterschiedlichen Feedbackverständnisse* der Einzelstudien in der Meta-Analyse keine Berücksichtigung gefunden hatten (s. z. B. Ruiz-Primo und Li 2013; Wisniewski, Zierer, und Hattie 2020). In der Tat unterscheiden sich Feedbackdesigns in vielerlei Hinsicht. So werden beispielsweise Begriffe wie «Video-Feedback» in der Literatur mit unterschiedlichen Bedeutungen verwendet, z. B. als Videografie des eigenen Unterrichtshandelns zum Zweck der Selbstreflexion und Professionalisierung (Tochon 2013) oder als Videografie der Feedbacknachricht, in welcher nur die sprechende Person und/oder eine Bildschirmaufzeichnung sichtbar ist (Mahoney, Macfarlane, und Ajjawi 2019). Oftmals bleibt die konkrete Ausgestaltung digitaler Feedbackdesigns jedoch hinter Sammelbegriffen wie «Video-Feedback» verborgen und wird in der Literatur selten explizit thematisiert oder hinterfragt (Schluer 2022). In Anbetracht dieser Begriffsheterogenität und der vielfältigen technologischen Möglichkeiten ist anzunehmen, dass auch mit umfassenden Forschungssynthesen keine aussagekräftigen Ergebnisse gewonnen werden können, sofern die unterschiedlichen didaktischen Designs in Primärstudien nicht ausreichend beschrieben und diese Differenzen somit in Meta-Analysen nicht berücksichtigt werden können.

Das Problem der begrifflichen Unschärfe beschränkt sich selbstverständlich nicht nur auf den Umgang mit dem «Feedback»-Begriff. Das auch unter der Beschreibung «Apples and Oranges» bekannte Uniformitätsphänomen betrifft vielmehr alle Sekundäranalysen und bezieht sich nicht nur auf unterschiedliche begriffliche und didaktische Konzepte, sondern umfasst unter anderem auch lehr- und forschungsmethodische Differenzen in den zusammengefassten Primärstudien (siehe hierzu z. B. Higgins 2021, 33; Eisend 2004, 22; Ritzhaupt et al. 2021, 2517). Es ist anzunehmen, dass eine grössere begriffliche und konzeptionelle Differenziertheit bei der Beschreibung der zu erforschenden pädagogischen Praxis die Nachvollziehbarkeit und Aussagekraft der Ergebnisse erhöht und zugleich differenziertere Rückschlüsse auf Korrelationen und Effektstärken ermöglicht.

Das Ziel des in diesem Beitrag vorgestellten qualitativ-inhaltsanalytischen Scope-Reviews ist nicht die statistische Analyse des Datenmaterials oder die evidenzbasierte Zusammenfassung von Primärstudien, sondern vielmehr eine möglichst umfassende qualitative Erfassung der didaktischen Designs in Feedbackstudien. Durch die qualitativ-inhaltsanalytische Kategorisierung der didaktischen Design-Elemente aus Primärstudien wurde eine Taxonomie erarbeitet, die in zukünftigen Studien als Grundlage einer transparenten und differenzierten Beschreibung von Feedbackdesigns dienen kann. In diesem Beitrag gilt es, basierend auf der exemplarischen Durchführung anhand der Feedbackdesigns in Primärstudien die Chancen und Grenzen qualitativ-inhaltsanalytischer Scope-Reviews zur Herausarbeitung von Taxonomien zu diskutieren.

2. Taxonomien als Hilfsmittel für begriffliche und konzeptionelle Präzision

In diesem Kapitel wird zunächst erörtert, was Taxonomien sind und welches Potenzial sie im Hinblick auf die Präzisierung und Systematisierung von Begriffen und Konzepten bergen.

2.1 Was sind Taxonomien?

Taxonomien sind Ordnungsrahmen, die ermöglichen, die vielfältigen Erscheinungsbilder eines Phänomens zu klassifizieren und nach einem einheitlichen Schema zu beschreiben und zu erklären (Baumgartner, Himpsl, und Kleindienst 2012, 13). In Anlehnung an Baumgartner, Himpsl und Kleindienst (2012) sowie Baumgartner (2006) können Taxonomien insgesamt acht Funktionen erfüllen:

1. *Integrationsfunktion* (Zusammenfassung/Klassifikation verschiedener Ausprägungen nach gemeinsamen Merkmalen),
2. *Orientierungsfunktion* (Überblick über die vorhandenen Möglichkeiten),
3. *Informationsfunktion* (Verbesserung der Kommunikation und Reduktion von Missverständnissen durch präzise begriffliche Abgrenzung),
4. *Kostensenkungsfunktion* (die Übertragbarkeit von Erkenntnissen wird durch eine gemeinsame Beschreibungsgrundlage erleichtert),
5. *Transferfunktion* (Klassifikationen ermöglichen die Unterscheidung von Grundtypen und weniger relevanten Variationen),
6. *Innovationsfunktion* (Verdeutlichung der didaktischen Vielfalt, welche auch Praktizierende dazu anregen kann, mit noch unbekanntem Szenarien zu experimentieren),
7. *Heuristische Funktion* (mögliche Leerstellen regen zu einer Suche nach «passenden» Szenarien und damit die Entwicklung neuer Arrangements an) und
8. *Theoriefunktion* (Validierung von Strukturmodellen und Weiterentwicklung der Theorie beim Auftreten neuer Szenarien).

Zusammengefasst sind Taxonomien folglich nicht nur Systematisierungen, die einen wichtigen Beitrag zur begrifflichen Präzisierung in wissenschaftlichen Diskussionen leisten und damit zu einer klaren Verständigung über den Untersuchungsbe- reich beitragen. Sie geben darüber hinaus auch einen Überblick über die prinzipiell vorhandenen Möglichkeiten und können die Theorieentwicklung und Praxis voranbringen.

2.2 Welches Potenzial bergen Taxonomien im Hinblick auf Primärstudien, Reviews und Meta-Analysen?

Taxonomien haben sich in empirischen Studien als wichtige *Orientierungsrahmen* erwiesen. Eines der bekanntesten Beispiele ist die Lernzieltaxonomie von Bloom (1974), welche (bzw. deren spätere Revisionen) die Basis für die Lernzieldefinition in zahlreichen Studien und auch Meta-Analysen bildet (z. B. Bredow et al. 2021; Witeck, Alves, und Bernardo 2021). Die Lernzieltaxonomie unterteilt Lernerfolg in verschiedene Stufen und erleichtert die entsprechende Kategorisierung der Primärstudien nach dem ihnen zugrundeliegenden Lernerfolgsverständnis.

Über die im vorangestellten Kapitel erörterten acht Grundfunktionen von Taxonomien nach Baumgartner hinaus lassen sich thesenartig weitere Chancen für den Einsatz von Taxonomien im Rahmen von Primärstudien, Reviews oder Meta-Analysen formulieren:

1. *These:* Indem Taxonomien die Mehrdimensionalität von Begriffen und Konzepten verdeutlichen und visualisieren, schaffen sie ein Problembewusstsein für die Begriffsheterogenität und leisten einen wichtigen Beitrag zur stärkeren begrifflichen Präzision in Primärstudien sowie zur Bewertung der Aussagekraft und Reichweite der Ergebnisse von Reviews und Meta-Analysen.
2. *These:* Taxonomien bilden die verschiedenen Dimensionen und Ausprägungsformen eines Begriffs ab und können dadurch die Basis für die Formulierung von Inklusions- und Exklusionskriterien für Reviews oder Meta-Analysen darstellen, die Teilaspekte eines bestimmten Phänomens fokussieren.
3. *These:* Taxonomien ermöglichen die Strukturierung von Wissen sowie eine Hierarchisierung von Begriffsdimensionen und -elementen. In Reviews und Meta-Analysen können sie bei der Systematisierung der Forschungsbefunde helfen.
4. *These:* Eine Systematisierung und Quantifizierung von Forschungsbefunden mithilfe von Taxonomien kann im Rahmen von Reviews und Meta-Analysen dabei helfen, Forschungstrends und Forschungsdesiderate aufzuzeigen sowie neue Thesen zu generieren.

Diese Vorüberlegungen lassen darauf schliessen, dass Taxonomien als Ausgangspunkt, Orientierungsrahmen sowie Strukturierungs- und Hierarchisierungshilfe dienen können und dass deren Nutzung in Primärstudien auch zu einer erhöhten Nachvollziehbarkeit und Aussagekraft der Ergebnisse in Reviews und Meta-Analysen führen kann.

3. Taxonomien zum Begriff «Feedback» – Ein Überblick

In den letzten Jahren wurde der Begriff «Feedback» im Rahmen zahlreicher konzeptioneller Arbeiten näher beleuchtet. Die Studien weisen dabei unterschiedliche Akzentuierungen auf und beschäftigen sich u. a. mit den Merkmalen der Feedbacknachricht (Haughney, Wakeman, und Hart 2020), den benötigten Kompetenzen aufseiten der Lehrenden (Boud und Dawson 2023) und Lernenden (Carless und Boud 2018; Little et al. 2023; Zhan 2022) sowie mit interpersonalen Faktoren (Gravett und Carless 2023). Durch die Pluralisierung der Gestaltungsmöglichkeiten infolge der zunehmenden Digitalisierung gerieten auch kontextuelle, insbesondere materielle, aber auch soziokulturelle Faktoren stärker in den Blick (Chong 2021, 2022; Gravett 2022). Die Forschungsergebnisse zeigen, dass Feedbackprozesse von einem komplexen Zusammenspiel zahlreicher dynamischer Variablen beeinflusst werden, die sich in konkreten Lernumgebungen in spezifischer Weise manifestieren (Nieminen et al. 2022; Tai et al. 2021). Aufgrund der Komplexität von Feedbackdesigns bilden bisher bestehende Taxonomien nur ausgewählte Teilaspekte des Gesamtkonstrukts ab. Baumgartner und Bergner (2016) konzentrieren sich mit ihrer «Feedbacktaxonomie» beispielsweise auf verschiedene Formen von Multiple-Choice-Tests für das Online-Lernen, während Winstone et al. (2017) mit ihrer Taxonomie die Beteiligungsformen von Feedbackempfangenden fokussieren.

Diese Problematik haben Lipnevich und Panadero (2021) bzw. Panadero und Lipnevich (2022) in ihren kürzlich erschienenen Reviews kritisch beleuchtet und den Versuch unternommen, bisherige Feedbackmodelle im MISCA-Modell zusammenzuführen. Sie benennen fünf Faktorengefüge: die Merkmale der Feedbacknachricht (Message), die Umsetzung (Implementation), die Eigenschaften der Lernenden (Student), kontextuelle Variablen (Context) und die Handlungen und Interaktionen der am Feedbackprozess teilhabenden Personen (Agents). Diese Dimensionen bedürfen jedoch noch einer inhaltlichen Ausdifferenzierung und empirischen Überprüfung (Panadero und Lipnevich 2022).

Wie umfassend eine differenzierte Darstellung an Feedbackdesigns sein kann, wird mit Blick auf eine vermeintlich konkrete didaktische Variante deutlich. In ihrem Review zu «Peer Assessment Design Elements» merken Alqassab et al. (2023) zum einen an, dass es eine Vielfalt von Peer-Assessment-Design-Möglichkeiten gibt, zum anderen stellen sie eine Vagheit in deren Beschreibung in den Primärstudien fest (siehe auch Topping 1998).

Somit fehlt bislang eine Taxonomie, die die tatsächliche Vielfalt der Merkmalsausprägungen abbildet und systematisiert. Ziel unseres qualitativ-inhaltsanalytischen Scope-Reviews ist es daher, zu einer möglichst umfassenden Taxonomie zu gelangen, welche eine differenzierte und konkrete Beschreibung der didaktischen Feedbackdesigns für künftige Forschungsstudien erleichtert.

4. Erarbeitung von Taxonomien mithilfe qualitativ-inhaltsanalytischer Scope-Reviews – methodische Vorüberlegungen

Für die Erstellung einer Taxonomie ist es zentral, alle Merkmale eines Phänomens zu erfassen. Dies erfordert einen umfassenden Überblick über das Forschungsfeld und die systematische Aufarbeitung dort benannter Begriffsverständnisse. Methodisch scheint es diesbezüglich naheliegend, ein systematisches Literaturreview durchzuführen (Fink 2014). Es gibt mittlerweile eine Vielzahl an verschiedenen Review-Arten (s. z. B. Barnes et al. 2022; Grant und Booth 2009), über die oftmals der aktuelle Forschungsstand (evidenzbasiert) dargestellt und zusammenzufasst werden soll. Zugleich können Reviews aber auch für theoretische und konzeptionelle Analysen genutzt werden (Grant und Booth 2009, 92). Scope-Reviews zielen darauf ab, die Bandbreite an verfügbaren Forschungsaktivitäten, Erkenntnissen und Evidenzen zu erfassen, um darauf aufbauend die Forschungslandschaft zu kartografieren und etwaige Forschungslücken zu identifizieren (Grant und Booth 2009, 101). Laut Anderson et al. (2008, 7–8) kann über Scope-Reviews auch analysiert werden, welche unterschiedlichen konzeptionellen Verständnisse hinter Begriffen stehen («concept map»). Nach Elm, Schreiber und Haupt (2019, 1) sind Scope-Reviews dann geeignet, wenn die Literatur eine sehr heterogene Problematik aufweist. Diese Kriterien treffen auf unser Forschungsvorhaben zu. Jedoch geht es in unserem Review nicht darum, den aktuellen Forschungsstand evidenzbasiert aufzuarbeiten, vielmehr interessiert uns eine qualitative Einsicht in die jeweiligen Feedbackdesigns. Daher wurde die «klassische» Vorgehensweise von Scope-Reviews (Anderson et al. 2008, 3; Arksey und O'Malley 2005, 21) um den Schritt einer qualitativen Inhaltsanalyse ergänzt, um die Feedbackdesigns in ihrer Komplexität und Diversität zu erfassen und diese anschliessend in einer Taxonomie abzubilden.

Bei der Taxonomieentwicklung geht es – in Abgrenzung zur Grounded Theory oder Thematischen Analyse – nicht darum, Muster aufzuarbeiten oder eine abgeschlossene Theorie zu entwickeln. Stattdessen ist das primäre Ziel, ein Kategoriensystem zu entwickeln, durch welches auch Einzelfälle abgebildet werden können. Auch ist eine abschliessende Quantifizierung von nachrangiger Bedeutung. Vielmehr gilt es, die vielfältigen Merkmale von Feedbackdesigns im Rahmen des qualitativ-inhaltsanalytischen Scope-Reviews nicht nur deskriptiv herauszuarbeiten, sondern diese auch zu systematisieren und zu kategorisieren. Eine qualitative Inhaltsanalyse kann genau das leisten: Das qualitative Datenmaterial (Vielfalt der Feedbackdesigns) wird mithilfe der Herausbildung eines Codesystems (verschiedene Merkmalsdimensionen und -ausprägungen) strukturiert. Im Anschluss erfolgt eine Kategorisierung und Systematisierung des Datenmaterials (Gläser und Laudel 2010; Kuckartz 2016). In diesem Sinne stellen Taxonomien «Nebenerzeugnisse» eines jeden qualitativ-inhaltsanalytischen Literaturreviews dar (Rädiker und Kuckartz 2019, 143). Jedoch werden Kategoriensysteme nach Abschluss einer qualitativen

Inhaltsanalyse meist nicht vollumfänglich publiziert. Dadurch geht die Chance für andere Forschende verloren, diese (Vorstufen von) «Taxonomien» zu reproduzieren bzw. weiterzuentwickeln und für die Theorieentwicklung sowie für künftige Forschung weiterzuverwenden (i. A. an Hocker 2021, 2).

Für die in diesem Forschungsprojekt beschriebene Zielsetzung bietet sich das Vorgehen der qualitativen Inhaltsanalyse nach Kuckartz (2016) sowie Gläser und Laudel (2010) an. In Abgrenzung zu Mayring wird hier nicht mit einem geschlossenen Kategoriensystem gearbeitet; vielmehr gilt das Prinzip der «Offenheit» für den gesamten Analyseprozess (Gläser und Laudel 2010, 199). Gläser und Laudel (2010, 205) legen die Merkmalsausprägungen der Kategorien nicht vorab fest, sondern stellen offene Fragen an den Text (z. B. Welche Medien werden im Rahmen von Feedbackprozessen verwendet?). Des Weiteren kann das Kategoriensystem während des Prozesses iterativ ergänzt und modifiziert werden. Durch diese Flexibilität des Vorgehens können auch Einzelfälle besser erfasst werden. Eine solche qualitative Inhaltsanalyse könnte somit ein vielversprechendes Analyseverfahren für die Entwicklung von Taxonomien im Rahmen von Scope-Reviews darstellen.

5. Auf dem Weg zu einer Taxonomie von Feedbackdesigns: Ein qualitativ-inhaltsanalytischer Scope-Review

Laut Arksey und O'Malley (2005) besteht ein Scope-Review aus fünf obligatorischen und einer optionalen Phase:

1. Formulierung der Forschungsfrage
2. Studiensuche
3. Studienselektion
4. Kategorisierung des Datenmaterials
5. Berichterlegung
6. Optional: Einbindung von Praxisvertreter:innen

Die Erarbeitung einer Taxonomie auf Basis eines qualitativ-inhaltsanalytischen Scope-Reviews beginnt demnach mit (1) der Festsetzung der Forschungsziele und -fragen sowie (2) der Auswahl geeigneter Datenbanken, Suchbegriffe und Selektionskriterien. Nach (3) der Identifikation des relevanten Datenmaterials wird dieses im Schritt (4) kategorisiert. Im Verhältnis zum «klassischen» Scope-Review kartografieren wir die Forschungsstudien jedoch nicht anhand allgemeiner empirischer Parameter (bspw. Studienzeitpunkt, Teilnehmende, Ergebnisse) mit anschließender narrativer Berichterstattung. Vielmehr erweitern wir Schritt (4) um eine qualitative Inhaltsanalyse der in den Studien beschriebenen didaktischen Designs. Hierzu wird das relevante Datenmaterial in der Software MAXQDA aufbereitet, codiert und

analysiert, bevor die Resultate dargestellt werden (u. a. in Anlehnung an Fink 2014; Rädiker und Kuckartz 2019, 207). Konkret wurden dabei folgende Schritte unternommen.

5.1 Forschungsfragen

Unsere Studie fokussiert folgende Fragestellungen:

1. Welche Merkmale von didaktischen Feedbackdesigns werden in der Forschungsliteratur genannt?
2. Wie lassen sich diese Merkmale systematisch ordnen (Taxonomie)?

5.2 Inklusionskriterien

Einbezogen werden alle

1. wissenschaftlichen Journalartikel (keine Tagungsbände o. ä.),
2. bei denen es sich um peer-reviewte Interventionsstudien handelt (keine Reviews/ Meta-Analysen oder Befragungen ohne Beschreibung der Feedbackpraxis),
3. die im Zeitraum 2018–2022
4. in deutscher oder englischer Sprache publiziert wurden,
5. über die Datenbank «ERIC» frei zugänglich sind und
6. Feedbackprozesse im Hochschulkontext fokussieren (der Begriff «Feedback» kommt im Abstract vor; der Fokus liegt auf der Untersuchung von Feedbackprozessen).

5.3 Methodisches Vorgehen

Der Forschungsprozess lässt sich in insgesamt drei Phasen gliedern: den Selektionsprozess, die inhaltsanalytische Auswertung sowie die Taxonomieaufbereitung (s. Abbildung 1). Im Folgenden werden die einzelnen Schritte differenziert dargelegt.

5.3.1 Selektionsprozess

Selektionsschritt 1: Zu Beginn des Reviews erfolgte die gefilterte Suche nach allen Artikeln in der ERIC-Datenbank, in denen das Wort «Feedback» im Abstract vorkommt und welche peer-reviewed, online frei verfügbar sowie in dem Zeitraum 01.2018–07.2022 erschienen sind und sich auf den Hochschulkontext fokussieren (Educational Level = Higher Education). Auf diesem Weg wurden insgesamt 1.134 Publikationen gefunden.

Selektionsschritt 2: In einem zweiten Durchgang wurden die Abstracts der 1.134 Artikel hinsichtlich ihrer Passung zu den weiteren Inklusionskriterien überprüft. Die Abstracts wurden unter den Forschenden gleichmässig aufgeteilt und in einem ersten Schritt als «inkludierte», «exkludierte» und «unklare» Studien gekennzeichnet. Die Passung der als «unklar» markierten Abstracts wurde von den Forschenden gemeinsam diskutiert. Zudem erfolgte eine stichprobenartige Überprüfung (5 % der Gesamtstudien, Auswahl über einen Zufallsgenerator) der «inkludierten» und «exkludierten» Studien durch die jeweils andere Forschende. Der Grad der Übereinstimmung lag bei 98 %. Ein strittiger Artikel wurde im Anschluss in die Auswertung integriert. Am Ende dieser Selektionsschritte waren 202 Journal-Artikel übrig.

Selektionsschritt 3: Im vorangegangenen Selektionsschritt war zunächst die Überprüfung der Inklusionskriterien auf Basis des Abstracts erfolgt. Dabei wurden auch Studien integriert, deren Abstract nicht eindeutig die Exklusionskriterien erfüllte. Es galt daher in diesem dritten Selektionsschritt, den methodischen Teil der Studien zu sichten und erneut hinsichtlich der Passung zu den Inklusionskriterien zu überprüfen. Der Liste wurde zudem um folgendes Exklusionskriterium erweitert:

1. Die Feedbackpraxis/ das jeweils zugrunde liegende Feedbackverständnis wird nicht (ausreichend) beschrieben.

135 Studien erfüllten alle Inklusionskriterien und wurden in die Analyse einbezogen.

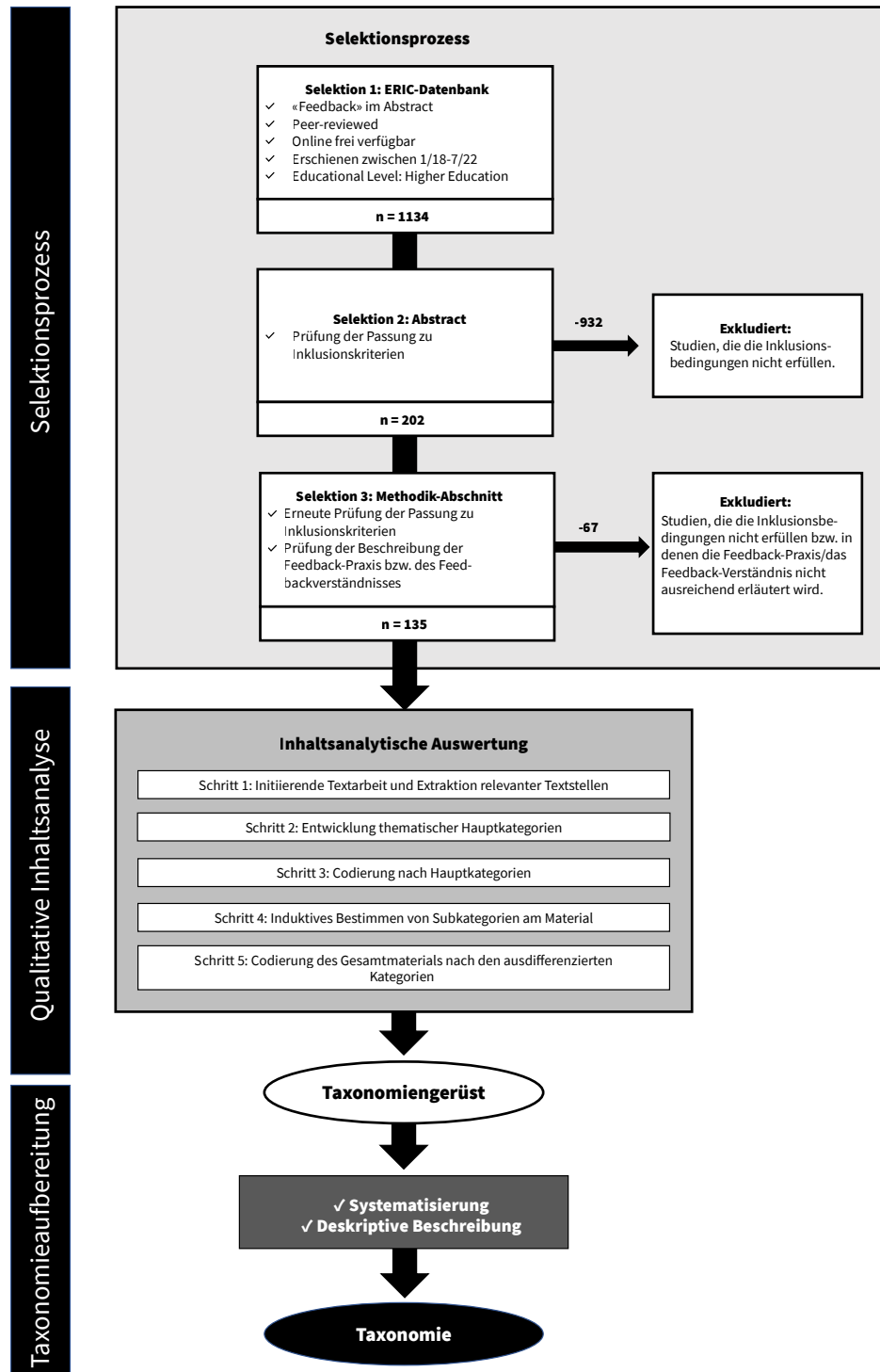


Abb. 1: Übersicht über das methodische Vorgehen.

5.3.2 Inhaltsanalytische Auswertung

Zur Taxonomieentwicklung wurde im Anschluss an die Selektionsphase eine inhaltlich strukturierende qualitative Inhaltsanalyse der 135 Studien durchgeführt. Der Ablauf gliedert sich – in Anlehnung an Kuckartz (2016, 101–15) – in fünf Schritte. Das Vorgehen wurde dabei vereinzelt um Elemente von Gläser und Laudel (2010) erweitert.

Schritt 1: Initiierende Textarbeit und Extraktion relevanter Textstellen

Im Zuge der initiierenden Textarbeit haben sich bezogen auf unsere Fragestellungen insgesamt acht Hauptkategorien herauskristallisiert, die sich durch die folgenden Fragen abbilden lassen:

1. WER gibt Feedback («Feedbackgebende/r»)?
2. WEM wird Feedback gegeben («Feedbacknehmende/r»)?
3. WORAUF bezieht sich das Feedback («Feedbackinhalt»)?
4. WAS wird mit dem Feedback bezweckt? («Feedbackzweck»)
5. WORÜBER und WIE wird Feedback gegeben («Feedbackmedium und -modus»)?
6. WANN wird Feedback gegeben («Feedbackzeitpunkt»)?
7. WIE OFT wird Feedback gegeben («Feedbackhäufigkeit»)?
8. WELCHE Rolle spielen die/der Feedbacknehmende und die/der Feedbackgebende im Prozess («Feedbackinteraktionen»)?

Diese acht iterativ gewonnenen Leitfragen bildeten die Basis für die Identifikation relevanter Textstellen. Letztere wurden aus den Journalartikeln extrahiert, in Word-Dokumente eingefügt und zur weiteren Analyse und Codierung in MAXQDA importiert. Die Extraktion relevanter Textstellen und damit auch eine konsequente Trennung vom Ursprungstext erfolgte somit vor der Codierung des Gesamtmaterials (siehe hierzu auch Früh 2007, 157; Gläser und Laudel 2010, 200).

Schritt 2: Entwicklung thematischer Hauptkategorien

Mit dem Ziel der Bewahrung des Prinzips der Offenheit wurde eine induktive Erweiterung und potenzielle Revision der Hauptkategorien nicht ausgeschlossen (Gläser und Laudel 2010, 203). In der Tat hat sich während des Codierprozesses herausgestellt, dass die Kategorien «Feedbackzeitpunkt» und «Feedbackhäufigkeit» meist in einem unmittelbaren Zusammenhang miteinander stehen und kaum voneinander zu trennen sind. Infolgedessen wurde die «Feedbackhäufigkeit» von einer Hauptkategorie in eine Unterkategorie umgewandelt. Alle weiteren Hauptkategorien wurden bestätigt und es erfolgten keine weiteren Modifikationen. Es verblieben demnach sieben thematische Hauptkategorien: Feedbackgebende/r, Feedbacknehmende/r, Feedbackinhalt, Feedbackziel, Feedbackmedium/-modus, Feedbackhäufigkeit sowie Feedbackinteraktionen.

Schritt 3: Codierung nach Hauptkategorien

Bei der Codierung des Datenmaterials wurden folgende Regeln eingehalten (i. A. an Kuckartz 2016, 102–4):

1. Jede Codiereinheit muss eine «Sinneinheit» sein, d. h. die Textstellen müssen inhaltlich zu einer der thematischen Hauptkategorien in Verbindung stehen und in sich geschlossen sein.
2. Jede Codiereinheit muss für sich allein ausreichend verständlich sein und kann einzelne Sinneinheiten oder mehrere Sätze umfassen.
3. Eine Textstelle kann mehreren thematischen Hauptkategorien zugeordnet werden, sofern in ihr mehrere Themen angesprochen werden.
4. Es erfolgt keine Codierung thematisch nicht relevanter Textstellen.

Schritt 4: Induktives Bestimmen von Subkategorien am Material

Im Anschluss an die thematische Strukturierung erfolgt in diesem Schritt nun das induktive Bestimmen der Subkategorien. Hierfür wurden anhand des Datenmaterials für jede Hauptkategorie die relevanten (Unter-)Dimensionen identifiziert und zu Subcodes zusammengefasst. Hierbei stellte sich heraus, dass es bei einem Teil der Subcodes notwendig war, diese weiter auszudifferenzieren, sodass eine Art «Hierarchie unter den Subcodes» entstand.

Dieser Schritt stellt das Kernstück der Taxonomiebildung dar, da hier alle Dimensionen bzw. Ausprägungsformen der vielfältigen Feedbackdesigns herausgearbeitet werden. Am Ende dieses Prozesses entsteht ein Taxonomiegerüst, welches abschliessend noch in eine (deskriptive) Taxonomie überführt werden muss.

Schritt 5: Codierung des Gesamtmaterials nach den ausdifferenzierten Kategorien

Durch eine abschliessende Codierung des Gesamtmaterials wird die Vollständigkeit der in Schritt 4 gebildeten Subkategorien überprüft. Hierdurch wird die Erfassung von Einzelfällen sichergestellt. Zugleich eröffnet dieser Schritt aber auch die Möglichkeit der späteren Quantifizierung des Vorkommens bestimmter Aspekte der Feedbackimplementierungen und unterstützt damit tiefergehende Analysen (z. B. Aufzeigen von Forschungstrends und -desideraten oder Visualisierung von Kategorienzusammenhängen).

5.3.3 Taxonomieaufbereitung

Das im Rahmen der qualitativen Inhaltsanalyse erstellte Taxonomiegerüst wird in diesem Schritt nochmals grundlegend strukturiert und systematisiert sowie um Deskriptionen ergänzt. Neben der Zusammenfassung bedeutungsgleicher Kategorien (Gläser und Laudel 2010, 203) galt es folglich, die Haupt- und Subkategorien in eine hierarchische Ordnung zu bringen.

5.4 Zentrale Ergebnisse: Die Taxonomie

Die im Zuge des qualitativ-inhaltsanalytischen Scope-Reviews gebildete Taxonomie zu didaktischen Feedbackdesigns umfasst insgesamt sieben Hauptkategorien, welche sich in 259 Subkategorien (54 erster Ordnung, 83 zweiter Ordnung, 102 dritter Ordnung, 8 vierter Ordnung) aufteilen (siehe Abbildungen 2-8). Damit verdeutlicht das Ergebnis eindrücklich die grosse Vielfalt der in den 135 Studien analysierten Feedbackdesigns über die sieben Hauptkategorien hinweg. An dieser Stelle kann nicht im Detail auf die einzelnen Kategorienstränge eingegangen werden. Es sei jedoch angemerkt, dass sich viele Feedbackdesigns nicht eindeutig nur einem Kategorienstrang zuordnen lassen. So gibt es beispielsweise Studien, bei denen sowohl die Dozierenden als auch die Peers den Studierenden Feedback geben (z. B. Farsani und Aghamohammadi 2021) oder das Feedback zu verschiedenen Zwecken erfolgt (z. B. Navarro Jover 2021). Eine klare Abgrenzung der Einzelkategorien ist daher nicht immer möglich, was die beachtliche Diversität der Primärstudien noch einmal verdeutlicht:

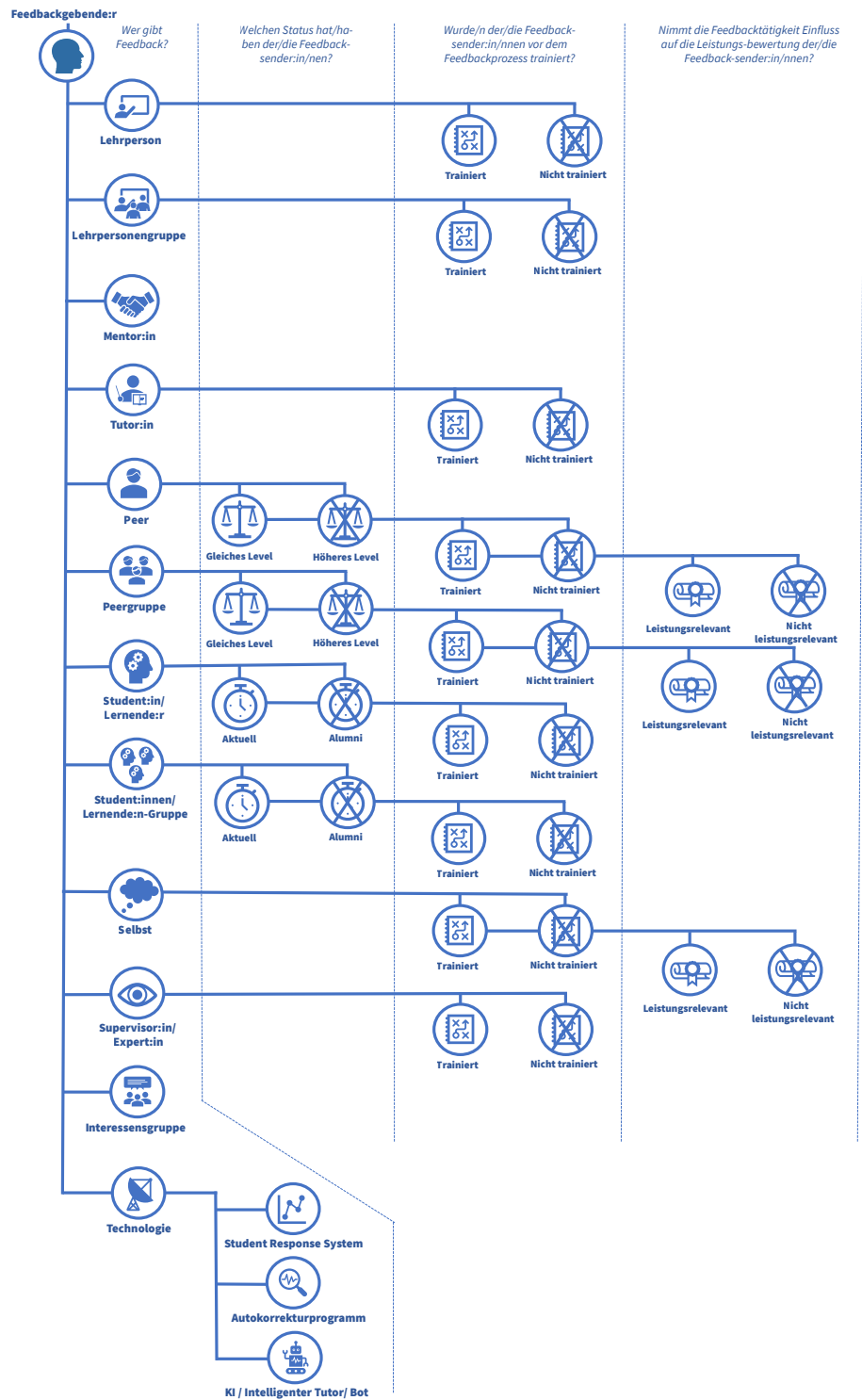


Abb. 2: Feedbacktaxonomie - Feedbackgebende:r.

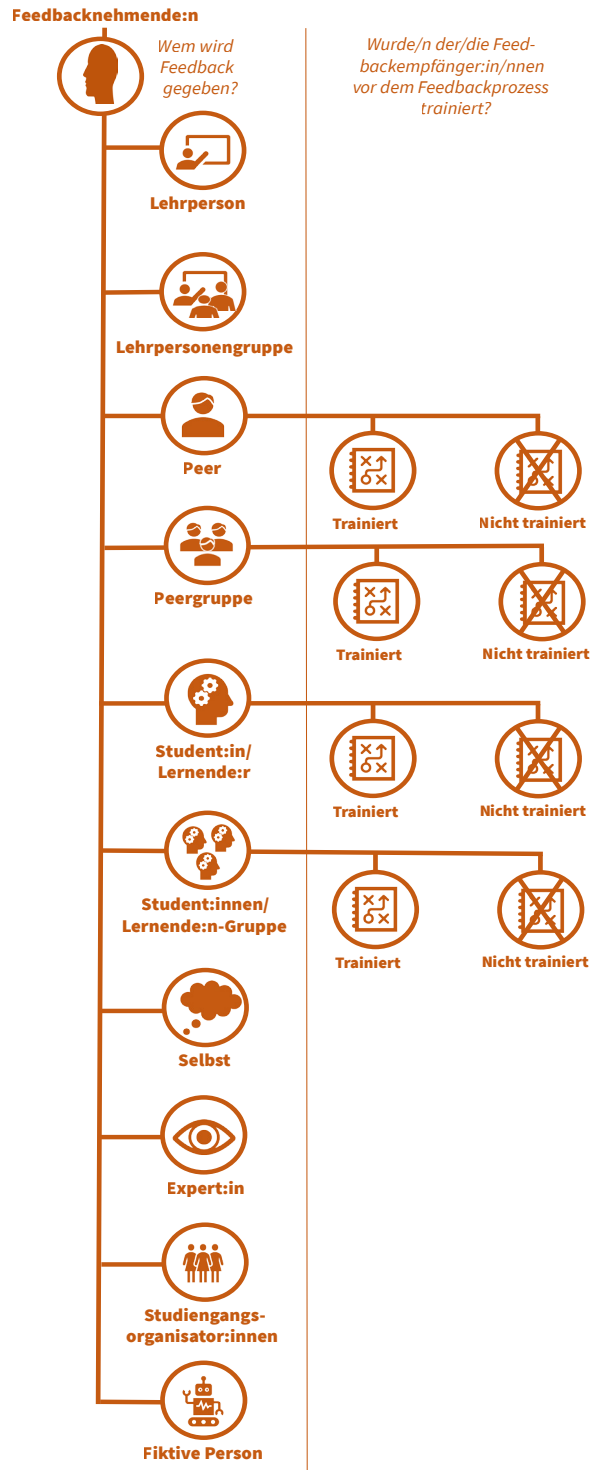


Abb. 3: Feedbacktaxonomie – Feedbackempfänger:innen.

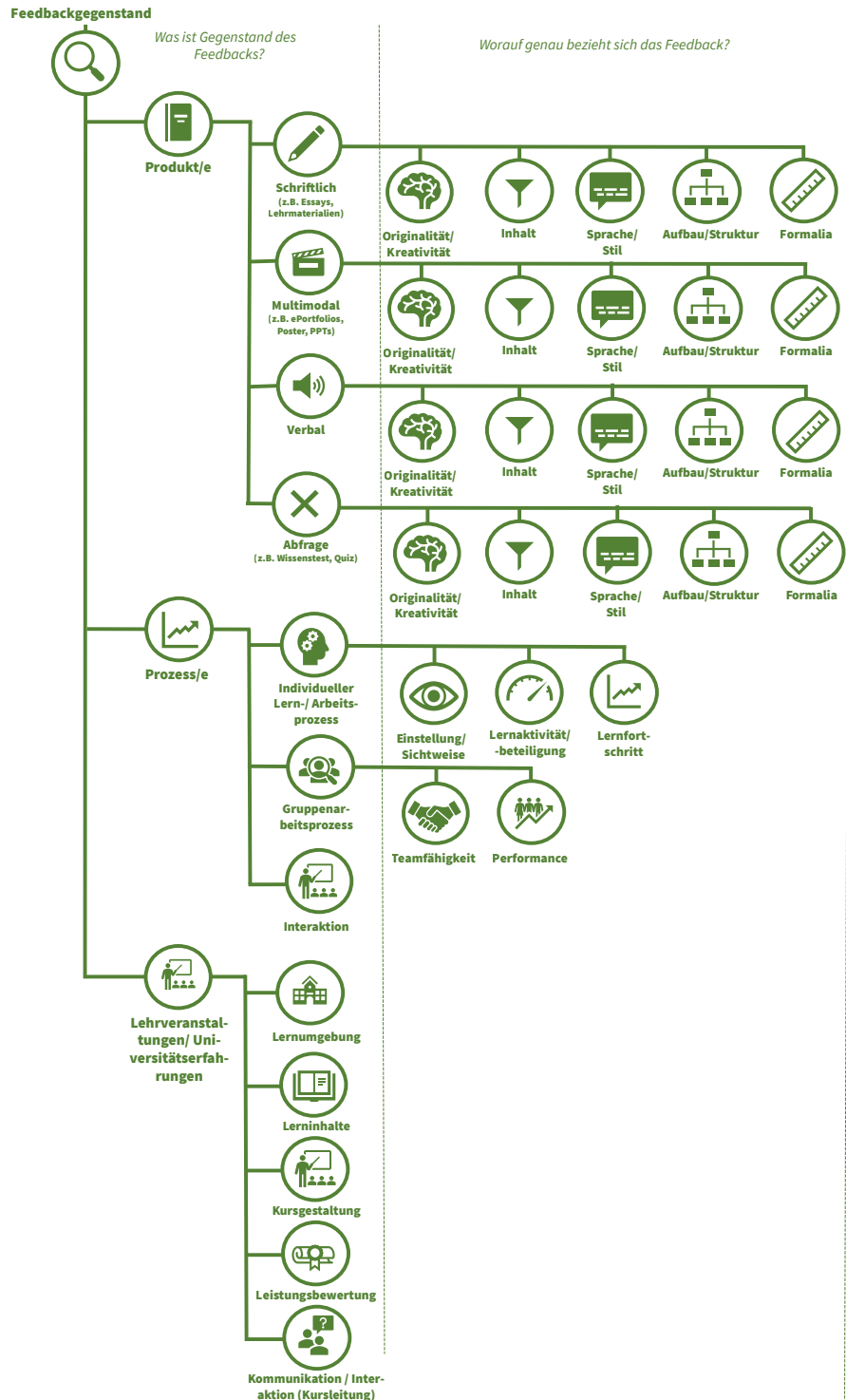


Abb. 4: Feedbacktaxonomie – Feedbackgegenstand.

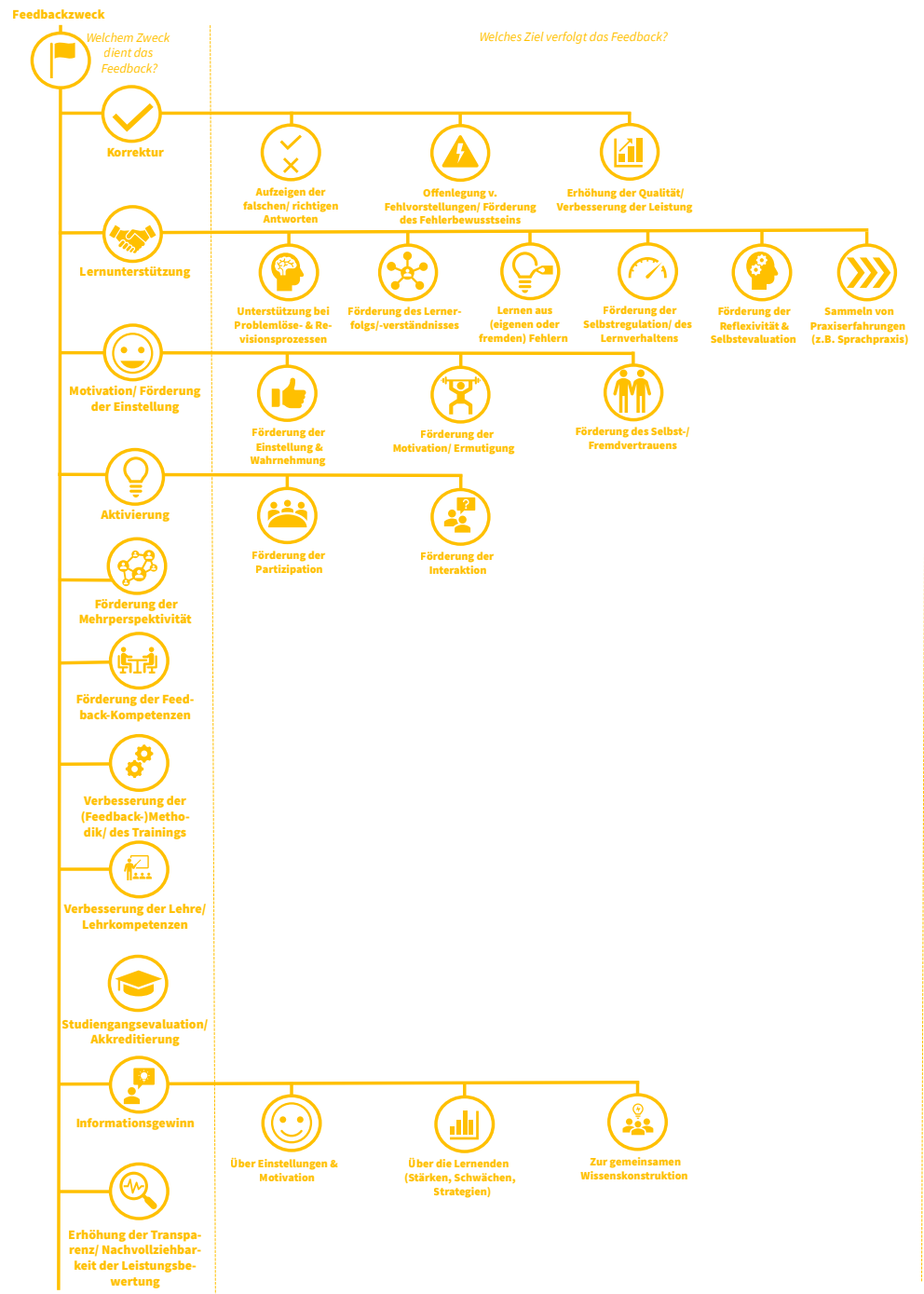


Abb. 5: Feedbacktaxonomie – Feedbackzweck.

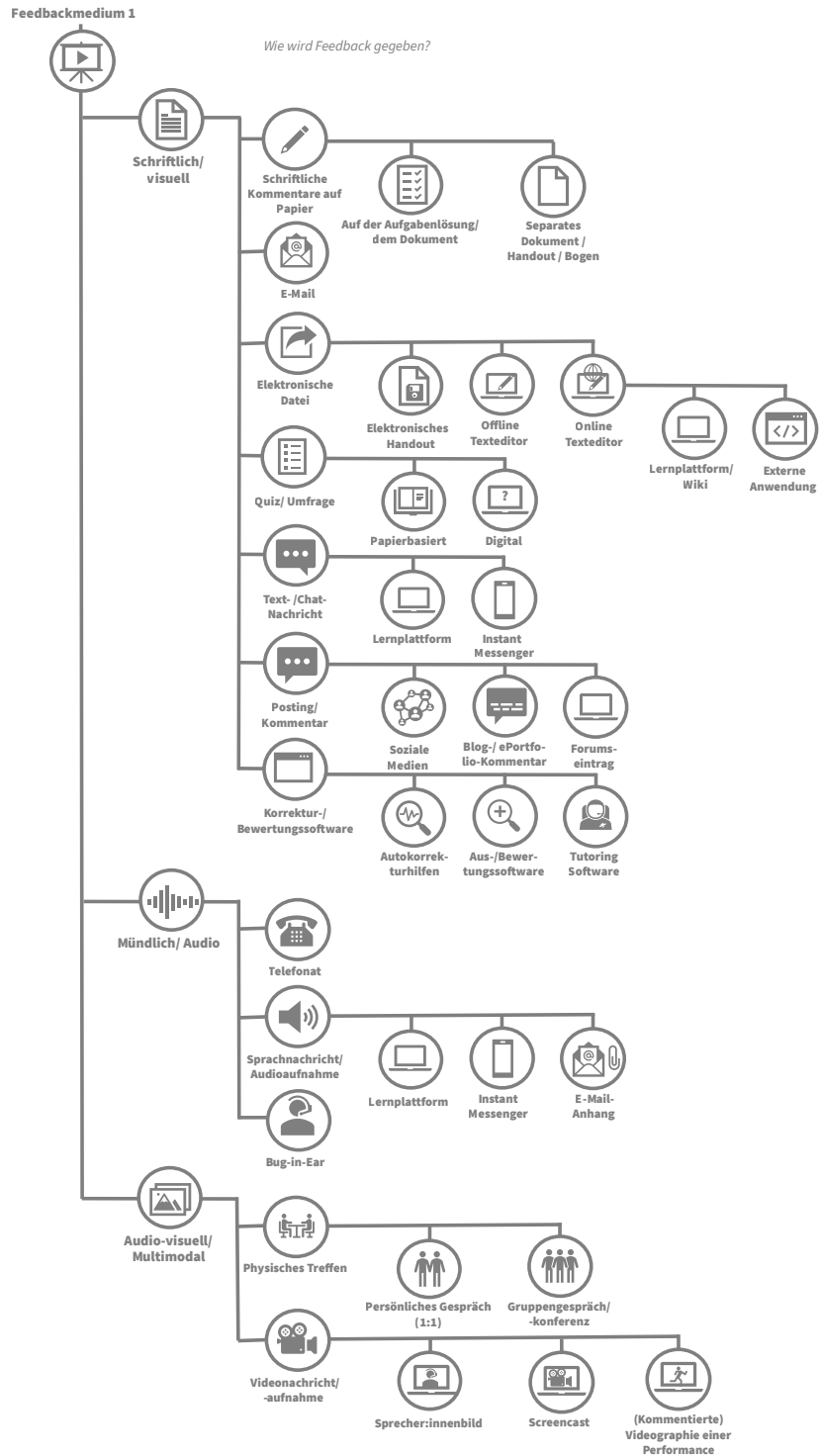


Abb. 6: Feedbacktaxonomie – Feedbackmedium 1.

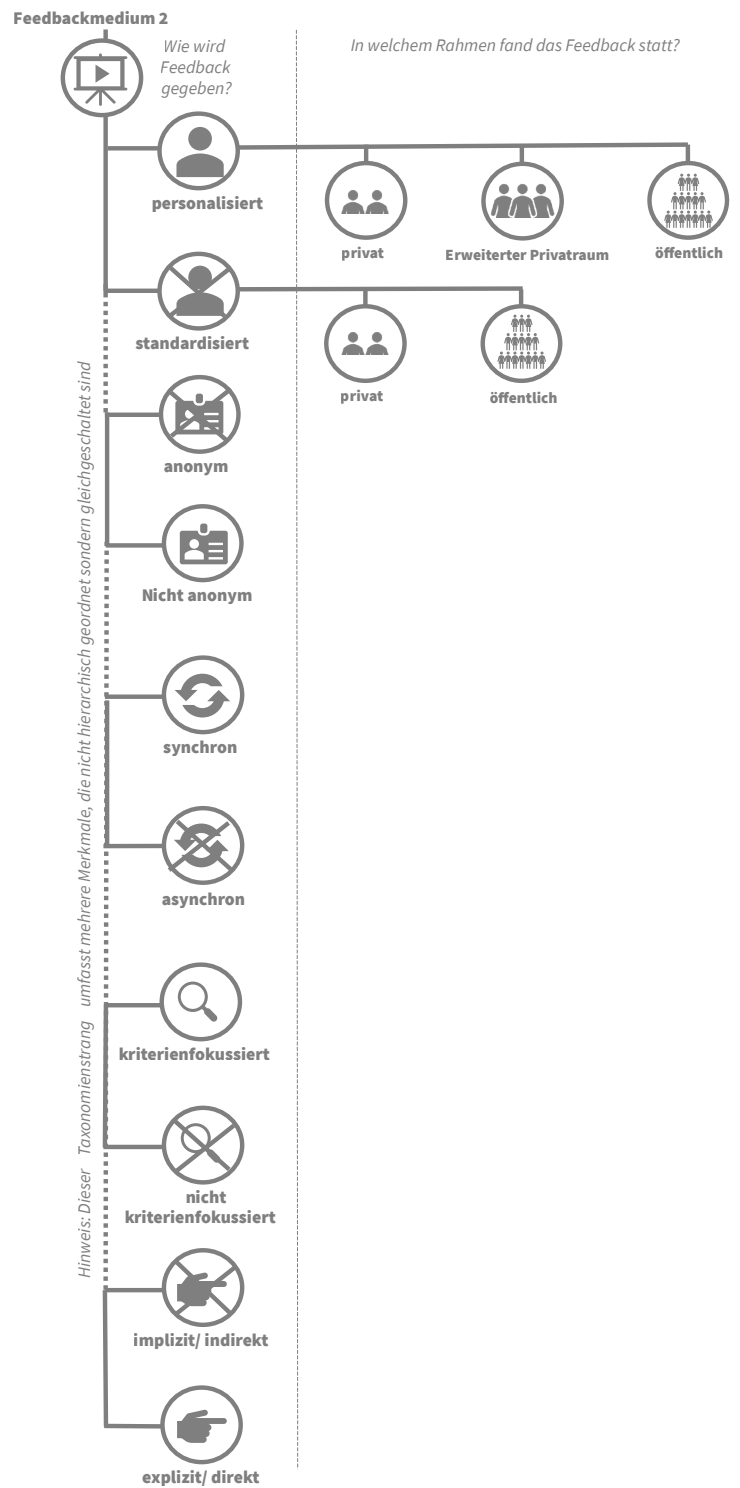


Abb. 7: Feedbacktaxonomie – Feedbackmedium 2.

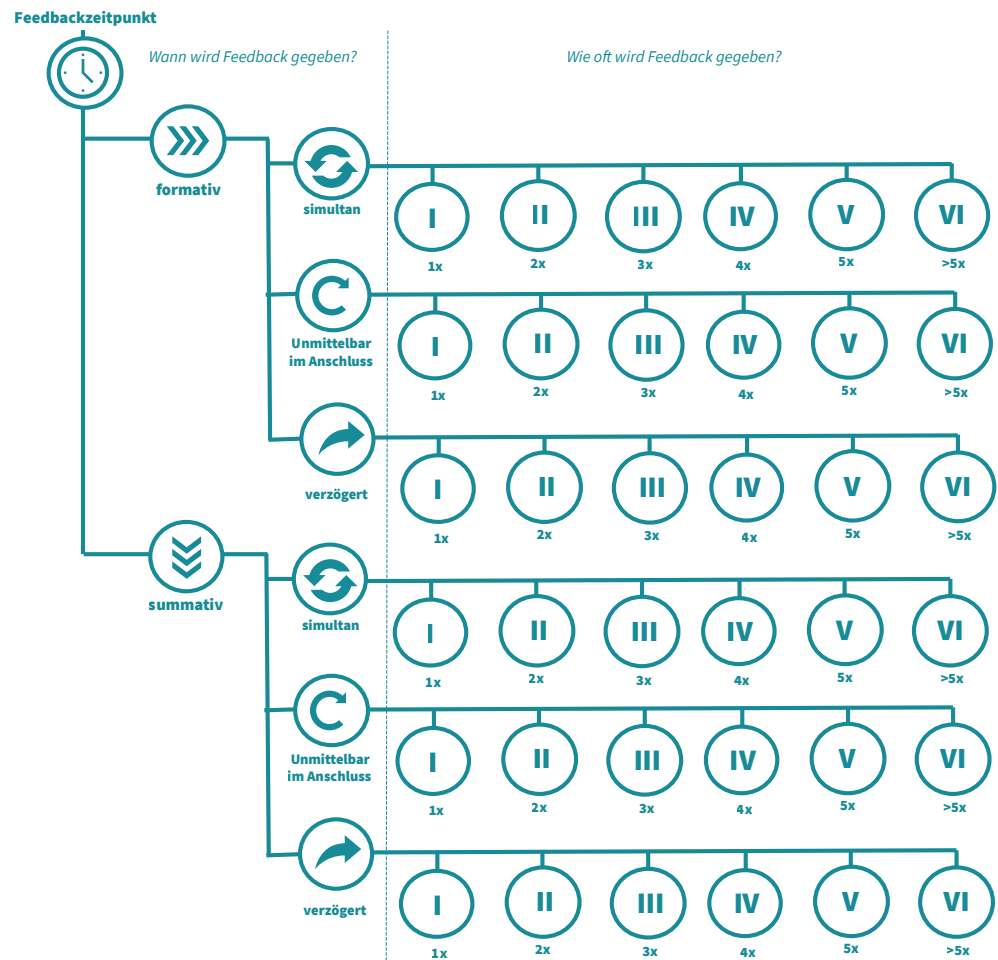


Abb. 8: Feedbacktaxonomie – Feedbackzeitpunkt.

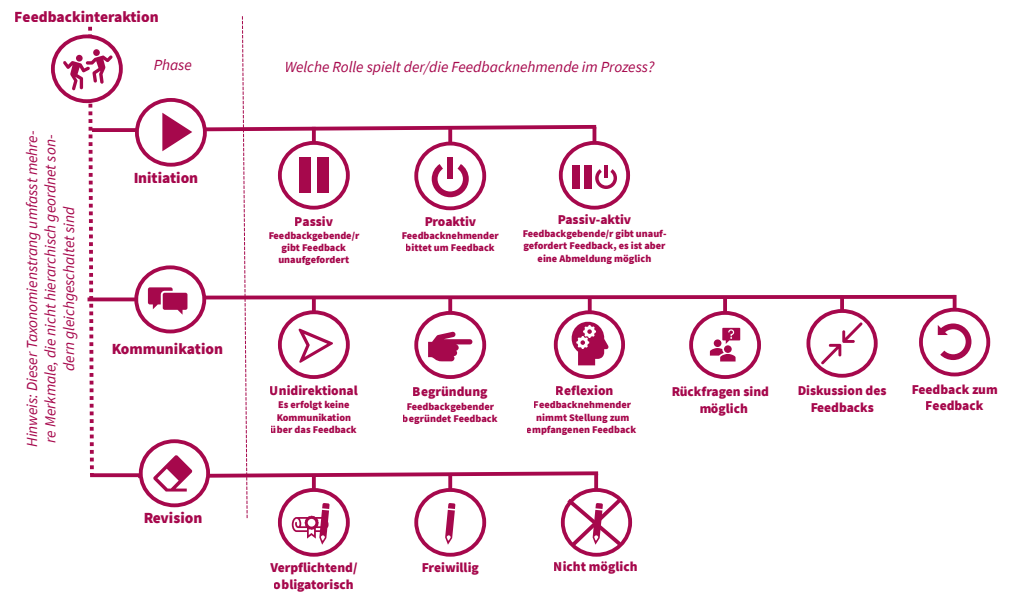


Abb. 9: Feedbacktaxonomie – Feedbackinteraktion.

5.5 Ausblick: Wie kann mit der Taxonomie gearbeitet werden?

Die Taxonomie kann in Forschungsstudien dazu genutzt werden, differenzierte und transparente Beschreibungen der Feedbackdesigns zu formulieren. Dabei ist die Taxonomie im Sinne eines «Baukastensystems» zu verstehen. Pro Kategorie gilt es zu entscheiden, welches Element sich am besten dazu eignet, das eigene Feedbackdesign zu beschreiben. Vereinzelt müssen ggf. noch Bausteine weiter expliziert werden (siehe Abbildung 4).

Beispiel 1:

Bei dem im Rahmen dieser Studie praktizierten Feedback gaben die aktuellen **Studierenden (untrainiert)** der **Lehrkraft** ein **Einzelfeedback**. Hierfür füllten sie schriftlich eine digitale Umfrage aus. Dies erfolgte **privat, standardisiert, kriterienfokussiert, anonym und asynchron**. Inhaltlich bezog sich das Feedback auf die **Lehrveranstaltung**, insbesondere auf die **Kursgestaltung** und die **Kursleitung** und zielte auf die **Verbesserung der Lehre**. Die Lehrkraft bat selbst **proaktiv** um Feedback, welches **unidirektional** erfolgte (Rückfragen und eine Diskussion waren nicht möglich). Es handelte sich um ein **summatives** Feedback, welches **einmalig** in **unmittelbarem Anschluss** an die letzte Seminarsitzung erfolgte. Alle Studierenden **mussten** an dem Feedbackprozess teilnehmen.

Beispiel 2:

Bei dem im Rahmen dieser Studie praktizierten Feedback gab die **Lehrperson** über einen „Bug-in-Ear“ einem bzw. einer **Studierenden** während einer Mikro-Lehrstunde ein **synchrones, persönliches, direktes** Feedback. Beide Akteurinnen erhielten vorab ein **Training** bezüglich der Nutzung dieser Technologie. Da sich die Lehrperson während des Feedbackprozesses im gleichen Raum befand, erfolgte das Feedback **nicht anonym** und im **erweiterten Privatraum** (ein paar weitere Studierende, die neben der Lehrkraft saßen, konnten das Feedback mithören). Inhaltlich bezog sich das Feedback auf die **Interaktion** der bzw. des Studierenden und zielte auf die **Korrektur** der Gestik und Mimik, mit dem Ziel der **Verbesserung des Kommunikationsverhaltens**. Jede/r Studierende erhielt bei **drei** verschiedenen Mikro-Lehrstunden auf diesem Weg ein **formatives-simultanes** Feedback. Im Anschluss an jede Sitzung erfolgten eine **gemeinsame Reflexion** und **Diskussion des Feedback-Prozesses**. Hier hatten die Studierenden auch die Möglichkeit, auf das Feedback zu reagieren. Dabei wurde ebenfalls erwartet, dass die **Kritik** in den nächsten Stunden **umgesetzt** wurde. Diese Revision des Verhaltens wurde auch im Rahmen der Leistungsbewertung berücksichtigt.

Abb. 9: Beispiele: Bildung von Realdefinitionen mithilfe der Taxonomie (Baukasten-System).

Die zwei Beispiele mit den farblich hervorgehobenen Kategorien verdeutlichen, wie eine differenziertere Beschreibung des Feedbackdesigns unter Einbezug der Taxonomie aussehen kann.

6. Diskussion des methodischen Vorgehens, «Lessons learned» und Limitationen

Bereits seit Jahrzehnten ist «Feedback» ein zentraler Gegenstand der pädagogischen Forschung. Die ungefilterte Suche nach dem Begriff «Feedback» lieferte allein in der Datenbank «ERIC» 42.243 Resultate.² Aufgrund der grossen Trefferanzahl erscheint es für die Anfertigung eines Reviews unabdingbar, strenge Exklusionskriterien zu formulieren, um zu einer handhabbaren Menge an Analysematerial zu gelangen. Die von uns gewählten Exklusionskriterien haben zu einer deutlichen Reduktion des Gesamtmaterials geführt. Dennoch war die Ausgangsbasis mit 1.134 Studien sehr umfangreich. Zugleich gehen die strengen Exklusionskriterien auch mit Beschränkungen einher: Die Fokussierung auf die Forschungsliteratur der letzten 5 Jahre führt dazu, dass nur die aktuellsten Entwicklungen erfasst wurden. Ausserdem zeigte sich während des Arbeitsprozesses, dass besonders aussergewöhnliche Studien häufig

² Zahl entnommen von <https://eric.ed.gov/?q=Feedback>, Stand 05.04.23, 11:04 Uhr.

das Inklusionskriterium «Interventionsstudie» nicht erfüllten. Die Exklusion dieser speziellen Studien hatte zur Folge, dass auch die Taxonomie diese «besonderen» Feedbackprozesse nicht umfasst. Exemplarisch können bezogen auf die Hauptkategorien «Feedbackgebende/r» und «Feedbacknehmend/e» die Studie von Kanuka und Sadowski (2020) (Peer-Feedback auf Dozierenden-Ebene) und die Studie von Vencille et al. (2021) (Feedback von Praktiker:innen aus der Arbeitswelt an Studierende) angeführt werden. Die Exklusionskriterien haben daher direkte Auswirkungen auf den Inhalt und die Vollständigkeit der mithilfe eines qualitativ-inhaltsanalytischen Scope-Reviews erstellten Taxonomie. Vor allem das Exklusionskriterium «Interventionsstudie» hat sich als ungünstig erwiesen. Mögliche Alternativen für die Reduktion des Analysematerials auf eine zumutbare Menge wären der Einsatz eines Zufallsgenerators oder eines Relevanzfilters (z. B. Mindesthäufigkeit der Verwendung des Feedbackbegriffs in der Publikation). Es kann jedoch nicht ausgeschlossen werden, dass auch dies zur Exklusion von Sonderfällen führt.

Wenn es gilt, die gesamte Bandbreite der Feedbackdesigns abzubilden, ist Einzelfällen eine besondere Bedeutung beizumessen. Dies bedeutet aber auch, dass eine theoretische Sättigung – wie sie z. B. im Rahmen der Grounded Theory beschrieben wird (Strauss und Corbin 1996) – nicht erreicht werden kann. Die im Rahmen des qualitativ-inhaltsanalytischen Scope-Reviews erstellte Taxonomie erfüllt nicht den Anspruch auf Vollständigkeit. Als Beispiel sei hier das Feedback per Videokonferenz genannt, das in den analysierten Studien und dementsprechend auch im Taxonomiegerüst nicht thematisiert wird. Laut einer Studierendenbefragung von Brück-Hübner (2022, 2023) stellte gerade das Feedback über Videokonferenztools jedoch die häufigste Feedbackart in der Hochschullehre während der Corona-Pandemie dar. Hier zeigt sich folglich eine eklatante Lücke in der Taxonomie – und womöglich auch in der Forschung –, die es noch zu schliessen gilt. Inhaltliche Lücken in der Taxonomie könnten u. a. durch die Integration weiterer Taxonomien geschlossen werden («Meta-Taxonomie»). So benennen z. B. Alqassab et al. (2023) weitere potenzielle Gestaltungsdimensionen zu «Peer-Feedback» (z. B. Gruppenkonstellation, Gruppenbildungsprozess), die zu einer weiteren Ausdifferenzierung der Taxonomie beitragen können. Auch der optionale Schritt in einem Scope-Review, der Praxis-Dialog, kann in dieser Hinsicht gewinnbringend sein (Arksey und O'Malley 2005). Gespräche mit Praxisvertreter:innen und anderen relevanten Stakeholdern können nicht nur Einblicke in die Praxistauglichkeit der Taxonomie gewähren, sondern auch zu deren Erweiterung beitragen. Es ist anzunehmen, dass auch keiner dieser alternativen methodischen Zugänge allein zu einer allumfassenden Taxonomie führt. Vielmehr bietet sich ein Mixed-Methods-Design an.

Trotz der relativ eingeschränkten Stichprobe verdeutlichen die Ergebnisse des qualitativ-inhaltsanalytischen Scope-Reviews eindrücklich, wie unterschiedlich und mehrdimensional die Konzepte, die sich hinter dem Begriff «Feedback» verbergen,

in der (Forschungs-)Praxis sind. Die Taxonomie stellt unseren Recherchen zufolge die aktuell umfassendste Darstellung verschiedener Feedbackdesigns dar und kann damit – vor allem auch bezüglich des Umgangs mit Primärstudien in Sekundäranalysen – ein Problembewusstsein für diese Begriffsheterogenität schaffen (s. Kapitel 3, These 1).

Ein weiteres Ziel von Taxonomien ist es, Phänomene zu strukturieren und zu systematisieren, um einen Überblick über die Komplexität eines Konstrukts zu gewinnen. Dazu sollen sie einerseits möglichst umfassend, andererseits aber auch in sich geschlossen und begrenzt sein (s. Kapitel 3). Im Rahmen der Taxonomienbildung galt es daher zu entscheiden, welche Aspekte integriert werden und welche nicht. Solche Entscheidungen wurden während des Forschungsprozesses konsensuell unter Abwägung der jeweiligen Relevanz für das Feedbackverständnis getroffen. Dennoch ist nicht von der Hand zu weisen, dass die Grenzziehung einer gewissen Beliebigkeit unterliegt. So ist z. B. eine der Subkategorien von Peer-Feedback das «Training». Hierzu wurde eine Unterscheidung zwischen «trainiert» und «nicht trainiert» vorgenommen. Das Datenmaterial zeigte jedoch, dass es auch bezüglich der Ausgestaltung der Feedbacktrainings deutliche Unterschiede geben kann (z. B. «peer feedback screencast training sessions» [Irwin 2019, 46] oder Präsenztrainings [Daweli 2018, 273]). Auch hier bietet sich in einem zweiten Schritt ein Abgleich mit anderen Taxonomien an, um zu entscheiden, bis zu welchem Grad die Differenzierung erfolgen sollte. So nehmen z. B. Alqassab et al. (2023, 18) in ihrer Taxonomie ebenfalls keine weitere Unterscheidung der Art und Weise des Feedbacktrainings vor und ziehen damit an der gleichen Stelle die Grenze wie wir.

Trotz der Grenzziehungen kann das Ergebnis des hier beschriebenen Forschungsprozesses nicht dem Anspruch der «Kompaktheit» und «Einfachheit» genügen. Mit 259 Kategorien wirkt die Taxonomie zunächst schwer zugänglich. Andererseits erleichtert die hierarchische Struktur und Auffächerung in Unterkategorien die Navigation durch das Kategoriensystem. Es kann daher durchaus für zukünftige Studien als Orientierungsrahmen und Strukturierungshilfe dienen. Die Taxonomie kann z. B. bei Reviews oder Meta-Analysen dabei helfen, sich durch die Wahl entsprechender Inklusions- und Exklusionskriterien gezielt auf bestimmte Teilaspekte und deren Zusammenhänge zu fokussieren (s. Kapitel 3, These 2 und 3).

In Anbetracht der Komplexität der Taxonomie gilt es auch, die grundlegende Frage zu diskutieren, ob Taxonomien nicht ein zu starres Konstrukt sind, um dynamische Interaktionen – wie z. B. in Feedbackimplementierungen – zu beschreiben. Panadero und Lipnevich (2022) weisen darauf hin, dass es eine Vielfalt an verschiedenen miteinander interagierenden (Sub-)Dimensionen gibt. Dieses Phänomen der inhaltlichen Überlappungen bzw. engen Verknüpfungen zwischen (Unter-)Kategorien konnten wir während des Codierprozesses ebenfalls feststellen. Beispielsweise dient formatives Feedback häufig der Lernunterstützung und an (hand)schriftlichen

Textprodukten werden zumeist schriftliche Korrekturen vorgenommen. Es erscheint daher sinnvoll, in weiterführenden Analysen die Häufigkeit gewisser Paarungen tiefergehend zu analysieren und mögliche feste Zusammenhänge aufzuarbeiten. Diese Ergebnisse können zu einer weiteren Verfeinerung der Taxonomie beitragen.

Trotz solcher Feinarbeiten können in Taxonomien die komplexen Interaktionen zwischen verschiedenen Kategorien vermutlich nicht vollständig abgebildet werden. Dies ist aber auch gar nicht der Anspruch einer Taxonomie, denn eine solche ist durch thematische Schwerpunktsetzungen per se immer begrenzt. Im Rahmen des hier beschriebenen Reviews wurde der Fokus ausschliesslich auf die Dimensionen gelegt, die den Feedbackbegriff in Form der «W-Fragen» umschreiben. Dimensionen jenseits dessen, z. B. der instruktionale Kontext oder affektive Aspekte (Gravett und Carless 2023, 2), blieben unberücksichtigt. Wie bereits erörtert, sind Grenzziehungen im Rahmen der Taxonomienbildung essenziell, denn nur so kann ein Werkzeug entstehen, welches auch in weiteren Forschungsarbeiten genutzt werden kann.

Eine weitere Erkenntnis, die wir gewinnen konnten, ist, dass es in den Primärstudien grosse Unterschiede bezüglich der Differenziertheit und Detailliertheit der Beschreibung der jeweiligen didaktischen Feedbackdesigns gibt. Damit sind unsere Erkenntnisse auch anschlussfähig an die von Alqassab et al. (2023). Die Studien, die die Feedbackdesigns nicht explizit darlegen, wurden bereits im Rahmen der Selektionsphase III exkludiert. Dennoch fehlten bei einigen inkludierten Studien Informationen zu ausgewählten Hauptkategorien oder aber darin wurden einzelne Hauptkategorien nur oberflächlich beschrieben, sodass keine eindeutige Zuordnung zu den Subcodes erfolgen konnte. Diese Feststellung führt zu der Grundsatzfrage, welche Informationen bei der Beschreibung einer Interventionspraxis (in unserem Fall einer Feedbackpraxis) in Primärstudien mindestens vorhanden sein sollten, um die Nachvollziehbarkeit des untersuchten Kontexts zu gewährleisten. Auch in diesem Sinne könnten Taxonomien bei der Formulierung adäquater Inklusions- und Exklusionskriterien helfen, sodass bestimmte fehlende Informationen zu einer Exklusion der Primärstudie führen. Um eine optimale Kompatibilität sicherzustellen, wäre es wünschenswert, dass bereits in Primärstudien eine an der Taxonomie orientierte adäquate Beschreibung des Feedbackdesigns erfolgt. Zugleich wirft ein solches Vorgehen aber auch die Frage auf, ob dies eine «Vereinheitlichung» der Studienlandschaft zur Folge haben könnte. Vor allem neue Ansätze und Ideen werden von Taxonomien (noch) nicht abgedeckt. Eine zu starke Fokussierung auf Taxonomien

kann den Blick demnach nicht nur weiten, sondern auch eingrenzen. Aufgrund dynamischer Entwicklungen (z. B. der medialen Transformationen) besteht daher die Notwendigkeit, Taxonomien in regelmässigen Abständen kritisch zu prüfen und ggf. zu modifizieren.

7. Fazit und Ausblick

Die Bildung von Begriffstaxonomien ist sehr komplex und erfordert umfassende Recherchen. Die Durchführung eines qualitativ-inhaltsanalytischen Scope-Reviews zur Taxonomieerstellung hat sich hierbei als bedingt zielführend erwiesen. Zum einen zeigt die Komplexität des Ergebnisses eindrucksvoll die Vielfalt der didaktischen Szenarien auf, leistet damit einen wichtigen Beitrag zur Schaffung eines Problembewusstseins beim Umgang mit didaktischen Konzepten und bietet sich für eine Weiternutzung und kritische Prüfung an. Zum anderen stellt sich jedoch die Frage, inwiefern ein solch komplexes und daher schwer zugängliches Konstrukt als ordnendes Element in Folgestudien verwendet werden kann. Die Herausarbeitung von Subtaxonomien (z. B. fokussiert auf einzelne Medien oder Personengruppen) könnte dabei helfen, die Komplexität zu reduzieren. Zugleich ist es aber – in Abgrenzung zu anderen Feedbacktaxonomien – gerade der Mehrwert der vorliegenden Studie, dass eben keine zu starke Schwerpunktsetzung vorgenommen, sondern vielmehr eine Vielfalt an Feedbackdesigns abgebildet wurde.

Dabei gilt es jedoch zu betonen, dass die im Rahmen des qualitativ-inhaltsanalytischen Scope-Reviews gebildete Taxonomie noch nicht als abgeschlossen angesehen werden kann. Der Vergleich mit anderen Taxonomien und Studien hat bereits an einigen Stellen inhaltliche Lücken offengelegt. Es besteht somit die Notwendigkeit der Erweiterung, z. B. durch die Integration von Kategorien aus anderen Taxonomie-Modellen («Meta-Taxonomie») oder aber durch einen Praxis-Dialog (Arksey und O'Malley 2005).

Im Rahmen des qualitativ-inhaltsanalytischen Scope-Reviews ist zugleich aber auch eine codierte Primärstudien-/Materialsammlung entstanden, die eine ausgezeichnete Basis für weitere Forschungsprojekte bildet. Mithilfe tiefergehender Analysen könnten typische Kategorienkonstellationen herausgebildet und seltene Einzelfälle separiert werden. Zugleich ermöglichen Quantifizierungen Einblicke in Forschungstrends und -desiderate. Während ersteres eine gute Basis für die Erstellung fokussierter Literaturreviews oder Meta-Analysen darstellt, kann letzteres dabei helfen, den Blick zu weiten und auf eher selten betrachtete Feedbackphänomene zu lenken, welche in Primärstudien noch einer tiefergehenden Analyse bedürfen. Der Prozess der Taxonomiebildung bietet somit über das ursprüngliche Forschungsanliegen hinaus ein grosses methodisches Potenzial für zukünftige Studien.

Literatur

- Alqassab, Maryam, Jan-Willem Strijbos, Ernesto Panadero, Javier Fernández Ruiz, Matthijs Warrens, und Jessica To. 2023. «A Systematic Review of Peer Assessment Design Elements». *Educational Psychology Review* 35 (1). <https://doi.org/10.1007/s10648-023-09723-7>.
- Anderson, Stuart, Pauline Allen, Stephen Peckham, und Nick Goodwin. 2008. «Asking the Right Questions: Scoping Studies in the Commissioning of Research on the Organisation and Delivery of Health Services». *Health research policy and systems* 6: 7. <https://doi.org/10.1186/1478-4505-6-7>.
- Arksey, Hilary, und Lisa O'Malley. 2005. «Scoping studies: towards a methodological framework». *International Journal of Social Research Methodology* 8 (1): 19–32. <https://doi.org/10.1080/1364557032000119616>.
- Barnes, Benjamin, Maike Buchmann, Rebekka Mumm, Enno Nowossadeck, Diana Peitz, Franziska Prütz, Benjamin Wachtler, und Antje Wienecke. 2022. «Evidenzsynthesen in Public Health: ein Überblick». [Evidence syntheses in public health: An overview]. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 175: 17–28. <https://doi.org/10.1016/j.zefq.2022.09.003>.
- Baumgartner, Peter. 2006. «Unterrichtsmethoden als Handlungsmuster – Vorarbeiten zu einer didaktischen Taxonomie für ELearning». In *DeLFI 2006, 4. e-Learning Fachtagung Informatik, 11.–14. September 2006*, herausgegeben von M. Mühlhäuser, G. Rößling, und R. Steinmetz, 51–62. Bonn: Gesellschaft für Informatik. <https://dl.gi.de/bitstream/handle/20.500.12116/15267/GI-Proceedings-87-4.pdf?sequence=1&isAllowed=y>.
- Baumgartner, Peter, und Ingrid Bergner. 2016. «Einige Feedback-Arten für Online-Lernen: Taxonomie und Realisierung von Feedback-Mustern für Multiple-Choice-Tests in Moodle». In *Digitale Medien: Zusammenarbeit in der Bildung*, herausgegeben von Josef Wachtler, Martin Ebner, Ortrun Gröbinger, Michael Kopp, Erwin Bratengeyer, Hans-Peter Steinbacher, Christian Freisleben-Teutscher, und Christine Kapper, 201–10. Medien in der Wissenschaft 71. Münster: Waxmann. <https://doi.org/10.25656/01:15793>.
- Baumgartner, Peter, Klaus Himpsl, und Silke Kleindienst. 2012. «Eine Taxonomie für E-Portfolios». https://portfolio.peter-baumgartner.net/files/pdf/2012/Baumgartner_2012_Eine%20Taxonomie%20fuer%20E-Portfolios.pdf.
- Bloom, Benjamin S. 1974. *Taxonomie von Lernzielen im kognitiven Bereich*. 4. Auflage. Weinheim, Basel: Beltz.
- Boud, David, und Phillip Dawson. 2023. «What feedback literate teachers do: an empirically-derived competency framework». *Assessment & Evaluation in Higher Education* 48 (2): 158–71. <https://doi.org/10.1080/02602938.2021.1910928>.
- Bredow, Carrie A., Patricia V. Roehling, Alexandra J. Knorp, und Andrea M. Sweet. 2021. «To Flip or Not to Flip? A Meta-Analysis of the Efficacy of Flipped Learning in Higher Education». *Review of Educational Research* 91 (6): 878–918. <https://doi.org/10.3102/00346543211019122>.

- Brück-Hübner, Annika. 2022. «Feedback digital: Besonderheiten und Herausforderungen von digitalen Feedbackprozessen in der Hochschullehre». Marburg, 2022. <https://doi.org/10.13140/RG.2.2.24765.74722>
- Brück-Hübner, Annika. 2023. «Was kennzeichnet «gutes» digitales Feedback? Eine empirische Studie zu den Gelingensbedingungen digitaler Feedbackprozesse in der Hochschullehre aus Studierendenperspektive.» In *Kompetenzen im digitalen Lehr- und Lernraum an Hochschulen*, herausgegeben von Katharina Hombach und Heike Rundnagel, 103–19. Bielefeld: wbv. <https://doi.org/10.3278/9783763973989>.
- Carless, David, und David Boud. 2018. «The development of student feedback literacy: Enabling uptake of feedback». *Assessment & Evaluation in Higher Education* 43 (8): 1315–25. <https://doi.org/10.1080/02602938.2018.1463354>.
- Chong, Sin Wang. 2021. «Reconsidering student feedback literacy from an ecological perspective». *Assessment & Evaluation in Higher Education* 46 (1): 92–104. <https://doi.org/10.1080/02602938.2020.1730765>.
- Chong, Sin Wang. 2022. «The role of feedback literacy in written corrective feedback research: From feedback information to feedback ecology». *Cogent Education* 9 (1): 1–13. <https://doi.org/10.1080/2331186X.2022.2082120>.
- Daweli, Talal Waleed. 2018. «Engaging Saudi EFL Students in Online Peer Review in a Saudi University Context». *Arab World English Journal* 9 (4): 270–80. <https://doi.org/10.24093/awej/vol9no4.20>.
- Eisend, M. 2004. «Metaanalyse – Einführung und kritische Diskussion. Diskussionsbeiträge des Fachbereichs Wirtschaftswissenschaften der Freien Universität Berlin». <https://www.econstor.eu/bitstream/10419/49908/1/668827637.pdf>.
- Elm, Erik von, Gerhard Schreiber, und Claudia Cornelia Haupt. 2019. «Methodische Anleitung für Scoping Reviews (JBI-Methodologie)». *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 143:1–7. <https://doi.org/10.1016/j.zefq.2019.05.004>.
- Farsani, Mohammad Amini, und Nahid Aghamohammadi. 2021. «Exploring Students' Engagement with Peer-and Teacher Written Feedback in an EFL Writing Course: A Multiple Case Study of Iranian Graduate Learners». *MEXTESOL Journal* 45 (4): 1–17.
- Fink, Arlene. 2014. *Conducting research literature reviews: From the internet to paper*. Fourth edition. Thousand Oaks: Sage.
- Früh, Werner. 2007. *Inhaltsanalyse: Theorie und Praxis*. 6. Auflage. Konstanz: UVK.
- Gläser, Jochen, und Grit Laudel. 2010. *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*. 4. Auflage. Wiesbaden: VS.
- 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>.
- Gravett, Karen. 2022. «Feedback literacies as sociomaterial practice». *Critical Studies in Education* 63 (2): 261–74. <https://doi.org/10.1080/17508487.2020.1747099>.

- Gravett, Karen, und David Carless. 2023. «Feedback literacy-as-event: relationality, space and temporality in feedback encounters». *Assessment & Evaluation in Higher Education*, 1–12. <https://doi.org/10.1080/02602938.2023.2189162>.
- Hattie, John. 2009. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge. <https://doi.org/10.4324/9780203887332>.
- Haughney, Kathryn, Shawnee Wakeman, und Laura Hart. 2020. «Quality of Feedback in Higher Education: A Review of Literature». *Education Sciences* 10 (3): 60. <https://doi.org/10.3390/educsci10030060>.
- Higgins, Steven. 2021. *Improving Learning: Meta-analysis of intervention research in education*. Cambridge: University Press.
- Hocker, Julian. 2021. *Entwicklung einer Domänen-Ontologie für qualitative Kategoriensysteme – QualiCO*, Nr. 25.10.2022. <https://hildok.bsz-bw.de/frontdoor/index/index/docId/1280>.
- Irwin, Bradley. 2019. «Enhancing Peer Feedback Practices through Screencasts in Blended Academic Writing Courses». *jatcalljournal* 15 (1): 43–59.
- Kanuka, Heather, und Cheryl Sadowski. 2020. «Reflective peer observations of university teaching: A Canadian case study». *Journal of University Teaching and Learning Practice* 17 (5): 165–78. <https://doi.org/10.53761/1.17.5.11>.
- Kuckartz, Udo. 2016. *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. 3. Auflage. Weinheim, Basel: Juventa.
- Lipnevich, Anastasiya A., und Ernesto Panadero. 2021. «A Review of Feedback Models and Theories: Descriptions, Definitions, and Conclusions». *Frontiers in Education* 6. <https://doi.org/10.3389/feduc.2021.720195>.
- Little, Tegan, Phillip Dawson, David Boud, und Joanna Tai. 2023. «Can students' feedback literacy be improved? A scoping review of interventions». *Assessment & Evaluation in Higher Education*, 1–14. <https://doi.org/10.1080/02602938.2023.2177613>.
- Mahoney, Paige, Susie Macfarlane, und Rola Ajjawi. 2019. «A qualitative synthesis of video feedback in higher education». *Teaching in Higher Education* 24 (2): 157–79. <https://doi.org/10.1080/13562517.2018.1471457>.
- Navarro Jover, José Manuel. 2021. «Auto-Feedback to Improve Academic Performance». *Journal of Technology and Science Education* 11 (1): 180–93. <https://doi.org/10.3926/jotse.1120>.
- Nieminen, Juuso Henrik, Joanna Tai, David Boud, und Michael Henderson. 2022. «Student agency in feedback: Beyond the individual». *Assessment & Evaluation in Higher Education* 47 (1): 95–108. <https://doi.org/10.1080/02602938.2021.1887080>.
- Panadero, Ernesto, und Anastasiya A. Lipnevich. 2022. «A review of feedback models and typologies: Towards an integrative model of feedback elements». *Educational Research Review* 35:100416. <https://doi.org/10.1016/j.edurev.2021.100416>.
- Rädiker, Stefan, und Udo Kuckartz. 2019. *Analyse qualitativer Daten mit MAXQDA: Text, Audio und Video*. Wiesbaden: Springer <https://doi.org/10.1007/978-3-658-22095-2>.

- Ritzhaupt, Albert D., Rui Huang, Max Sommer, Jiawen Zhu, Anita Stephen, Natercia Valle, John Hampton, und Jingwei Li. 2021. «A meta-analysis on the influence of gamification in formal educational settings on affective and behavioral outcomes». *Education Tech Research Dev* 69 (5): 2493–2522. <https://doi.org/10.1007/s11423-021-10036-1>.
- Ruiz-Primo, Maria Araceli, und Min Li. 2013. «Examining Formative Feedback in the Classroom Context: New Research Perspectives». In *SAGE Handbook of Research on Classroom Assessment*, herausgegeben von James McMillan, 215–32. California: SAGE Publications, Inc. <https://doi.org/10.4135/9781452218649.n13>.
- Schluer, Jennifer. 2022. *Digital feedback methods*. Tübingen: Narr Francke Attempto. <https://doi.org/10.24053/9783823395324>.
- Strauss, Anselm, und Juliet Corbin. 1996. *Grounded theory: Grundlagen qualitativer Sozialforschung*. Weinheim: Psychologie Verlags Union.
- Tai, Joanna, Margaret Bearman, Karen Gravett, und Elizabeth Molloy. 2021. «Exploring the notion of teacher feedback literacies through the theory of practice architectures». *Assessment & Evaluation in Higher Education* 62 (1): 1–13. <https://doi.org/10.1080/02602938.2021.1948967>.
- Tochon, François. 2013. «A Brief History of Video Feedback and its Role in Foreign Language Education». *CALICO Journal* 25 (3): 420–35. <https://doi.org/10.1558/cj.v25i3.420-435>.
- Topping, Keith. 1998. «Peer Assessment Between Students in Colleges and Universities». *Review of Educational Research* 68 (3): 249–76. <https://doi.org/10.3102/00346543068003249>.
- Vencille, Annie, Tina Kosteci, Bernardine Lynch, Elisabeth Santhanam, und Anthony Whitty. 2021. «Formalizing feedback in work-integrated learning partnerships : Opportunities for collaboration.». *International Journal of Work-Integrated Learning* 22 (1): 17–22. https://www.ijwil.org/files/IJWIL_22_1_17_23.pdf.
- Winstone, Naomi E., Robert A. Nash, Michael Parker, und James Rowntree. 2017. «Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes». *Educational Psychologist* 52 (1): 17–37. <https://doi.org/10.1080/00461520.2016.1207538>.
- Wisniewski, Benedikt, Klaus Zierer, und John Hattie. 2020. «The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research». *Frontiers in psychology* 10:3087. <https://doi.org/10.3389/fpsyg.2019.03087>.
- Witeck, Gabriela R., Anabela C. Alves, und Mariana H. Bernardo. 2021. «Bloom Taxonomy, Serious Games and Lean Learning: What Do These Topics Have in Common?». In *Learning in the Digital Era: 7th European Lean Educator Conference, ELEC 2021, Trondheim, Norway, October 25–27, 2021, Proceedings*. Bd. 610, herausgegeben von Daryl J. Powell, Erlend Alfnes, Marte D. Q. Holmemo, und Eivind Reke, 308–16. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-92934-3_31.
- Zhan, Ying. 2022. «Developing and validating a student feedback literacy scale». *Assessment & Evaluation in Higher Education* 47 (7): 1087–1100. <https://doi.org/10.1080/02602938.2021.2001430>.

Anhang: Liste der analysierten Studien

- Abri, Abdullah Al. 2021. «Exploring EFL Learners' Comments on Web-Based Peer Feedback: Local and Global Revisions». *English Language Teaching* 14 (6): 114–24. <https://doi.org/10.5539/elt.v14n6p114>.
- AbuSa'aleek, Atef O., und Mohammad Shariq. 2021. «Innovative Practices in Instructor E-feedback: A Case Study of E-feedback given in Three Linguistic Courses during the COVID 19 Pandemic». *AWEJ* 1: 183–98. <https://doi.org/10.24093/awej/covid.14>.
- Alharbi, Mohammed Abdullah. 2021. «Impact of teacher written vs. audio feedback on EFL undergraduates' writing». *CJES* 16 (3): 1141–53. <https://doi.org/10.18844/cjes.v16i3.5836>.
- Al-Hazzani, Nahla, und Sultan Altalhab. 2018. «Can Explicit Written Corrective Feedback Develop Grammatical and Lexical Accuracy of Saudi EFL Learners?». *International Journal of Education and Literacy Studies* 6 (4): 16–24. <https://doi.org/10.7575/aiac.ijels.v.6n.4p.16>.
- Alnasser, Suliman Mohammed Nasser. 2018. «Exploring Student-Writers' Views on Replacing Teacher Feedback with Peer Feedback and Computer-Based Feedback». *Arab World English Journal* 9 (3): 345–66. <https://doi.org/10.24093/awej/vol9no3.23>.
- Al-Wadi, Hasan Mohsen. 2018. «Facilitating in-Service English Language Teacher Trainees' Supervision Through Written Feedback: Action Research». *International Education Studies* 11 (9): 1–11. <https://doi.org/10.5539/ies.v11n9p1>.
- Andrade, Maureen, Ronald Mellado Miller, und Morgan Ogden. 2020. «Teamwork for Business Majors – the Impact of Peer Evaluation». *e-Journal of Business Education & Scholarship of Teaching* 14 (2): 1–18.
- Aridah, Aridah, und Weningtyas Parama Iswari. 2021. «The Effect of Indirect Feedback on Students' Writing Performance Across Different Learning Strategies». *CJES* 16 (3): 1021–35. <https://doi.org/10.18844/cjes.v16i3.5820>.
- Au, Hong Yu Connie. 2020. «An Analysis of the Effect of Peer and Teacher Feedback on EFL Learners' Oral Performances and Speaking Self-Efficacy Levels». *International Online Journal of Education and Teaching (IOJET)* 7 (4): 1453–68. <http://iojet.org/index.php/IOJET/article/view/895>.
- Barbosa, Ana, und Isabel Vale. 2021. «You've Got Mail! – Written Communication and Feedback in Mathematics». *International Journal on Social and Education Sciences* 3 (3): 563–75. <https://doi.org/10.46328/ijonses.234>.
- Bayrak, Fatma, und Halil Yurdugül. 2018. «Regulation of Cognition as a Mediator in the Relationship Between Knowledge of Cognition and Perceived Self-Intervention». *European Journal of Open, Distance and E-Learning* 21 (2): 1–13. <https://doi.org/10.2478/eurodl-2018-0003>.
- Beccaria, Lisa, Megan Kek, und Henk Huijser. 2019. «Using 'Just in Time' Online Feedback to Improve First Year Undergraduate Nursing Students' Essay Writing Performance». *Journal of University Teaching and Learning Practice* 16 (4): 91–109. <https://doi.org/10.53761/1.16.4.7>.

- Bent, Marije, Erick Velazquez-Godinez, und Frank de Jong. 2021. «Becoming an Expert Teacher: Assessing Expertise Growth in Peer Feedback Video Recordings by Lexical Analysis». *Education Sciences* 11 (11): 665. <https://doi.org/10.3390/educsci11110665>.
- Bi, Zhou, Tan Mengyu, Yuan Dying, und Pan Liandi. 2019. «Investigating the Validity of Different Peer Groupings in the Assessment of English Writings». *International Education Studies* 12 (12): 55–61. <https://doi.org/10.5539/ies.v12n12p55>.
- Bongiovanni, Ivano, und Dayana Balgabekova. 2021. «Ask Me If I Am Engaged: A Design-Led Approach to Collect Student Feedback on Their University Experience». *Design and Technology Education* 26 (1): 89–117.
- Bostanci, Hanife Bensen, und Fatma Sengul. 2018. «Who Is the Most Effective Agent When Giving Indirect Written Corrective Feedback?». *EJER*, Nr. 76: 73–92. <https://doi.org/10.14689/ejer.2018.76.4>.
- Britton, Emma R., und Theresa Y. Austin. 2020. «That’s Just How We Say It: Understanding L2 Student Writers’ Responses to Written and Negotiated Corrective Feedback Through Critical Incidents». *TESL Canada* 37 (2): 103–27. <https://doi.org/10.18806/tesl.v37i2.1332>.
- Byrne, Virginia L., und Alice E. Donlan. 2020. «Presenting a Validated Mid-Semester Evaluation of College Teaching to Improve Online Teaching». *OLJ* 24 (2): 94–110. <https://doi.org/10.24059/olj.v24i2.2126>.
- Çakiroglu, Ünal, Samet Atabas, Dogukan Sariyalcinkaya, und Ibrahim Enes Öner. 2020. «Learning programming online: Influences of various types of feedback on programming performances». *IJCSES* 3 (3): 3–18. <https://doi.org/10.21585/ijcses.v3i3.57>.
- Cañabate, Dolors, Lluís Nogué, Teresa Serra, und Jordi Colomer. 2019. «Supportive Peer Feedback in Tertiary Education: Analysis of Pre-Service Teachers’ Perceptions». *Education Sciences* 9 (4): 1–11. <https://doi.org/10.3390/educsci9040280>.
- Canals, Laia, Gisela Granena, Yucel Yilmaz, und Aleksandra Malicka. 2020. «Second Language Learners’ and Teachers’ Perceptions of Delayed Immediate Corrective Feedback in an Asynchronous Online Setting: An Exploratory Study». *TESL Canada Journal* 37 (2): 181–209. <https://doi.org/10.18806/tesl.v37i2.1336>.
- Carnegie, Jacqueline, und Joanne Savory. 2021. «Students Work in Groups to Create and Peer-Evaluate Newsletters Pertaining to Current Health-Related Topics». *HAPS Educator* 25 (2): 124–31. <https://doi.org/10.21692/haps.2021.010>.
- Carpenter, Tara S., Lisa Carter Beall, und Linda C. Hodges. 2020. «Using the LMS for Exam Wrapper Feedback to Prompt Metacognitive Awareness in Large Courses». *Journal of Teaching and Learning with Technology* 9 (1): 79–91. <https://doi.org/10.14434/jotlt.v9i1.29156>.
- Cavaleri, Michelle, Satomi Kawaguchi, Bruno Di Biase, und Clare Power. 2019. «How recorded audio-visual feedback can improve academic language support». *Journal of University Teaching and Learning Practice* 16 (4): 71–90. <https://doi.org/10.53761/1.16.4.6>.
- Chand Dayal, Hem, und Reema Alpana. 2020. «Secondary Pre-Service Teachers’ Reflections on Their Micro-Teaching: Feedback and Self-Evaluation». *Waikato Journal of Education* 25 (1): 73–83. <https://doi.org/10.15663/wje.v25i0.686>.

- Chen, Jialiang. 2021. «Research on the Effect of Peer Feedback Training in English Writing Teaching – A Case Study of Students in Business English Major». *ELT* 14 (6): 12. <https://doi.org/10.5539/elt.v14n6p12>.
- Chen, Wenling. 2018. «The Effects of Corrective Feedback Strategies on English Majors' Writing». *ELT* 11 (11): 55–64. <https://doi.org/10.5539/elt.v11n11p55>.
- Chuaphalakit, Ken, Bhornsawan Inpin, und Prarthana Coffin. 2019. «A Study of the Quality of Feedback Via the Google Classroom-mediated-Anonymous Online Peer Feedback Activity in a Thai EFL Writing Classroom». *International Journal of Progressive Education* 15 (5): 103–18. <https://doi.org/10.29329/ijpe.2019.212.8>.
- Cleary, Ann, Brid Delahunt, Claire Fox, Moira Maguire, Lorna O'Connor, und Jamie Ward. 2018. «Promoting Student Engagement with Academic Literacy Feedback: An Institute Wide Initiative». *Practitioner Research in Higher Education* 11 (1): 101–9.
- Cobbold, Christian, und Louise Wright. 2021. «Use of Formative Feedback to Enhance Summative Performance». *INT J INSTRUCTION* 6 (1): 109–16. <https://doi.org/10.29333/aje.2021.619a>.
- Colpitts, Bradley D. F. 2019. «A comparison of computer-mediated peer corrective feedback between high and low-proficiency learners in a Japanese EFL writing classroom». *JALT CALL Journal* 15 (2): 23–39.
- Cutting, Chelsea, und Kevin Larkin. 2021. «The Impact of Weekly Formative Video Feedback on Pre- Service Teachers' Experiences in Online Mathematics Education». *Mathematics Teacher Education and Development* 23 (1): 74–90.
- Daweli, Talal Waleed. 2018. «Engaging Saudi EFL Students in Online Peer Review in a Saudi University Context». *Arab World English Journal* 9 (4): 270–80. <https://doi.org/10.24093/awej/vol9no4.20>.
- Deale, Cynthia S. 2020. «A Team Approach to Using Student Feedback to Enhance Teaching and Learning». *ij-sotl* 14 (2): Art. 6. <https://doi.org/10.20429/ijstl.2020.140206>.
- Demonacos, Constantinos, Steven Ellis, und Jill Barber. 2019. «Student Peer Assessment Using Adaptive Comparative Judgment: Grading Accuracy Versus Quality of Feedback». *Practitioner Research in Higher Education* 12 (1): 50–60.
- Dennis, Diane, Anne Furness, Dawn Hall-Bibb, und Shylie Mackintosh. 2020. «Teaching Physiotherapy Students to Provide Feedback Using Simulation». *Journal of Peer Learning* 13: 36–47.
- Ducasse, Ana Maria, und Kathryn Hill. 2019. «Developing Student Feedback Literacy Using Educational Technology and the Reflective Feedback Conversation». *Practitioner Research in Higher Education* 12 (1): 24–37.
- Eather, Narelle, Nicholas Riley, Drew Miller, und Scott Imig. 2019. «Evaluating the Impact of Two Dialogical Feedback Methods for Improving Pre-Service Teacher's Perceived Confidence and Competence to Teach Physical Education Within Authentic Learning Environments». *Journal of Education and Training Studies* 7 (8): 32–46. <https://doi.org/10.11114/jets.v7i8.4053>.

- El Mortaji, Latifa. 2022. «Public Speaking and Online Peer Feedback in a Blended Learning EFL Course Environment: Students' Perceptions». *ELT* 15 (2): 31–49. <https://doi.org/10.5539/elt.v15n2p31>.
- Elumalai, Kesavan Vadakalur. 2019. «Teacher Constructed Corrective Feedback Enhancing Students Writing Skills in EFL Classroom». *Advances in Language and Literary Studies* 10 (5): 103–6. <https://doi.org/10.7575/aiac.all.v.10n.5p.103>.
- Ene, Estela, und Jia Yao. 2021. «How Does that Make You Feel: Students' Affective Engagement with Feedback». *LTRQ* 25:66–83. <https://doi.org/10.32038/ltrq.2021.25.04>.
- Farsani, Mohammad Amini, und Nahid Aghamohammadi. 2021. «Exploring Students' Engagement with Peer-and Teacher Written Feedback in an EFL Writing Course: A Multiple Case Study of Iranian Graduate Learners». *MEXTESOL Journal* 45 (4): 1–17.
- Fernández-Michels, Pedro, und Laia Canals Fornons. 2021. «Learner engagement with corrective feedback using think-aloud protocols». *The JALT CALL Journal* 17 (3): 203–32. <https://doi.org/10.29140/jaltcall.v17n3.461>.
- Flora, Raja, und Mahpul Patuan. 2020. «Discovery Learning Strategy: Integrating Think-Pair-Share and Teacher's Corrective Feedback to Enhance Students' Writing Language Accuracy». *International Journal of Education and Practice* 8 (4): 733–45. <https://doi.org/10.18488/journal.61.2020.84.733.745>.
- González-Gómez, David, Jin Su Jeong, und Florentina Cañada-Cañada. 2020. «Examining the Effect of an Online Formative Assessment Tool (OFAT) of Students' Motivation and Achievement for a University Science Education». *Journal of Baltic Science Education* 19 (3): 401–14. <https://doi.org/10.33225/jbse/20.19.401>.
- Grier, Donovan, Suzanne F. Lindt, und Stacia C. Miller. 2021. «Formative Assessment with Game-based Technology». *International Journal of Technology in Education and Science* 5 (2): 193–202. <https://doi.org/10.46328/ijtes.97>.
- Gündüz Kartal, Mine, und Derin Atay. 2019. «Corrective Feedback on Writing in EFL Context: Comparison of Two Approaches». *Eurasian Journal of Applied Linguistics* 5 (3): 385–401. <https://doi.org/10.32601/ejal.651390>.
- Gürkan, Serkan. 2018. «The Effect of Feedback on Instructional Behaviours of Pre-Service Teacher Education». *Universal Journal of Educational Research* 6 (5): 1084–93. <https://doi.org/10.13189/ujer.2018.060530>.
- Hart, Laura C., und Christine Robinson. 2019. «From the Mouths of Babes: Using a Charrette Model to Assess Student Learning and Engage External Stakeholders». *Research & Practice in Assessment* 14: 18–29.
- Havis, Leanne R. 2019. «The Exam Autopsy: An Integrated Post-Exam Assessment Model». *International Journal for the Scholarship of Teaching and Learning* 13 (1). <https://doi.org/10.20429/ijstl.2019.130104>.
- Hentasmaka, Daning, und Bambang Yudi Cahyono. 2021. «Peer Feedback Uptakes and Outcomes Across EFL Students' Proficiency Levels: A Study at Tertiary Education in Indonesia». *International Journal of Instruction* 14 (3): 271–86. <https://doi.org/10.29333/iji.2021.14316a>.

- Hill, Jennifer, Kathy Berlin, Julia Choate, Lisa Cravens-Brown, Lisa McKendrick-Calder, und Susan Smith. 2021. «Can Relational Feed-Forward Enhance Students' Cognitive and Affective Responses to Assessment?». *Teaching & Learning Inquiry* 9 (2): 1–21. <https://doi.org/10.20343/teachlearninqu.9.2.18>.
- Ho, Pham Vu Phi, Luong Thi Kim Phung, Tran Thi Thuy Oanh, und Nguyen Quang Giao. 2020. «Should Peer E-Comments Replace Traditional Peer Comments?». *International Journal of Instruction* 13 (1): 295–314. <https://doi.org/10.29333/iji.2020.13120a>.
- Hung, Do Minh, und Tran Thi Thuy Diem. 2020. «Does Conferencing Feedback Improve Non-English Majors' Paragraph-Writing? A Case Study in Vietnam». *Arab World English Journal* 11 (4): 78–95. <https://doi.org/10.24093/awej/vol11no4.6>.
- Hussain, Sajjad, Kelum A. A. Gamage, Wasim Ahmad, und Muhammad A. Imran. 2019. «Assessment and Feedback for Large Classes in Transnational Engineering Education: Student-Staff Partnership-Based Innovative Approach». *Education Sciences* 9 (3): 221. <https://doi.org/10.3390/educsci9030221>.
- Iksan, Halimatussaadiah, und Huzaina Abdul Halim. 2018. «The Effect of E-Feedback via Wikis on ESL Students' L2 Writing Anxiety Level». *Malaysian Online Journal of Educational Sciences* 6 (3): 30–48.
- Iraj, Hamideh, Anthea Fudge, Huda Khan, Margaret Faulkner, Abelardo Pardo, und Vitomir Kovanović. 2021. «Narrowing the Feedback Gap: Examining Student Engagement with Personalized and Actionable Feedback Messages». *JLA* 8 (3): 101–16. <https://doi.org/10.18608/jla.2021.7184>.
- Irwin, Bradley. 2019. «Enhancing Peer Feedback Practices through Screencasts in Blended Academic Writing Courses». *jatcalljournal* 15 (1): 43–59.
- Isnawati, Ida, Gunadi Harry Sulisty, Utami Widiati, und Nunung Suryati. 2019. «Impacts of Teacher-Written Corrective Feedback with Teacher-Student Conference on Students' Revision». *International Journal of Instruction* 12 (1): 669–84. <https://doi.org/10.29333/iji.2019.12143a>.
- Jang, Sung-Soo. 2020. «The Efficacy of Different Types of Metalinguistic Information in L2 Written Corrective Feedback». *English Teaching* 75 (4): 33–56. <https://doi.org/10.15858/engtea.75.4.202012.33>.
- Jeon, Eun Young. 2018. «The Effect of Learner-Centered EFL Writing Instruction on Korean University Students' Writing Anxiety and Perception». *TESOL International Journal* 13 (3): 100–12.
- Kahraman, Emine, und Ozlem Koray. 2020. «The Use of Standardized Feedback for Teaching Material Preparation: The Opinions of Preservice Science Teachers». *Eurasian Journal of Educational Research* 20 (90): 1–20. <https://doi.org/10.14689/ejer.2020.90.5>.
- Kennette, Lynne N., und Dawn McGuckin. 2018. «Using the Immediate Feedback Assessment Technique (IFAT) for Non-Assessments: Student Perceptions and Performance». *Psychology Teaching Review* 24 (2): 64–69.

- Ketchum, Cheri, Daria S. LaFave, Chelsey Yeats, Elaine Phompheng, und James H. Hardy. 2020. «Video-based Feedback on Student Work: An Investigation into the Instructor Experience, Workload, and Student Evaluations». *Online Learning* 24 (3). <https://doi.org/10.24059/olj.v24i3.2194>.
- Khadawardi, Hanadi Abdulrahman. 2020. «The Effect of Implicit Corrective Feedback on English Writing of International Second Language Learners». *ELT* 14 (1): 123. <https://doi.org/10.5539/elt.v14n1p123>.
- Kim, Jeongyeon, und Victoria Kim. 2021. «Rediscovering Feedback and Experiential Learning in the English-Medium Instruction Classroom». *JUTLP* 18 (4): 286–305. <https://doi.org/10.53761/1.18.4.19>.
- Kim, Ji Hyun. 2018. «Recasts and Prompts in Dyadic Interaction: Explicitness of Feedback and Learner Proficiency». *English Teaching* 73 (4): 3–28. <https://doi.org/10.15858/engtea.73.4.201812.3>.
- Kim, Ji Hyun. 2019. «Relative Effects of Direct Focused and Unfocused WCF on the Accuracy Development of Two Language Forms». *English Teaching* 74 (4): 29–50. <https://doi.org/10.15858/engtea.74.4.201912.29>.
- Kim, Victoria. 2018. «Technology-Enhanced Feedback on Student Writing in the English-Medium Instruction Classroom». *English Teaching* 73 (4): 29–53. <https://doi.org/10.15858/engtea.73.4.201812.29>.
- Kita, Yoko. 2021. «Using Forms Quizzes to Enhance Students' Motivation in Language Learning at a Japanese University». *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)* 5 (2): 219–35. <https://doi.org/10.21093/ijeltal.v5i2.661>.
- Klash, Erin F., und Kate D. Simmons. 2020. «The Virtual Avatar Lab (VAL): Tapping into Virtual Live Environments to Practice Classroom Feedback Conversations». *Journal of Curriculum and Teaching* 9 (3): 149–60. <https://doi.org/10.5430/jct.v9n3p149>.
- Kusumaningrum, Shirly Rizki, Bambang Yudi Cahyono, und Johannes Ananto Prayogo. 2019. «The Effect of Different Types of Peer Feedback Provision on EFL Students' Writing Performance». *International Journal of Instruction* 12 (1): 213–24. <https://doi.org/10.29333/iji.2019.12114a>.
- Kuyyogsuy, Sirikarn. 2019. «Promoting Peer Feedback in Developing Students' English Writing Ability in L2 Writing Class». *International Education Studies* 12 (9): 76–90. <https://doi.org/10.5539/ies.v12n9p76>.
- Lee, Jeong-Won, und Kyeong-Ok Yoon. 2020. «Effects of Written Corrective Feedback on the Use of the English Indefinite Article in EFL Learners' Writing». *English Teaching* 75 (2): 21–40. <https://doi.org/10.15858/engtea.75.2.202006.21>.
- Lei, Zhibin, und Issra Pramoolsook. 2020. «Written Supervisory Feedback Strategies on Bachelor's Theses: Chinese EFL Supervisors' Beliefs and Practices». *rEFlections* 27 (2): 162–88.
- Lenkaitis, Chesla Ann, Shannon M. Hilliker, und Kayla Roumeliotis. 2020. «Teacher Candidate Reflection and Development Through Virtual Exchange». *IAFOR Journal of Education* 8 (2): 125–39. <https://doi.org/10.22492/ije.8.2.07>.

- Lewis, Sally, Georgina Heath, Lisa Lim, und Rosie Roberts. 2021. «I'm Not a Number, I'm Someone to Them»: Supporting Commencing University Students' Through Technology-Mediated Personalised Communication». *Student Success* 12 (1): 24–34. <https://doi.org/10.5204/ssj.1623>.
- Lyman, Michael, und Christopher Keyes. 2019. «Peer-Supported Writing in Graduate Research Courses: A Mixed Methods Assessment». *International Journal of Teaching and Learning in Higher Education* 31 (1): 11–20.
- Mahmoudi, Farzaneh, und Cemile Buğra. 2020. «The Effects of Using Rubrics and Face to Face Feedback in Teaching Writing Skill in Higher Education». *International Online Journal of Education and Teaching (IOJET)* 7 (1): 150–58.
- Majid, Nadia, und Muhammad Islam. 2021. «Effectiveness of Peer Assessment and Peer Feedback in Pakistani Context: A Case of University of the Punjab». *Bulletin of Education and Research* 43 (2): 101–22.
- Marshall, David T., Savanna Love, und LaRon Scott. 2020. «It's Not Like He Was Being a Robot: Student Perceptions of Video-Based Writing Feedback in Online Graduate Coursework». *International Journal for the Scholarship of Teaching and Learning* 14 (1). <https://doi.org/10.20429/ijstl.2020.140110>.
- McCuin, Deborah. 2021. «Online Discussion Threads: An Exercise in Civil Discourse». *Educational Research* 32 (1): 30–42.
- Metcalfe, Janet, und Teal S. Eich. 2019. «Memory and Truth: Correcting Errors with True Feedback Versus Overwriting Correct Answers with Errors». *Cognitive Research: Principles and Implications* 4 (1): 1–18. <https://doi.org/10.1186/s41235-019-0153-8>.
- Meza, Angélica, Ingrid Rodríguez, und Lorena Caviedes. 2021. «Fostering EFL Preservice Teachers' Academic Writing Skills Through Reflective Learning». *Profile: Issues in Teachers' Professional Development* 23 (1): 89–106. <https://doi.org/10.15446/profile.v23n1.85145>.
- Mohsen, Mohammed Ali, und Abdulaziz Alshahrani. 2019. «The Effectiveness of Using a Hybrid Mode of Automated Writing Evaluation System on EFL Students' Writing». *Teaching English with Technology* 19 (1): 118–31.
- Moxon, Simon. 2021. «Exploring the Effects of Automated Pronunciation Evaluation on L2 Students in Thailand». *IAFOR Journal of Education* 9 (3): 41–57. <https://doi.org/10.22492/ije.9.3.03>.
- Nakagawa, Hiroyo, und Ambrose Leung. 2020. «The Effects of Implicit Learning on Japanese EFL Junior College Students' Writing». *International Journal of Instruction* 13 (1): 637–52. <https://doi.org/10.29333/iji.2020.13141a>.
- Navarro Jover, José Manuel. 2021. «Auto-Feedback to Improve Academic Performance». *Journal of Technology and Science Education* 11 (1): 180–93. <https://doi.org/10.3926/jotse.1120>.
- Nguyen, Thi Thuy Loan. 2018. «The Effect of Combined Peer-Teacher Feedback on Thai Students' Writing Accuracy». *Iranian Journal of Language Teaching Research* 6 (2): 117–32.
- Nicholson, Dawn Theresa. 2018. «Enhancing Student Engagement Through Online Portfolio Assessment». *Practitioner Research in Higher Education* 11 (1): 15–31.

- Nikolaeva, Sofiya, und Tetiana Korol. 2021. «Formative Assessment in the Translation Classroom: Closing a Feedback Loop». *International Journal of Evaluation and Research in Education (IJERE)* 10 (2): 738–46. <https://doi.org/10.11591/ijere.v10i2.21274>.
- Noordin, Nooreen, und Laleh Khojasteh. 2021. «The Effects of Electronic Feedback on Medical University Students' Writing Performance». *IJHE* 10 (4): 124. <https://doi.org/10.5430/ijhe.v10n4p124>.
- Okyar, Hatice. 2019. «Training Students in Peer Interaction and Peer Feedback to Develop Competence in L2 Forms». *PASAA* 58: 62–94.
- Parra G., Lorena, und Ximena Calero S. 2019. «Automated Writing Evaluation Tools in the Improvement of the Writing Skill». *International Journal of Instruction* 12 (2): 209–26. <https://doi.org/10.29333/iji.2019.12214a>.
- Pearson, John. 2018. «Engaging Practical Students Through Audio Feedback». *Practitioner Research in Higher Education* 11 (1): 87–94.
- Pereira, Íris Susana Pires, Eva Lopes Fernandes, und Maria Assunção Flores. 2021. «Teacher Education during the COVID-19 Lockdown: Insights from a Formative Intervention Approach Involving Online Feedback». *Education Sciences* 11 (8): 400. <https://doi.org/10.3390/educsci11080400>.
- Perks, Bradley, Bradley Colpitts, und Matthew Michaud. 2021. «The Role of Individual Preferences in the Efficacy of Written Corrective Feedback in an English for Academic Purposes Writing Course». *Australian Journal of Teacher Education* 46 (10): Art. 1. <https://doi.org/10.14221/ajte.2021v46n10.1>.
- Persaud, Vineeta, und Rita Persaud. 2019. «Increasing Student Interactivity Using a Think-Pair-Share Model with a Web-Based Student Response System in a Large Lecture Class in Guyana». *International Journal of Education and Development using Information and Communication Technology (IJEDICT)* 15 (2): 117–31.
- Pow, Wai Cheong Jacky, und Kwok Hung Lai. 2021. «Enhancing the quality of student teachers' reflective teaching practice through building a virtual learning community». *Journal of Global Education and Research* 5 (1): 54–71. <https://doi.org/10.5038/2577-509X.5.1.1088>.
- Quierke-Bolt, Nigel. 2020. «The effects of peer feedback, within an 'assessment as learning' approach, on the learning and development of student-teachers». *Practitioner Research in Higher Education Journal* 13 (2).
- Rastgou, Ali, Neomy Storch, und Ute Knoch. 2020. «The Effect of Sustained Teacher Feedback on CAF, Content and Organization in EFL Writing». *Iranian Journal of Language Teaching Research* 8 (2). <https://doi.org/10.30466/ijltr.2020.120888>.
- Rawle, Fiona, Mindy Thuna, Ting Zhao, und Michael Kaler. 2018. «Audio Feedback: Student and Teaching Assistant Perspectives on an Alternative Mode of Feedback for Written Assignments». *The Canadian Journal for the Scholarship of Teaching and Learning* 9 (2): Art. 2. <https://doi.org/10.5206/cjsotl-rcacea.2018.2.2>.
- Reinholz, Daniel L. 2018. «Three Approaches to Focusing Peer Feedback». *ij-sotl* 12 (2): Art. 10. <https://doi.org/10.20429/ij-sotl.2018.120210>.

- Richards, Laura. 2020. «Head in the Clouds – Innovating Classroom Practice Using Online File Sharing for Collaboration and Feedback». In *Innovative language teaching and learning at university: treasuring languages*, herausgegeben von Alessia Plutino, Kate Borthwick, und Erika Corradini, 33–39: Research-publishing.net. <https://research-publishing.net/manuscript?10.14705/rpnet.2020.40.1063>. <https://doi.org/10.14705/rpnet.2020.40.1063>.
- Riyanto, Riyanto, und Diah Aryulina. 2020. «Implementing Immediate Feedback with Unlimited Plus Bonus Points to Increase College Student Learning Motivation and Achievement». *International Journal of Instruction* 13 (3): 387–400. <https://doi.org/10.29333/iji.2020.13327a>.
- Rodas, Elisabeth L., und Laura Colombo. 2021. «Self-Managed Peer Writing Groups for the Development of EFL Literacy Practices». *The Electronic Journal for English as a Second Language* 24 (4): 1–22.
- Sabbaghan, Soroush, Murray Peglar, und M. Gregory Tweedie. 2019. «The Multimodal Effects of Voice-Based Asynchronous Technology-Mediated Communication on EAP Speaking Performance». *TESL Canada* 36 (3): 82–109. <https://doi.org/10.18806/tesl.v36i3.1322>.
- Şahin, Muhittin. 2019. «Classroom Response Systems as a Formative Assessment Tool: Investigation into Students' Perceived Usefulness and Behavioural Intention». *International Journal of Assessment Tools in Education* 6 (4): 693–705. <https://doi.org/10.21449/ijate.576249>.
- Said, Mohamed Mohamed Tolba, Vasudeva Rao Aravind, Debra Ferdinand James, und Krishnan Umachandran. 2020. «Dissecting Assessment: A Paradigm Shift Towards Technology-Enhanced Assessments». *World Journal on Educational Technology: Current Issues* 11 (2): 162–70. <https://doi.org/10.18844/wjet.v11i2.4256>.
- Saiful, Saiful, Wardani Dwi Wihastyanang, Gunadi Harry Sulisty, und Nur Mukminatien. 2019. «Confronting Effect of Online Teacher and Peer Feedback on the Students' Writing Performance». *Problems of Education in the 21st Century* 77 (5): 650–66. <https://doi.org/10.33225/pec/19.77.650>.
- Sambell, Ros, Amanda Devine, Johnny Lo, und Tanya Lawlis. 2020. «Work-Integrated Learning Builds Student Identification of Employability Skills: Utilizing a Food Literacy Education Strategy». *International Journal of Work-Integrated Learning* 21 (1): 63–87.
- Serrano-Aguilera, Juan Jose, Alicia Tocino, Sergio Fortes, Cristian Martín, Pere Mercadé-Melé, Rafael Moreno-Sáez, Antonio Muñoz, Sara Palomo-Hierro, und Antoni Torres. 2021. «Using Peer Review for Student Performance Enhancement: Experiences in a Multidisciplinary Higher Education Setting». *Education Sciences* 11: Art. 71. <https://doi.org/10.3390/educsci11020071>.
- Sevcikova, Beata Lewis. 2018. «Online Open-Source Writing Aid as a Pedagogical Tool». *ELT* 11 (8): 126–42. <https://doi.org/10.5539/elt.v11n8p126>.
- Simpson, Rachel, und Catherine Reading. 2019. «Developing as a Student Peer Reviewer: Enhancing Students' Graduate Attributes of Producing Evaluative Judgements and Oral Feedback Communication». *Practitioner Research in Higher Education* 12 (1): 38–49.

- Smith, Matthew, und Cassie Lowe. 2021. «DIY assessment feedback: Building engagement, trust and transparency in the feedback process». *JUTLP* 18 (3): 121–35. <https://doi.org/10.53761/1.18.3.9>.
- Solhi, Mehdi, und İlknur Eginli. 2020. «The effect of recorded oral feedback on EFL learners' writing». *Dil ve Dilbilimi Çalışmaları Dergisi* 16 (1): 1–13. <https://doi.org/10.17263/jlls.712628>.
- Sükür, Murat, und Fatma Esra Demircan. 2020. «The Effects of Intensive and Extensive Recasts on Learning the Third Person Singular -s». *Novitas-ROYAL (Research on Youth and Language)* 14 (2): 58–75.
- Tatsanajamsuk, Pratabjai, und Saksit Saengboon. 2021. «The Roles of Indirect Asynchronous Electronic Feedback and Direct Corrective Feedback in Improving Students' L2 Writing: A Multiple Case Study of Thai Undergraduate Students». *LEARN Journal: Language Education and Acquisition Research Network* 14 (2): 397–426.
- Thompson, David, und Nicky Meer. 2021. «Blurring the Boundaries of Formative and Summative Assessment for Impact on Learning». *Practitioner Research in Higher Education* 14 (1): 28–40.
- Torres, J. T. 2022. «Feedback as Open-Ended Conversation: Inviting Students to Co-Regulate and Metacognitively Reflect During Assessment». *Journal of the Scholarship of Teaching and Learning* 22 (1). <https://doi.org/10.14434/josotl.v22i1.31232>.
- Tran, Dung. 2020. «ESL Students' Comments on Teacher's Written Corrective Feedback in a Freshman Composition Class». *MEXTESOL Journal* 44 (4): 1–13.
- Tur, Gemma, Santos Urbina, und Dolores Forteza. 2019. «Rubric-based Formative Assessment in Process Eportfolio: Towards Self-regulated Learning». *Digital Education Review* (35): 18–35.
- Uymaz, Eren. 2019. «The Effects of Peer Feedback on the Essay Writing Performances of EFL Students». *International Journal of Curriculum and Instruction* 11 (2): 20–37.
- Valizadeh, Mohammadreza. 2020. «The Effect of Comprehensive Written Corrective Feedback on EFL Learners' Written Syntactic Accuracy». *Advances in Language and Literary Studies* 11 (1): 17–26. <https://doi.org/10.7575/aiac.all.v.11n.1p.17>.
- Wettstein, Stephanie G. 2018. «Self-Paced, Active Problem-Solving Using Immediate Feedback (IF-AT; Scratch-Off) Forms in Large Classes». *Advances in Engineering Education* 6 (3): 1–18.
- Wihastyanang, Wardani Dwi, Shirly Rizki Kusumaningrum, M. Adnan Latief, und Bambang Yudi Cahyono. 2020. «Impacts of Providing Online Teacher and Peer Feedback on Students' Writing Performance». *Turkish Online Journal of Distance Education*, 178–89. <https://doi.org/10.17718/tojde.728157>.
- Wilken, Jayme Lynn. 2017. «Perceptions of L1 Glossed Feedback in Automated Writing Evaluation: A Case Study». *CALICO Journal* 35 (1): 30–48. <https://doi.org/10.1558/cj.26383>.
- Williams, Amanda S. 2018. «Self-Regulation and Feedback in an Educational Statistics Course». *Journal of Teaching and Learning with Technology* 7 (1): 25–42.

- Xie, Qin, und Junting Yuan. 2020. «English Writing Anxiety and Preservice Teacher's Written Corrective Feedback». *Language Education and Assessment* 3 (2): 58–84. <https://doi.org/10.29140/lea.v3n2.357>.
- Yamalee, Eason, und Supong Tangkiengsirisin. 2019. «Effects of Integrated Feedback on Academic Writing Achievement». *Arab World English Journal* 10 (3): 250–70. <https://doi.org/10.24093/awej/vol10no3.17>.
- Yang, Chi Cheung Ruby. 2019. «Student Responses to Online Peer Assessment in Tertiary English Language Classrooms». *The Electronic Journal for English as a Second Language* 23 (1): 1–24.
- Yigitoglu Aptoula, Nur. 2021. «Pre-service teachers' perceptions about the efficacy of different types of feedback on micro-teaching activities». *Kuramsal Eğitimbilim* 14 (2): 79–92. <https://doi.org/10.30831/akukeg.752214>.
- Yonesaka, Suzanne Marie. 2019. «Feedback on Local Intelligibility by Same-L1 Learners: Face-to-Face or Computer-Mediated?». *The Electronic Journal for English as a Second Language* 24 (2): 1–25.
- Zeglen, Eric, und Joseph A. Rosendale. 2018. «Increasing Online Information Retention: Analysing the Effects of Visual Hints and Feedback in Educational Games». *Journal of Open, Flexible and Distance Learning* 22 (1): 22–33.
- Ziegenfuss, Donna H., und Cynthia M. Furse. 2021. «Flipping the Feedback: Formative Assessment in a Flipped Freshman Circuits Class». *Practical Assessment, Research & Evaluation* 26: Art. 8. <https://doi.org/10.7275/007T-DJ06>.






Themenheft Nr. 54:

Forschungssynthesen in der Mediendidaktik. Ansätze und Herausforderungen

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

Methodological Approaches to Evidence Synthesis in Educational Technology

A Tertiary Systematic Mapping Review

Katja Buntins¹ , Svenja Bedenlier² , Victoria Marín³ , Marion Händel⁴ 
und Melissa Bond⁵ 

¹ Universität Duisburg-Essen

² Friedrich-Alexander-Universität Erlangen-Nürnberg

³ Universitat de Lleida

⁴ Hochschule Ansbach

⁵ University College London; University of Stavanger

Abstract

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



Methodische Ansätze zu Evidenzsynthesen in der Bildungstechnologie. Eine tertiäre Übersichtsarbeit

Zusammenfassung

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

1. Introduction

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

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

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

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

1.1 Transparency and replicability in systematic reviews

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

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

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

2. Synthesis of Evidence: General Process Steps

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

2.1 Developing a research question

The research question defines the problems that will be addressed in the subsequent analysis and conclusions. Researchers can approach the development of research questions from a variety of angles, including examining intervention effects, exploring causal relationships, addressing research gaps, and synthesizing existing knowledge (Shea et al. 2007). In evidence synthesis, the formulation of precise and focused research questions plays a crucial role in guiding the entire research process and ensuring meaningful results. Approaches such as the PICO schema (Population,

Intervention, Comparator, Outcome) have been developed to formulate research questions precisely. In this approach, the research question operationalizes the steps that follow, such as search and screening, based on the four constituent elements in the acronym. It aims to standardize the formulation of research questions and subsequent steps to maximize the accuracy of the results (Methley et al. 2014).

2.2 Search and screening

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

This stage includes defining not only the research question but also the inclusion and exclusion criteria applied to identify relevant studies for analysis during the screening process (Abrami, Cohen, and d’Apollonia 1988). Inclusion criteria refer to characteristics that must occur for a study to be included in the analysis, thereby

defining the relevant sample of studies. Exclusion criteria describe the conditions under which a study is excluded from the analysis. They are often contrary to the inclusion criteria but may also include other aspects (Meline 2006).

2.3 Coding and data extraction

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

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

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

2.4 Synthesis

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

3. Research Questions

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

- RQ1: How transparent and comprehensible is the reporting of evidence synthesis methods in reviews in the field of educational technology?*
- RQ2: How many systematic reviews, meta-analyses, and literature reviews are fully replicable?*
- RQ3: Are there differences in replicability depending on the type of evidence synthesis (i.e., between systematic reviews, meta-analyses, or literature reviews?)*

4. Methodology

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

4.1 Search strategy

4.1.1 Search strings

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

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

Tab. 1: Search term in English.

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

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

Tab. 2: Search term in Spanish.

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

Tab. 3: Search term in German.

4.1.2 Search strategy and study selection

The searches were conducted in February and March 2022. As the author team is multilingual, studies in English, Spanish, and German were considered for possible inclusion. The English language databases and platforms searched were ERIC, Scopus, Web of Science, and Google Scholar. For Spanish studies, Dialnet was used, and for German-speaking countries, the FIS database was searched. These databases

were selected because they provide broad coverage of the state of research suitable for evidence synthesis (Gusenbauer and Haddaway 2020), with the Spanish and German databases specifically selected because they are more comprehensive for the Spanish- and German-speaking contexts.

4.1.3 Screening and sampling procedures

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

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

This article specifically focuses on the three types of review previously described; literature reviews (“traditional review family”), systematic reviews and meta-analyses (“systematic review family”; Sutton et al. 2019), with these synthesis approaches being located on opposing ends of the spectrum of transparency and

replicability, and at the same time being the most frequently conducted syntheses in the corpus. As such, 379 evidence syntheses were included in the final corpus for this article.

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

Tab. 4: Inclusion and exclusion criteria.

4.2 Data extraction

In addition to operationalizing the research process research process (see Table 5), a range of data were extracted (see OSF¹ for the complete coding scheme), including publication and authorship information (e.g., publication type and name), review type (self-declared by the authors and informed by the typology of Sutton et al. 2019), specific educational and participant context (e.g., undergraduates, Health and Welfare), and EdTech thematic focus (e.g., blended learning). In regard to methodological characteristics, which are particularly pertinent to the present analysis, data were extracted based on an adaptation of both the Database of Abstracts and Reviews of Effects (DARE) tool (Centre for Reviews and Dissemination 1995), which has been used in previous tertiary reviews (e.g., Lai and Bower 2020), and the AMSTAR 2 tool (Shea et al. 2017). When interpreting these categories, it should be noted that replication is only possible if reporting was complete (yes). If only examples are given, this helps the reader to better understand the scope of the study but does not help to verify or update the results of the study.

¹ <https://osf.io/83vp7/>

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

Tab. 5: Extracted methodological data.

4.3 Synthesis

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

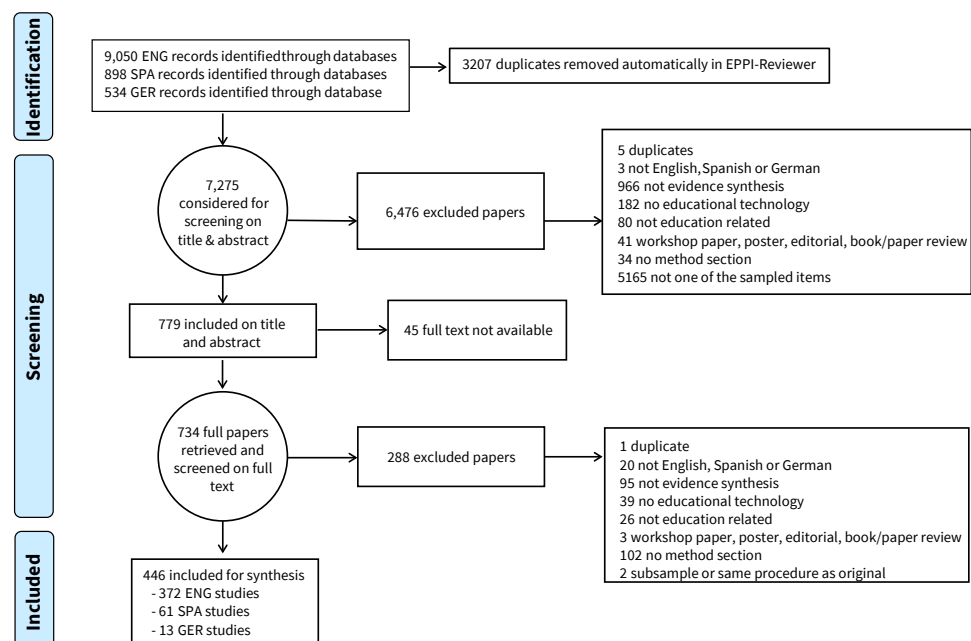


Fig. 1: PRISMA flow diagram.

4.4 Methodological limitations

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

The review was conducted by five reviewers covering three languages, and although attempts were made to reduce inconsistency and bias, the human element remains. Data were only extracted and coded if the information could be found in the article itself, that is, no further research was done to determine, for example, the author's discipline or country of origin if this information could not be found in the manuscript itself. Furthermore, while a protocol was not pre-registered with an official systematic review registry such as Prospero, the full search details can be found on the OSF.² In the future, the authors will be able to use the newly created International Database of Education Systematic Reviews³, which is now accepting protocols for all education disciplines.

4.5 Features of the study

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

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

² <https://osf.io/83vp7/>.

³ <https://idesr.org>.

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

Tab. 6: Self-reported types of verification.

5. Results

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

5.1 *Replicability of the research question*

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

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

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

5.2 *Transparency of the search strategy*

As our study is concerned with the transparency and reproducibility of reviews, we only consider criteria that are important for this: namely, specifying the search string, the search period, the databases searched, and the presence of a PRISMA diagram (see Page et al. 2021). In 87% of reviews, an example of the search string

or the complete search string was given, with 44% reporting the full search string (n=192) and 43% (n=196) providing examples. In 13% of the research syntheses, the search string was not provided. The proportion of reported search terms is highest for systematic reviews, followed by literature reviews and meta-analyses. This order remains the same even when incomplete data are included (see Table 8). This shows that even if, in many cases, no direct replication of the search is possible, the reader at least gets an idea of what is meant by the constructs searched for. Six literature reviews (16%), 24 meta-analyses (18%) and 72 systematic reviews (34%) are still replicable at this point.

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

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

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

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

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

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

5.3 Transparency of screening and data extraction

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

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

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

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

Twenty-one percent of studies reported inter-rater reliability in screening or coding of primary studies (n=92). Publication bias was reported, corrected, or estimated in 19% of the studies (n=86), although it should be noted that this is only

mathematically possible for quantitative reviews. The aim of a quality assessment is to exclude studies with inferior methods. This is not equally necessary for all types of review (e.g., scoping reviews), which partly explains why a quality or critical appraisal was only carried out in 24% (n=105) of cases.

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

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

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

5.4 Replicability of the synthesis

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

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

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

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

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

Tab. 13: Flow Table of Replication.

6. Discussion

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

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

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

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

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

6.1 Further research

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

6.2 Concluding remarks

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

References

- Abrami, Philip C., Peter A. Cohen, and Sylvia d'Apollonia. 1988. «Implementation Problems in Meta-Analysis». *Review of Educational Research* 58 (2): 151–79. <https://doi.org/10.3102/00346543058002151>.
- Alexander, Patricia A. 2020. «Methodological Guidance Paper: The Art and Science of Quality Systematic Reviews». *Review of Educational Research* 90 (1): 6–23. <https://doi.org/10.3102/0034654319854352>.
- Atlam, Hany F., Muhammad Ajmal Azad, Madini O. Alassafi, Abdulrahman A. Alshdadi, and Ahmed Alenezi. 2020. «Risk-Based Access Control Model: A Systematic Literature Review». *Future Internet* 12 (6). <https://doi.org/10.3390/fi12060103>.
- Baumeister, Roy F., and Mark R. Leary. 1997. «Writing Narrative Literature Reviews». *Review of General Psychology* 1 (3): 311–20. <https://doi.org/10.1037/1089-2680.1.3.311>.
- Belur, Jyoti, Lisa Tompson, Amy Thornton, and Miranda Simon. 2018. «Interrater Reliability in Systematic Review Methodology: Exploring Variation in Coder Decision-Making». *Sociological Methods & Research* 50 (2): 837–65. <https://doi.org/10.1177/0049124118799372>.

- Bond, Melissa. 2020. «Schools and Emergency Remote Education During the COVID-19 Pandemic: A Living Rapid Systematic Review». *Asian Journal of Distance Education* 15 (2): 191–247. <https://doi.org/10.5281/zenodo.4425683>.
- Bond, Melissa, Katja Buntins, Svenja Bedenlier, Olaf Zawacki-Richter, and Michael Kerres. 2020. «Mapping Research in Student Engagement and Educational Technology in Higher Education: A Systematic Evidence Map». *International Journal of Educational Technology in Higher Education* 17 (1): 1–30. <https://doi.org/10.1186/s41239-019-0176-8>.
- Borenstein, Michael, ed. 2009. *Introduction to Meta-Analysis*: John Wiley & Sons, Ltd.
- Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2010. «A Basic Introduction to Fixed-Effect and Random-Effects Models for Meta-Analysis». *Research Synthesis Methods* 1 (2): 97–111. <https://doi.org/10.1002/jrsm.12>.
- Borrego, Maura, Margaret J. Foster, and Jeffrey E. Froyd. 2014. «Systematic Literature Reviews in Engineering Education and Other Developing Interdisciplinary Fields». *Journal of Engineering Education*. 103 (1): 45–76. <https://doi.org/10.1002/jee.20038>.
- Bramer, Wichor M., Dean Giustini, and Bianca Kramer. 2016. «Comparing the Coverage, Recall, and Precision of Searches for 120 Systematic Reviews in Embase, MEDLINE, and Google Scholar: A Prospective Study». *Systematic Reviews* 5 (1): 39. <https://doi.org/10.1186/s13643-016-0215-7>.
- Bramer, Wichor M., Dean Giustini, Bianca Kramer, and P. F. Anderson. 2013. «The Comparative Recall of Google Scholar Versus PubMed in Identical Searches for Biomedical Systematic Reviews: A Review of Searches Used in Systematic Reviews». *Systematic Reviews* 2 (1): 115. <https://doi.org/10.1186/2046-4053-2-115>.
- Braun, Virginia, and Victoria Clarke. 2006. «Using Thematic Analysis in Psychology». *Qualitative Research in Psychology* 3 (2): 77–101. <https://doi.org/10.1191/1478088706qp0630a>.
- Bulfin, Scott, Michael Henderson, Nicola F. Johnson, and Neil Selwyn. 2014. «Methodological Capacity Within the Field of «Educational Technology» Research: An Initial Investigation». *British Journal of Educational Technology* 45 (3): 403–14. <https://doi.org/10.1111/bjet.12145>.
- Buntins, Katja, Michael Kerres, and Anna Heinemann. 2022. «A Scoping Review of Research Instruments for Measuring Student Engagement: In Need for Convergence». *International Journal of Educational Research Open* 2. <https://doi.org/10.1016/j.ijedro.2021.100099>.
- Buntins, Katja, Miriam Mulders, and Nadine Schröder. 2023. «Cognitive Biases in Screening Processes – Search Strategies in Educational Technology Research a Systematic Review on Learning with Virtual Reality». *MedienPädagogik* 54: 103–124. <https://doi.org/10.21240/mpaed/54/2023.11.19.X>.
- Campbell, Allison, Brian Taylor, Jessica Bates, and Una O'Connor-Bones. 2018. «Developing and Applying a Protocol for a Systematic Review in the Social Sciences». *New Review of Academic Librarianship* 24 (1): 1–22. <https://doi.org/10.1080/13614533.2017.1281827>.
- CASP. 2022. «CASP Checklists. Critical Appraisal Skills Programme». <https://casp-uk.net/casp-tools-checklists/>.

- Cooper, Chris, Rebecca Lovell, Kerryn Husk, Andrew Booth, and Ruth Garside. 2018a. «Supplementary Search Methods Were More Effective and Offered Better Value Than Bibliographic Database Searching: A Case Study from Public Health and Environmental Enhancement». *Research Synthesis Methods* 9 (2): 195–223. <https://doi.org/10.1002/jrsm.1286>.
- Cooper, Chris, Joanna Varley-Campbell, Andrew Booth, Nicky Britten, and Ruth Garside. 2018b. «Systematic Review Identifies Six Metrics and One Method for Assessing Literature Search Effectiveness but No Consensus on Appropriate Use». *Journal of Clinical Epidemiology* 99:53–63. <https://doi.org/10.1016/j.jclinepi.2018.02.025>.
- Cooper, Harris, and L. V. Hedges, eds. 1994. *Handbook of Research Synthesis*: Russell Sage Foundation.
- Crompton, Helen, Mildred V. Jones, and Diane Burke. 2022. «Affordances and Challenges of Artificial Intelligence in K-12 Education: A Systematic Review». *Journal of Research on Technology in Education*, 1–21. <https://doi.org/10.1080/15391523.2022.2121344>.
- Durlak, J. A., and M. W. Lipsey. 1991. «A Practitioner’s Guide to Meta-Analysis». *American Journal of Community Psychology* 19 (3): 291–332. <https://doi.org/10.1007/BF00938026>.
- Eysenbach, G., J. Tuische, and T. L. Diepgen. 2001. «Evaluation of the Usefulness of Internet Searches to Identify Unpublished Clinical Trials for Systematic Reviews». *Medical Informatics and the Internet in Medicine* 26 (3): 203–18. <https://doi.org/10.1080/14639230110075459>.
- Glass, Gene V., Barry McGaw, and Mary Lee Smith. 1981. «Meta-analysis in social research». SAGE Publications.
- Glass, Gene V., and M. L. Smith. 1979. «Meta-Analysis of Research on Class Size and Achievement». *Educational Evaluation and Policy Analysis* 1 (1): 2–16. <https://doi.org/10.3102/01623737001001002>.
- Gough, David, Phil Davies, Gro Jamtvedt, Etienne Langlois, Julia Littell, Tamara Lotfi, Edoardo Masset et al. 2020. «Evidence Synthesis International (ESI): Position Statement». *Systematic Reviews* 9 (1): 155. <https://doi.org/10.1186/s13643-020-01415-5>.
- Hammersley, Martyn. 2001. «On ‘Systematic’ Reviews of Research Literatures: A ‘Narrative’ Response to Evans & Benefield». *British Educational Research Journal* 27 (5): 543–54. <https://doi.org/10.1080/01411920120095726>.
- Hartling, Lisa, Andrea Milne, Michele P. Hamm, Ben Vandermeer, Mohammed Ansari, Alexander Tsertsvadze, and Donna M. Dryden. 2013. «Testing the Newcastle Ottawa Scale Showed Low Reliability Between Individual Reviewers». *Journal of Clinical Epidemiology* 66 (9): 982–93. <https://doi.org/10.1016/j.jclinepi.2013.03.003>.
- Hew, Khe Foon, Min Lan, Ying Tang, Chengyuan Jia, and Chung Kwan Lo. 2019. «Where Is the ‘Theory’ Within the Field of Educational Technology Research?» *British Journal of Educational Technology* 50 (3): 956–71. <https://doi.org/10.1111/bjet.12770>.
- Higgins, Julian P. T., Douglas G. Altman, Peter C. Gøtzsche, Peter Jüni, David Moher, Andrew D. Oxman, Jelena Savović, Kenneth F. Schulz, Laura Weeks, and Jonathan A. C. Sterne. 2011. «The Cochrane Collaboration Tool for Assessing Risk of Bias in Randomised Trials». *BMJ* 343. <https://doi.org/10.1136/bmj.d5928>.

- Higgins, Julian P. T., J. Thomas, J. Chandler, M. Cumpston, T. Li, Matthew J. Page, and Vivian A. Welch. 2022. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 6.3: Cochrane.
- Ioannidis, John P. A., Marcus R. Munafò, Paolo Fusar-Poli, Brian A. Nosek, and Sean P. David. 2014. «Publication and Other Reporting Biases in Cognitive Sciences: Detection, Prevalence, and Prevention». *Trends in Cognitive Sciences* 18 (5): 235–41. <https://doi.org/10.1016/j.tics.2014.02.010>.
- Katrak, Persis, Andrea E. Bialocerkowski, Nicola Massy-Westropp, V. S. Kumar, and Karen A. Grimmer. 2004. «A Systematic Review of the Content of Critical Appraisal Tools». *BMC Medical Research Methodology* 4 (1): 1–11. <https://doi.org/10.1186/1471-2288-4-22>.
- Kimmons, Royce, and Joshua M. Rosenberg. 2022. «Trends and Topics in Educational Technology, 2022 Edition». *TechTrends* 66 (2): 134–40. <https://doi.org/10.1007/s11528-022-00713-0>.
- Kolbe, Richard H., and Melissa S. Burnett. 1991. «Content-Analysis Research: An Examination of Applications with Directives for Improving Research Reliability and Objectivity». *Journal of Consumer Research* 18 (2): 243–50. <https://doi.org/10.1086/209256>.
- Krippendorff, Klaus. 2018. *Content Analysis: An Introduction to Its Methodology*: SAGE Publications.
- Lai, Jennifer W. M., and Matt Bower. 2020. «Evaluation of Technology Use in Education: Findings from a Critical Analysis of Systematic Literature Reviews». *Journal of Computer Assisted Learning* 36 (3): 241–59. <https://doi.org/10.1111/jcal.12412>.
- Landis, J. Richard, and Gary G. Koch. 1977. «The Measurement of Observer Agreement for Categorical Data». *Biometrics* 33 (1): 159–74. <https://doi.org/10.2307/2529310>.
- Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. «Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability». *Human Communication Research* 28 (4): 587–604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>.
- Marín, Victoria I., Katja Buntins, Svenja Bedenlier, and Melissa Bond. 2023. «Invisible Borders in Educational Technology Research? A Comparative Analysis». *Educational Technology Research and Development: ETR & D* 71: 1349–1370. <https://doi.org/10.1007/s11423-023-10195-3>.
- Meline, Timothy. 2006. «Selecting Studies for Systemic Review: Inclusion and Exclusion Criteria». *Contemporary Issues in Communication Science and Disorders* 33 (Spring): 21–27. https://doi.org/10.1044/cicsd_33_S_21.
- Methley, Abigail M., Stephen Campbell, Carolyn Chew-Graham, Rosalind McNally, and Sudeh Cheraghi-Sohi. 2014. «PICO, PICOS and SPIDER: A Comparison Study of Specificity and Sensitivity in Three Search Tools for Qualitative Systematic Reviews». *BMC Health Services Research* 14:579. <https://doi.org/10.1186/s12913-014-0579-0>.
- Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. «Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement». *PLoS Medicine* 6 (7): e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer et al. 2021. «The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews». *BMJ* 372. <https://doi.org/10.1136/bmj.n71>.

- Pigott, Therese D., and Joshua R. Polanin. 2020. «Methodological Guidance Paper: High-Quality Meta-Analysis in a Systematic Review». *Review of Educational Research* 90 (1): 24–46. <https://doi.org/10.3102/0034654319877153>.
- Polanin, Joshua R., Brandy R. Maynard, and Nathaniel A. Dell. 2017. «Overviews in Education Research: A Systematic Review and Analysis». *Review of Educational Research* 87 (1): 172–203. <https://doi.org/10.3102/0034654316631117>.
- Pollock, Alex, and Eivind Berge. 2018. «How to Do a Systematic Review». *International Journal of Stroke* 13 (2): 138–56. <https://doi.org/10.1177/1747493017743796>.
- Shea, Beverley J., Jeremy M. Grimshaw, George A. Wells, Maarten Boers, Neil Andersson, Candyce Hamel, Ashley C. Porter, Peter Tugwell, David Moher, and Lex M. Bouter. 2007. «Development of AMSTAR: A Measurement Tool to Assess the Methodological Quality of Systematic Reviews». *BMC Medical Research Methodology* 7 (1): 10. <https://doi.org/10.1186/1471-2288-7-10>.
- Singh, Jatinder. 2013. «Critical Appraisal Skills Programme». *Journal of pharmacology and Pharmacotherapeutics* 4 (1): 76. <https://doi.org/10.4103/0976-500X.107697>.
- Sotola, Lukas K. 2022. «Garbage In, Garbage Out? Evaluating the Evidentiary Value of Published Meta-Analyses Using Z-Curve Analysis». *Collabra: Psychology* 8 (1): 32571. <https://doi.org/10.1525/collabra.32571>.
- Stock, William A. 2008. «Systematic Coding for Research Synthesis». *The Handbook of Research Synthesis* 236: 125–38. Russell Sage Foundation.
- Tamim, Rana M., Robert M. Bernard, Eugene Borokhovski, Philip C. Abrami, and Richard F. Schmid. 2011. «What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study». *Review of Educational Research* 81 (1): 4–28. <https://doi.org/10.3102/0034654310393361>.
- Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., Bond, M., and Koryakina, A. 2023. «EPPI Reviewer: advanced software for systematic reviews, maps and evidence synthesis». EPPI Centre Software. UCL Social Research Institute. London. <https://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4>.
- Viera, Anthony J., and Joanne M. Garrett. 2005. «Understanding Interobserver Agreement: The Kappa Statistic». *Family Medicine* 37 (5): 360–63. <https://pubmed.ncbi.nlm.nih.gov/15883903/>.
- Zawacki-Richter, Olaf, Michael Kerres, Svenja Bedenlier, Melissa Bond, and Katja Buntins, eds. 2022. *Systematic Reviews in Educational Research*. Springer Fachmedien: Wiesbaden. <https://doi.org/10.1007/978-3-658-27602-7>.

Acknowledgments

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