ONLINE LEARNING
Volume 30 · Issue SI2 · 2026 · e6749
https://doi.org/10.24059/olj.v30iSI2.6749

Towards responsible AI-augmented evidence synthesis in Open, Distance, and Digital Education (ODDE)

Olaf Zawacki-Richter
Carl von Ossietzky Universität Oldenburg, Germany
Florence Martin
North Carolina State University, USA
Melissa Bond
University College London, UK
Richard E. West
Brigham Young University, USA

The growing importance of evidence synthesis in ODDE and quality issues

Over the past two decades, open, distance, and digital education (ODDE) has developed into a broad and increasingly diverse field of research (Zawacki-Richter & Jung, 2023). The expansion of online and blended learning, the proliferation of digital educational technologies, and, more recently, the rapid emergence of generative artificial intelligence (GenAI) have contributed to a substantial increase in empirical research. ODDE is now in the mainstream of higher education (Xiao, 2018) At the same time, the growing volume and fragmentation of this literature have created an increasing need for systematic approaches to synthesizing research findings, identifying patterns and gaps, and building cumulative knowledge, in order to inform evidence-based decisions in the digital transformation of higher education (Martin et al., 2023).

Systematic reviews, meta-analyses, and qualitative research syntheses play a crucial role in this process. Unlike traditional narrative literature reviews, systematic approaches to evidence synthesis aim to identify, evaluate, and integrate existing research through explicit, transparent, and reproducible procedures (Gough et al., 2017; Zawacki-Richter et al., 2020; Martin, 2025). They provide an important foundation for theory development, evidence-informed educational practice, and policy decisions. In a field characterized by rapid technological change and considerable diversity in institutional contexts, learner populations, and pedagogical approaches, such syntheses are particularly valuable.

Evidence synthesis and systematic reviews in particular are on the rise in ODDE. In an umbrella mapping review covering 576 ODDE reviews published between 2018 and 2022 the annual number increased from 35 to 251, more than sevenfold between the two endpoints (Zawacki-Richter et al., 2025). However, the growing popularity of systematic reviews has also raised important questions about their methodological quality and rigor. Recent investigations into the state of evidence synthesis in educational technology and ODDE have identified considerable variation in methodological approaches, reporting standards, and the transparency of review procedures (Forsström et al., 2026).

For example, Bond (2024) examined 66 reviews of AI in higher education, published between 2018 and July 2023, revealing concerning findings. Just over half of the reviews in the

Corresponding author: Olaf Zawacki-Richter, olaf.zawacki.richter@uni-oldenburg.de
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

corpus did not report agreement in screening or coding, 45% included no quality appraisal, and about a third did not discuss any methodological limitations. Overall, based on a quality appraisal across 10 key reporting areas, 65% received a score ranging from critically low to medium. Likewise, in a tertiary mapping review of 446 evidence syntheses in the field of educational technology, Buntins et al. (2023) found that only 44% reported the complete search string, 62% outlined the inclusion/exclusion criteria, 37% provided the data extraction coding scheme, and only 26% of systematic reviews performed a quality assessment. Zawacki-Richter et al.'s (2025) umbrella review points in a similar direction. Only about 8% of the 576 systematic reviews in ODDE reported a review protocol, around 19% reported inter-rater reliability, and 26% reported quality appraisal. Furthermore, only about two-thirds provided a full search string or a PRISMA flow chart. Such gaps make it difficult to understand and reproduce the review process, thereby seriously undermining the trustworthiness of the reported evidence.

These findings raise a fundamental question: How can we ensure that the increasing quantity of systematic reviews translates into a corresponding improvement in the quality and usefulness of the evidence they produce? This question provided the starting point for the present special issue of the *Online Learning Journal*. Our aim was to bring together high quality exemplary systematic reviews, meta-analyses, and meta-syntheses addressing substantive research questions in ODDE, as well as methodological contributions that advance the standards of evidence synthesis in ODDE.

In developing the call for papers, we deliberately adopted a broad understanding of evidence synthesis. While systematic reviews and meta-analyses are well-established approaches, the field also benefits from scoping reviews, mapping reviews, qualitative meta-syntheses, umbrella reviews, and other forms of research synthesis. These approaches serve different purposes and address different kinds of research questions. What unites them is the commitment to systematic, transparent, and methodologically justified procedures for integrating existing knowledge.

For assessing the quality and rigor of the submitted review manuscripts, we applied the Quality of Evidence Synthesis Tool (QuEST), developed by Bond (2024). It was adapted from the AMSTAR 2 (Shea et al., 2017) and DARE (Centre for Reviews and Dissemination, 1995) appraisal tools, to enable the appraisal of review reporting quality across a range of evidence synthesis approaches, rather than just systematic reviews or meta-analyses, especially given the heterogeneity of education research. The QuEST has eleven criteria that evaluate the clarity of research questions, eligibility criteria, search coverage, search strategy, agreement between reviewers, and the disclosure of digital tools.

The contributions to this special issue also emerge at a particularly interesting moment: AI is beginning to transform not only the educational practices we investigate in ODDE but also the primary and secondary research methods we employ.

AI and the future of evidence synthesis

AI applications offer considerable opportunities to support systematic review processes, particularly in research fields characterized by rapidly expanding publication volumes such as AI in education (Karlstrøm, 2024). AI tools can potentially assist researchers at several stages of the systematic review process, including literature identification, screening, classification, data extraction, and synthesis (Scheelbeek et al., 2024). Machine learning approaches have been particularly relevant to the prioritization and screening of large numbers of potentially relevant publications. These developments raise the prospect of making systematic reviews more efficient and scalable (Burgard & Bittermann, 2023). Researchers may be able to screen larger bodies of literature, identify relevant publications more quickly, and reduce the time required for repetitive tasks.

However, increased efficiency does not necessarily translate into improved methodological quality. The application of AI to evidence synthesis introduces a range of challenges concerning accuracy, reliability, transparency, bias, and reproducibility. AI-generated classifications and extracted information may contain errors, omit relevant contextual information, or reflect biases embedded in training data and underlying algorithms (Jansen et al., 2026). Generative AI systems may also produce plausible but inaccurate interpretations or references (Seifi & Seyfi, 2026). Even when factually accurate, AI-generated syntheses may remain generic and descriptive, lacking the conceptual depth, originality, and critical interpretation needed to advance knowledge. Developing a meaningful and theoretically informed narrative from the available evidence therefore requires the judgment and domain expertise of human researchers.

Consequently, the integration of AI into systematic reviews should not be understood simply as a matter of replacing human activities with automated procedures. Rather, it requires a reconsideration of how human expertise and computational capabilities can be combined within a rigorous methodological framework, as suggested by the Responsible use of AI in evidence SynthEsis recommendations (RAISE Guidelines; Flemyng et al., 2025; Thomas et al., 2026).

From AI automation to human-in-the-loop evidence synthesis

One promising approach to integrating AI into systematic reviews is the development of *human-in-the-loop* procedures (Lazaros et al., 2026). Rather than replacing researchers, these approaches distribute tasks between AI systems and human reviewers, with the intention of combining computational efficiency with human subject matter expert judgment.

This principle is illustrated by the contribution of Ji Hyun Yu, Fengjiao Tu, Li Cheng, Shasha Li, Zhizi Zheng in this special issue. Their systematic review investigates how generative AI affects self-regulated learning and metacognitive processes. However, the study is also methodologically noteworthy because the authors themselves employ generative AI in conducting their review. Following the DISCAR framework for systematic reviews (West & Martin, 2024) and PRISMA 2020 reporting guidelines, Yu et al. implemented a semi-automated screening procedure based on a *confidence-based human-in-the-loop* protocol. Their initial searches across six bibliographic databases identified 10,243 records. Following automated filtering and duplicate removal, 2,909 records remained for title and abstract screening. The authors employed GPT-4o to classify records against predefined inclusion and exclusion criteria. The AI system assessed records across four screening dimensions and generated confidence scores for its classifications. Rather than accepting all automated decisions, the research team implemented a differentiated verification strategy that directed human attention towards potentially uncertain classifications.

The methodological importance of this approach becomes apparent in the reported screening results. Manual inspection of 823 low-confidence exclusions identified three incorrectly excluded studies. In contrast, a random audit of approximately 10% of 893 high-confidence exclusions identified no additional false negatives. The authors also conducted full manual audits at other screening stages, including educational context, generative AI relevance, and empirical study status. Ultimately, 65 empirical studies were included in the review. This example illustrates how AI-assisted screening can be combined with targeted human verification

From a methodological perspective, the contribution of Yu et al. demonstrates that the value of AI-assisted evidence synthesis lies not simply in reducing the number of records that require manual inspection, but in designing a transparent division of labor between human researchers and AI systems. This raises a broader question for the future of systematic reviews: Which activities can reasonably be delegated to AI, which require human judgment, and how should the reliability of this division of responsibilities be established?

Towards responsible AI-augmented evidence synthesis

Taken together, the methodological discussion by Karlstrøm (2024) and the contribution of Yu et al. in this issue suggest that the future of systematic reviews should not be framed as a choice between traditional human-led approaches and increasingly automated AI-driven procedures. Instead, the challenge is to develop responsible AI-augmented evidence synthesis, in which AI capabilities are integrated into review processes without compromising their epistemological foundations.

We suggest that such an approach should be guided by five interrelated principles, supported by the RAISE Guidelines (Flemyng et al., 2025; Thomas et al., 2026) and other global Digital Evidence Synthesis Tool projects and standards (Farotimi et al., 2026; Scheelbeek et al., 2024).

1. *Methodological criteria must precede automation*. Review questions, eligibility criteria, coding frameworks, and quality appraisal standards should be established independently of the outputs generated by AI systems and under full control of human researchers. These elements provide the foundation against which automated decisions can subsequently be evaluated.
2. *The division of labor between researchers and AI systems must be transparent.* Review authors should document which tasks were performed by AI, which decisions were made by human reviewers, and where human intervention occurred. This includes reporting the AI tools and model versions employed, the instructions or prompts used, and the procedures through which outputs were assessed.
3. *AI-generated decisions require systematic validation*. Human oversight should be operationalized through explicit procedures rather than invoked as a general assurance of quality. Depending on the review task, these procedures may include independent human coding, sampling and auditing of automated decisions, agreement assessments, and analyses of classification errors. Particular attention should be paid to the possibility that apparently confident AI outputs may nevertheless be incorrect or lack novelty or usefulness.
4. *Fourth, researchers must retain substantive epistemic responsibility.* AI may support the identification, organization, and synthesis of evidence, but the interpretation of findings, assessment of methodological limitations, and formulation of conclusions remain scholarly responsibilities of the subject matter experts. This is particularly important in interdisciplinary fields such as ODDE, where concepts, educational settings, and research traditions vary considerably.
5. *AI-assisted review procedures must be sufficiently reproducible and auditable*. Transparency about AI use is a necessary starting point, but reproducibility also requires access, wherever possible, to search strategies, screening decisions, coding schemes, validation procedures, and relevant research data. The potential variability of generative AI outputs makes careful documentation particularly important.

The implications of these principles extend beyond the efficiency of individual reviews. As AI-assisted evidence synthesis becomes more widespread, the field will need empirical studies comparing human-led, AI-assisted, and more extensively automated review procedures, although this work has already begun to be collated (see e.g. Eisele-Metzger et al., 2026; Scheelbeek et al., 2024). Such research should investigate not only time savings but also sensitivity, specificity, coding reliability, reproducibility, and the consequences of errors for substantive conclusions.

This special issue therefore provides an opportunity to consider the future of evidence synthesis in ODDE from two complementary perspectives: the continued development of rigorous methodological standards and the critical evaluation of emerging technologies that may transform how those standards are implemented.

A brief overview of the contributions to this special issue

The 14 contributions to this special issue reflect the diversity of contemporary evidence synthesis in open, distance, and digital education (ODDE) and can be broadly grouped into four thematic areas: artificial intelligence, AI literacy, and self-regulated learning; learner success, engagement, and support; theoretical perspectives and developments in ODDE; and methodological advances in evidence synthesis.

AI Literacy and Self-Regulated Learning

The first four contributions focus on artificial intelligence, AI literacy, and self-regulated learning, examining how AI is transforming educational practices, the competencies learners need to engage with these technologies, and the pedagogical conditions under which AI can support meaningful learning.

In their paper, *"Refining Collaborative AI Literacy: A Systematic Review of Human-AI Collaboration Competency and AI Literacy Measures",* Ji Won Yang, Junseo Kim, Soojeong Jeong and Jaeseo Lim synthesised 17 studies exploring human-AI collaboration competency and AI literacy in higher education. The review indicates that the conceptualisation and measurement of human-AI collaboration competency remain at a nascent stage, with competency rarely assessed explicitly within existing AI literacy measures, let alone as a distinct construct. The authors propose three key dimensions of collaborative AI literacy: coordination, communication and metacognition. By distinguishing knowledge about AI and proficiency in its use from the competencies needed to collaborate with it, the review provides a conceptual foundation for developing and validating more targeted measures of collaborative AI literacy. In doing so, it highlights the need to assess not simply whether students can use AI, but how effectively they coordinate, communicate and exercise judgement when working with it.

Ji Hyun Yu, Fengjiao Tu, Li Cheng, Shasha Li, and Zhizi Zheng investigate the relationship between generative AI, self-regulated learning, and metacognition in *"Metacognitive Laziness or Pedagogical Scaffolding? A Systematic Review of GenAI's Impact on SRL and Metacognition."* Based on a systematic review and Epistemic Network Analysis of 65 empirical studies published between 2022 and 2026, the authors examine how different forms of pedagogical scaffolding shape learners' engagement with generative AI. Their findings suggest that instructor mediation plays an important role in supporting proactive planning and sustained self-regulation, whereas reflective prompting alone may encourage reactive monitoring without deeper regulatory engagement. Particularly noteworthy is the absence of explicit feedback literacy scaffolding across the reviewed studies, pointing to a significant gap in preparing learners to critically evaluate AI-generated outputs. Beyond its substantive findings, the study also illustrates the potential of AI-assisted screening combined with systematic human verification, as discussed in the preceding section.

Zhuo Zhang, Belle Li, Qian Xu, Zui Cheng, and Richard E. West examine the effectiveness of artificial intelligence in supporting self-regulated learning in *"A Meta-Analysis of AI-Assisted Self-Regulated Learning."* Synthesizing 34 empirical studies and 91 effect sizes, the authors report a positive overall effect of AI-based pedagogical interventions on self-regulated learning (Hedges' $g = 0.76$). Positive effects are also identified for cognitive, metacognitive, and resource management strategies, with the largest effect observed for resource management. Moderator analyses indicate that the effectiveness of AI interventions varies significantly according to educational level, the pedagogical role of AI, and the type of AI application. These findings provide quantitative evidence of the potential of AI to support learners' self-regulation while emphasizing that its effectiveness depends on how technologies are integrated into educational contexts. Together with the contribution by Yu et al., this meta-analysis offers complementary perspectives on AI-supported self-regulated learning, combining quantitative evidence of

intervention effects with a more detailed examination of the pedagogical conditions under which meaningful self-regulation may emerge.

Lok Tin Cheung, Haley Vu, Brooke Osborne, Nayana Parange and Sandhya Maranna synthesized research on GenAI specific to tertiary health education in "The integration of Generative Artificial Intelligence into Tertiary Health Education: A Scoping Review". Their review identified 20 studies which were organized to three categories, 1) application of GenAI in tertiary health education, 2) perceptions of GenAI use in health education, and 3) ethical and responsible use of GenAI in health education. Their findings indicated that GenAI is used to automate academic tasks, enhance course design and support critical thinking skills development among tertiary health students. Concerns included misinformation, over-reliance, and data confidentiality. The review also revealed limited structured training in AI literacy, underscoring the need for structured AI literacy programs and clear ethical guidelines to help students use GenAI.

Learner Success, Engagement, and Support

The following six contributions address learner success, engagement, and support in online education, synthesizing evidence on the factors that influence academic achievement, persistence, and retention, as well as the instructional and institutional strategies that can support diverse learner populations.

Jessica Sylvester and Melinda Kulick focus on a central challenge for online higher education in *"Mapping Evidence on Persistence and Retention Strategies for Non-Traditional Students in Online Higher Education: A Scoping Review."* Guided by the Joanna Briggs Institute framework, their scoping review maps research published between 2013 and 2025 on how persistence and retention among non-traditional students are conceptualized, investigated, and measured. The authors identify considerable variation in definitions and outcome measures, including a frequent lack of distinction between students' continued participation in education and institutional retention metrics. Their analysis also identifies five interconnected thematic domains encompassing instructional, learner-related, and institutional factors. While technology- and AI-mediated approaches are increasingly discussed, the evidence concerning their effectiveness remains limited. The review thus provides an important foundation for developing more coherent research approaches and evidence-informed institutional strategies to support non-traditional learners, particularly those balancing academic study with employment and family responsibilities.

In *"Online Learner Success Dimensions and Factors: A Framework Development Through Systematic Review,"* Florence Martin and Jered Borup offer a comprehensive synthesis of the factors associated with success in online learning. Drawing on 188 studies, the authors develop a framework comprising 30 success factors organized into four dimensions: learners, courses and instructors, programs and organizations, and communities. Their analysis reveals that course- and instructor-related factors have received the greatest research attention, followed by learner characteristics, while organizational and community dimensions have been investigated less extensively. By integrating findings across these different levels, the review moves beyond isolated explanations of learner success and emphasizes its multidimensional nature. The resulting framework provides a valuable conceptual foundation for future research and offers practical guidance for the design, implementation, and evaluation of online learning programs.

Bobbie Sandberg and Richard West synthesized the literature to investigate parent perspectives on how instructional modality affects type and amount of support parents provide to their adolescent online learners in "Modality Matters to Parents Who Support Adolescent Online Learners: A Literature Synthesis". From the 48 studies that were included in the review, the synthesis demonstrates that design decisions regarding modality can affect parental support roles,

parental perception of online school support, other aspects of parents' lives outside the online course, and students' online learning support needs. They recommend a Bichronous design, blending synchronous and asynchronous content to keep the parents involved.

Andrew Walker, Kaelyn Loveless, Naomi Aguiar, Yu-Chun Kuo, Raechel Sociher, and Jessica Resign examined the effectiveness of research quality as a moderator between student motivation and academic performance in *Research quality moderates the relationship between student motivation and academic performance in online and blended courses: A meta-analysis.* The meta-analysis included 64 studies with 117 outcomes where measurement and methodological quality moderated the relationship between student motivation and academic performance in online and blended learning. Their findings resulted in modest correlational values (r=0.14) indicating motivational factors can be overstated when research quality is taken into account.

Curtis Bonk, Meina Zhu and Min Young Doo reviewed the research on retention and learning engagement in online learning in Systematic Review of Retention and Learning Engagement in Online Learning from 2015-2025. They analyzed the 19 studies that met their inclusion criteria. Factors of engagement and retention, instructor-related factors or instructional strategies were most frequently examined, and this was followed by learner-related individual factors and course characteristics. They also found that behavioral engagement was most studied as a single dimension while multiple dimensions of engagement were most frequently examined.

Who is AI supporting? A systematic review of AI-enabled learner support in higher education is another review demonstrating strong utility for practitioners. In this review, Berrin Cefa, Olaf Zawacki-Richter, and Felicitas Macgilchrist, who reviewed 61 publications from 2010-2025 to understand how higher education institutions are using AI-enabled tools to support learners. They identified many potential benefits for learner support and crucially identified critical blind spots that need further attention. In our current AI-integrated age, their review provides key suggestions for future research and strong strategies for practice about how to use AI effectively to support learners as well as instructors.

Theoretical Perspectives and Developments in ODDE

The following three contributions highlight the potential of evidence synthesis to revisit, critically examine, and extend theoretical perspectives in ODDE, demonstrating how systematic reviews can contribute not only to consolidating existing knowledge but also to advancing theory and informing educational practice.

A main potential contribution of literature syntheses is the ability to reconsider theories in light of new research. In *Exploring the influence of generative AI on student's learning presence*, Min Young Doo and Yeongjeong Park provide a good example of this as they reviewed 21 peer-reviewed studies specifically to review evidence on how AI is affecting teaching, cognitive, and social presence within the Community of Inquiry framework. They found that generative AI may not be universally strengthening the COI framework and the three types of presence, but that there are ways AI can impact within nuanced, specific instructional contexts.

Chungsoo Na, Fan Yang, and Andrew Walker continued a theoretical deep dive into the Community of Inquiry framework by exploring learning presence as a fourth type of presence that could be added to the framework. In their article, *Revisiting learning presence in the Community of Inquiry presence: Insights from a systematic review,* they identified and reviewed 51 studies on learning presence over a 15 year span, and found this to be a commonly discussed type of presence—often studied as a type of learning regulation. Their meta-analysis on these studies found strong positive associations between LP and teaching presence (r = .55), social presence (r = .62), cognitive presence (r = .66), learning outcomes (r = .53), and satisfaction (r = .54). Their

review stands as another example of how a systematic review, and in this case a meta-analysis, can support and extend theory within existing theoretical frameworks.

Jennifer Darling-Aduana, Qurrat ul Ain Rasheed, Ijeoma Mbaezue, and Kristin Hemingway, in their systematic review also deepened theoretical understanding as they reviewed how online education can affect critical consciousness through a critical theory lens. In *Fostering critical consciousness online: Strategies, opportunities, and gaps in the literature,* they reviewed 96 studies published over a 13-year period and found that there are strategies that can help online students raise their critical consciousness and many tools have affordances to support this. Their article provides practical suggestions and best practices for online educators, showcasing how quality literature syntheses should contribute to both theory and practice.

Methodological Advances in Evidence Synthesis

Finally, Carl Westine, Chang Tong, Ting Sun, and Tong Wu address an important methodological challenge in their paper, *"Managing Overlapping Review Studies in a Second-Order Meta-Analysis."* As the number of systematic reviews and meta-analyses continues to increase, second-order reviews, or overviews of reviews, offer opportunities to integrate findings across existing evidence syntheses. However, when primary studies appear in multiple reviews, their overlapping inclusion can distort the interpretation of aggregated findings. Drawing on an existing dataset from an ODDE overview, the authors examine and compare different strategies for identifying, quantifying, visualizing, and managing overlapping studies. Their contribution highlights the methodological decisions involved in conducting second-order evidence syntheses and provides valuable guidance for researchers seeking to improve the rigor and trustworthiness of umbrella reviews and other higher-order syntheses in ODDE.

Conclusion

Altogether, these contributions demonstrate the diversity and growing significance of evidence synthesis in ODDE. They illustrate how systematic reviews, meta-analyses, and other forms of research synthesis can not only consolidate existing knowledge but also identify critical research gaps, advance theoretical understanding, and inform educational policy and practice. As artificial intelligence increasingly transforms both education and educational research, maintaining methodological rigor, transparency, and critical scholarly judgment will be essential to ensuring that evidence synthesis produces trustworthy, meaningful, and actionable knowledge.

References

Bond, M. (2024). Quality of evidence synthesis tool (QuEST). Zenodo. https://doi.org/10.5281/zenodo.14211692

Buntins, K., Bedenlier, S., Marín, V., Händel, M., & Bond, M. (2023). Methodological Approaches to Evidence Synthesis in Educational Technology. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, *54*, 167–191. https://doi.org/10.21240/mpaed/54/2023.12.20.X

Burgard, T., & Bittermann, A. (2023). Reducing literature screening workload with machine learning: A systematic review of tools and their performance. *Hotspots in Psychology, 231*(1), 3–15. https://doi.org/10.1027/2151-2604/a000509

Centre for Reviews and Dissemination (UK). (1995). *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews.* https://www.ncbi.nlm.nih.gov/books/NBK285222 . Accessed 4 January 2023.

Eisele-Metzger, A., Bond, M., Gopal, I., Lieberum, J., Mansilla, C., Siemens, W., … Metzendorf, M. (2026, February 23). Large Language Models to support conducting systematic reviews: Protocol for a living scoping review. https://doi.org/10.17605/OSF.IO/7QHM2

Farotimi, O., Dunn, A., Van Lissa, C. J., Polanin, J. R., Mavridis, D., & Pigott, T. D. (2026). Guidance for manuscript submissions testing the use of generative AI for systematic review and meta-analysis. *Research Synthesis Methods*, *17*(2), 237–239. https://doi.org/10.1017/rsm.2025.10058

Flemyng, E., Norl-Storr, A., Macura, B., Gartlehner, G., Thomas, J., Meerpohl, J. J., Jordan, Z., Minx, J., Eisele-Metzger, A., Hamel, C., Jemioło, P., Porritt, K., & Grainger, M. (2025). Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence 2025. *Cochrane Database Systematic Reviews*. https://doi.org/10.1002/14651858.ED000178

Forsström, S., Bond, M., & Njå, M. (2026). *A meta review of programming and robotics in schools*. *Review of Education*, 14, e70231. https://doi.org/10.1002/rev3.70231

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd edition). Sage Publishing.

Jansen, T., Liebenow, L. W., Schaller, N.-J., Hattie, J., & Möller, J. (2026). Automated Data Extraction by Large Language Models: Assessing Accuracy in Comparison to Human Experts Using the Example of Visible Learning. *Educational Psychology Review*, *38*(1), 38. https://doi.org/10.1007/s10648-026-10136-5

Karlstrøm, H. (2024). Uses of artificial intelligence and machine learning in systematic reviews of education research. *London Review of Education*, *22*(1). https://doi.org/10.14324/LRE.22.1.40

Lazaros, K., Vrahatis, A. G., & Kotsiantis, S. (2026). Human-in-the-loop artificial intelligence: A systematic review of concepts, methods, and applications. *Entropy*, *28*(4), 377. https://doi.org/10.3390/e28040377

Martin, F. (2025). *Systematic reviews in educational research: A practical guide for research synthesis*. Sage Publishing.

Martin, F., Dennen, V. P., & Bonk, C. J. (2023). Systematic reviews of research on online learning: An introductory look and review. *Online Learning Journal*, *27*(1), 1-14.

Scheelbeek, P., Bond, M., Callaghan, M., Minx, J. C., Hadida, G., O'Mara-Eves, A., Thomas, J., McDonald, S., & Schuster, T. (2024). *Digital evidence synthesis tools for climate & health.* The Wellcome Trust. https://doi.org/10.21955/wellcomeopenres.1115400.1

Seifi, A., & Seyfi, A. (2026). Hallucination Rate of Peer-Reviewed Citations Generated by Large Language Models in Neurocritical Care. Critical Care Explorations, 8(9), e1474. https://doi.org/10.1097/CCE.0000000000001474

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J. et al. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both *BMJ,* 358. https://doi.org/10.1136/bmj.j4008

Thomas, J., Flemyng, E., Noel-Storr, A. (2026). Responsible AI in Evidence SynthEsis (RAISE): guidance and recommendations. https://doi.org/10.17605/OSF.IO/FWAUD

West, R. E., & Martin, F. (2024). What type of paper are you writing? A taxonomy of review and theory scholarship distinguished by their summary and advocacy arguments. *Educational Technology Research and Development*, *72*, 2443–2476. https://doi.org/10.1007/s11423-023-10233-0

Xiao, J. (2018). On the margins or at the center? Distance education in higher education. *Distance Education*, *39*(2), 259–274. https://doi.org/10.1080/01587919.2018.1429213

Zawacki-Richter, O., Cefa, B., & Bai, J. Y. H. (2025). Towards reproducible systematic reviews in Open, Distance, and Digital Education—An umbrella mapping review. *Review of Education*, *13*(1), e70031. https://doi.org/10.1002/rev3.70031

Zawacki-Richter, O., & Jung, I. (Eds). (2023). *Handbook of Open, Distance, and Digital Education*. Springer. https://link.springer.com/referencework/10.1007/978-981-19-0351-9

Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (Eds). (2020). *Systematic reviews in educational research: Methodology, perspectives and application*. Springer. http://link.springer.com/10.1007/978-3-658-27602-7