ONLINE LEARNING
THE OFFICIAL JOURNAL OF OLC

SPECIAL ISSUE

Evidence Synthesis in Open, Distance, and Digital Education

NOVEMBER 2026

E-ISSN: 2472-5730

ONLINE LEARNING

Volume 30 · Issue SI2 · 2026

Contents

Evidence Synthesis in Open, Distance, and Digital Education

e6749 Towards Responsible AI-Augmented Evidence Synthesis in Open, Distance, and Digital Education (ODDE)

e5641 Who Is AI Supporting? A Systematic Review of AI-Enabled Learner Support in Higher Education

e5691 A Meta-Analysis of AI-Assisted Self-Regulated Learning

e5742 Online Learner Success Dimensions and Factors: A Framework Development Through Systematic Review

e5747 The Integration of Generative Artificial Intelligence Into Tertiary Health Education: A Scoping Review

e5764 Mapping Evidence on Persistence and Retention Strategies for Non-Traditional Students in Online Higher Education: A Scoping Review

e5781 Fostering Critical Consciousness Online: A Systematic Review of Strategies, Opportunities, and Gaps in the Literature

e5809 Modality Matters to Parents Who Support Adolescent Online Learners: A Literature Synthesis

e5816 Refining Collaborative AI Literacy: A Systematic Review of Human–AI Collaboration Competency and AI Literacy Measures

e5820 Revisiting Learning Presences in the Community of Inquiry Framework: Insights from a Systematic and Meta-Analytic Review

e5822 Research quality moderates the relationship between student motivation and academic performance in online and blended courses: A meta-analysis

e5823 Verification Without Criteria: An Epistemic Network Analysis of Metacognitive Laziness in Generative AI Research

e5880 Generative AI and the Community of Inquiry Framework: A Scoping Review of Teaching, Social, and Cognitive Presence

e5893 Systematic Review of Retention and Learning Engagement in Online Learning from 2015–2025

e5895 Managing Overlapping Review Studies in a Second-Order Meta-Analysis: Describing and Exploring Existing Strategies

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

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5641
https://doi.org/10.24059/olj.v30iSI2.5641

Who is AI supporting? A systematic review of AI-enabled learner support in higher education

Berrin Cefa
Olaf Zawacki-Richter
Felicitas Macgilchrist
Carl von Ossietzky Universität Oldenburg

Abstract

Learner support is a cornerstone of open, distance, and digital education (ODDE), yet providing effective, scalable, and equitable support remains a persistent challenge. With the rise of artificial intelligence (AI) and increasing publications on it, it is important to see how higher education institutions are adopting AI-enabled tools to support their learners. This study systematically reviewed and synthesized 61 publications to examine how AI has been applied to learner support, what practices and constraints are reported, and whether current approaches meet diverse learner needs. The included studies, identified without applying time restrictions, span publications from 2010 to 2025. Using the Inclusive Support for Student Persistence Model (ISSPM) and an AI typology, the review finds that adaptive systems and personalization, particularly chatbots and recommender systems, are the most common applications, followed by profiling and prediction. Reported benefits include timely interventions, personalized guidance, and workload reduction for staff; however, technical limitations, privacy concerns, limited inclusivity, and reduced human interaction pose significant challenges. The review finds critical blind spots, especially in addressing diverse learner needs and in the underexplored transition phase of the student journey. Notably, diversity in population, such as marginalized groups or neurodivergent learners, was absent. By offering a structured synthesis from an institutional perspective, this review contributes to the established learner support research about the role of AI in shaping equitable and sustainable learner support in digital higher education.

Keywords: Learner support, Artificial intelligence, Student services, Digital education, Online education, Distance education, diversity

Introduction

Learner support is essential for learner success, retention, and a well-rounded learning experience (Moore & Kearsley, 2012; Rotar, 2022a; Simpson, 2004, 2013), particularly when the learner is detached from the time and space of instruction (Brindley, 2014). Open and distance education (ODE), in this sense, has long faced low student retention rates, and academic support interventions have continued to be one of the major concerns of the field (Cefa & Zawacki-Richter, 2025; Zawacki-Richter & Bozkurt, 2023). Learner support has been critical for learner success in distance education (Brindley & Paul, 1996; Moore & Kearsley, 2012), especially for learners getting engaged with online learning (Rumble, 2000). The flexibility and independence in time and place of open and distance learning are widely adopted by conventional higher education institutions, but in more digital settings to reach out to their students (Xiao, 2018). This shift accelerated during the global emergency remote teaching (ERT) period (Bozkurt & Sharma, 2020; Hodges et al., 2020), when institutions worldwide moved online, often without preparation. Conventional universities were suddenly immersed in ODE challenges. The widely enabled online practices and digitally enhanced setting also in conventional universities brought a new term encapsulating all forms of HE as defined by Zawacki-Richter and Jung (2023), as an umbrella term: open, distance, and digital education (ODDE), encompassing digitally enhanced education also in conventional education, including technology-enhanced learning or blended learning, laying their roots in ODE practices.

Corresponding author: Dr. Berrin Cefa, berrin.cefa@uol.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)

The expansion of ODE in the 1990s spurred more research on student support due to quality problems that emerged with the growth of distance teaching institutions (Zawacki-Richter & Naidu, 2016). In the 2010s, with the rise of artificial intelligence (AI) and big data, attention turned to intelligent support systems (ITS) (Zawacki-Richter & Bozkurt, 2022). The expansion of access to higher education and the increasing digitalization of learning, first through ERT, then through ongoing technological advances, have made ODDE a reality for almost all universities (Arnhold & Bassett, 2021). Yet, despite its flexibility, ODDE demands strong motivation, self-regulation, time and task management, goal orientation, and the ability to work independently and collaboratively (Brindley, 2014). Learner support in ODDE will therefore remain crucial for helping students succeed, whether in technology-enhanced classrooms or fully online environments.

Providing effective learner support at scale is challenging for ODDE institutions serving geographically dispersed and diverse populations (Daniel et al., 2009). It requires careful cost analysis, substantial administrative resources, robust systems (Rumble, 2000), although it might prove more cost-effective for the university in the long term (Burke, 2019; Simpson, 2023). To streamline these support processes, the advent of the Internet played a key role (Rotar, 2022b) in enabling more personalized and automated support mechanisms (Dollinger et al., 2020; Walsh et al., 2020) alongside broader communication paths to be utilized (Zawacki-Richter & Anderson, 2014).

More recently, the techno-feudal rise of big tech companies making generative artificial intelligence (GAI) publicly accessible at scale fueled new learner support possibilities despite criticism (Cefa et al., 2025). Already, California State University system proudly announced integrating ChatGPT Edu across 23 campuses, serving over 460,000 students and 63,000 faculty at no cost with the promise of offering personalized tutoring, study guidance, and research help (DiPierro, 2025).

At this stage, we aim to critically explore the state-of-the-art in higher education learner support systems and pinpoint the AI-powered support interventions being applied or designed to help learners by means of a systematic literature review. Therefore, this study addresses the following research questions:

RQ 1: What is the scope of research on AI use in higher education for learner support?

RQ 2: What practical experiences and constraints are documented in the literature?

RQ 3: To what extent do current AI-driven learner support services address the diverse needs of learners?

Notably, no directly comparable systematic reviews focusing specifically on AI-enabled learner support in higher education were identified through our search strategy, highlighting a significant gap in the literature and the need for a structured synthesis of this emerging field. By this review, we do not only demonstrate the current institution-based systemic AI applications and their functions in learner support according to the learner journey, but also critically engage with the homogeneity of learner profiles in the datafication process.

Conceptual Framework and Literature Review

Learner Support in ODDE

Learner support is recognized as an institutional dimension (Cefa & Zawacki-Richter, 2025) defined through the relationship between the institution and the learner. Consequently, the responsibility of higher education institutions is to delineate the areas of "service" provided to students. Tait (2000) conceptualizes three interrelated dimensions: cognitive, affective, and systemic activities. Cognitive support focuses on facilitating students' mastery of course content and the development of academic skills; affective support addresses the emotional and motivational aspects of learning; and systemic support encompasses administrative, technical, and organizational processes. Accordingly, "the main institutional systems involved in learner support are teaching and tutoring, advising and counselling, and information and administrative" (Brindley et al., 2004, p. 13). Building on this conceptualization, this study adopts the definition by Cefa (2025):

> […] The learner support ecosystem constitutes communication that allows learners to self-direct, find guidance, connect and build a network, and engage with resources (human or inhuman) across cognitive, affective, and systemic domains, while pursuing diversity, inclusivity, and mitigating the historical education divide (p. 12).

To achieve this, the timing of the support intervention matters. Thorpe (2003) distinguishes support before, during, and after learning; Simpson (2013) differentiates reactive (after difficulties emerge) and proactive (before disengagement) approaches. As with most sociological structures, it is challenging to divide multifaceted systems into clearly separated layers. Tait (2000) highlights six design factors: student characteristics, programme demands, geography, technology, scale, and management. Simpson (2002) separates academic (course mastery, feedback, learning skills, enrichment) from non-academic (advising, practical help, advocacy, administration) support. Floyd and Casey-Powell's (2004) Inclusive Student Services Process Model (ISSPM), developed for online education, aligns support with five phases of the learner journey:

(1) Learner intake: initial recruitment, orientation, and preparation for study;
(2) Learner intervention: targeted actions to address emerging academic or non-academic issues;
(3) Learner support: ongoing assistance, both academic and administrative, throughout the study period;
(4) Learner transition: guidance for progression, completion, or movement to further study or employment; and
(5) Measurement: systematic evaluation of support effectiveness to inform continuous improvement. By structuring interventions along this timeline, the model enables institutions to plan proactive and integrated support strategies that respond to students' evolving needs.

Rotar (2022a), using the ISSPM as a framework for her systematic review on the effectiveness of online academic student support, emphasizes its usefulness as the model outlines the key tasks associated with each stage of the student learning journey.

Before addressing the methodology of this study, the following section will first provide a conceptual overview of the significance and scope of learner support in ODDE, with particular attention to emerging AI applications.

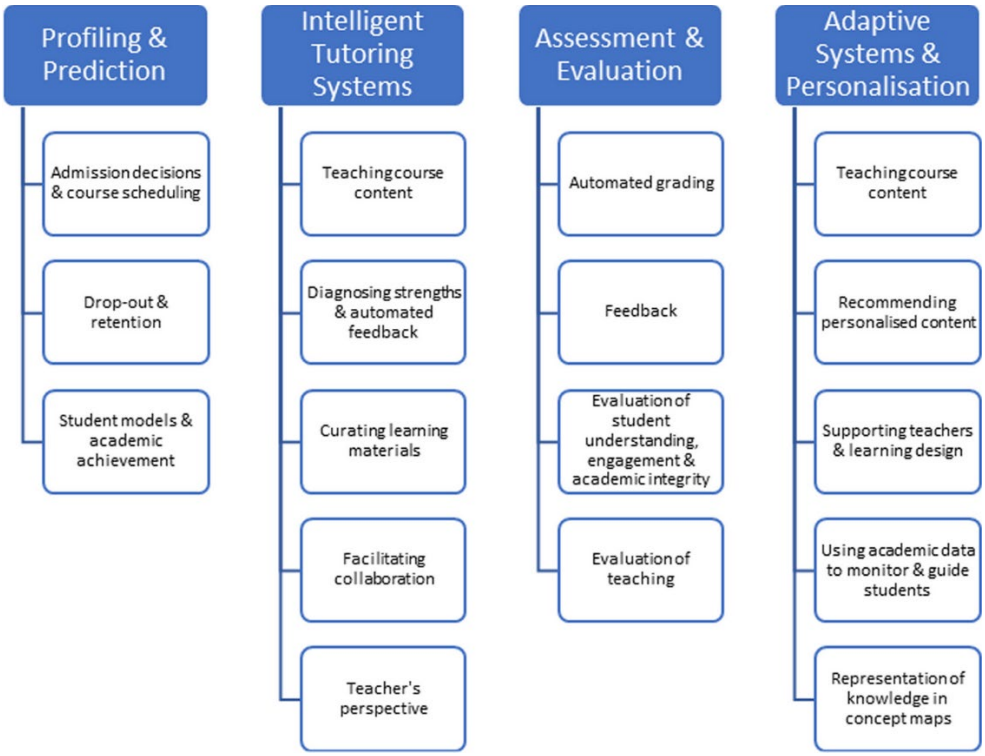
AI in Education (AIEd) and Learner Support

The application of AI for learner support is not a recent development; its roots extend back several decades, with early systems exploring automation in tutoring. The first ITS reported was the SCHOLAR system, launched in 1970, which allowed the reciprocal exchange of questions between teacher and student, but did not hold a continuous conversation (Welham, 2008). However, the growing availability of computing power, data, and advances in machine learning have expanded the scope and precision of such interventions, prompting renewed examination of their role in supporting students.

AI is a complex theme with a lot of technicalities. To keep the focus on learning and education rather than technical developments, the literature is discussed under the typology developed by Zawacki-Richter et al. (2019), with more recent developments in large language models (LLMs) located within the fourth area, 'adaptive systems and personalization' (Figure 1).

Figure 1

AIEd typology by Zawacki-Richter et al. (2019), visualized by Bond et al. (2024, p. 4)



Profiling and prediction encompass predictive models that use student data to predict and profile students on their performance, prospective program completion rates, and time; more importantly, the risk of dropping out (Jayaprakash et al., 2014). With the advancement of AI as well as the wide usage of LMS, especially in ODDE, a variety of student data became available, which opened the path for learning analytics (Joksimović et al., 2015; Siemens, 2013). Despite the long history of prediction-based analysis, it was not in use in education until 2010 (Joksimović et al., 2019). Learning analytics is defined as "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long & Siemens, 2011, p. 34). Through advanced computational techniques such as Bayesian modelling, cluster analysis, and predictive learning analytics (Herodotou et al., 2019), institutions provide learner support (Joksimović et al., 2019) with early prediction of students' course or program completion and academic performance (Calvert, 2014; Gašević et al.,

2015; Tempelaar et al., 2021). The predictive power enables institutions to optimize learning (Macfadyen & Dawson, 2010), and identify patterns that should lead to taking proactive action to change the pattern (Duval, 2011), as well as personalize the learning process (Chatti et al., 2012).

ITS are pedagogically designed non-human tutoring systems that deliver domain-specific information with step-by-step guidance. Huang and Chen (2016, p. 341) outline several models commonly incorporated into ITS: the student model, which contains information on the learner's knowledge level, e.g. cognitive abilities or motivation; the teacher model, which analyzes students' current status, selects suitable teaching strategies and methods, and offers help and guidance; and the domain model, which represents the knowledge held by both students and teachers; and the diagnosis model, which evaluates errors and deficiencies based on the domain model.

Assessment and evaluation are fields largely focus on teaching and learning processes, mostly focusing on feedback, automated assessment, or teaching evaluation (Bai et al., 2022; Zawacki-Richter et al., 2019). This category utilizes mostly training models through various algorithms such as supervised learning.

Adaptive systems and personalization, the fourth category, is another broad term without an agreed definition. The educational activities compose creation of personal content, recommender systems, monitoring and guidance during learning. Within this study, chatbots were also taken under this category with particular reference to its function in support systems. Actually, starting from the 1950s, continuing with ELIZA in 1966 by Weizenbaum (Han & Lee, 2022), chatbots has long been the most well-known tool of adaptive systems. The emergence of generative AI in the public domain, particularly through widely accessible chatbot interfaces, has sparked significant interest across various fields, including education, with the hope that there is a system capable of supporting learning, providing personalized content, and delivering timely feedback (Cefa et al., 2025; Jensen et al., 2024; Lo, 2023). Chatbots, however, vary in their underlying architectures, interaction styles, and capacities for adaptation, ranging from rule-based dialogue systems to large-scale, transformer-based generative models (Arun et al., 2019; Suta et al., 2020). These differences influence not only the quality and accuracy of responses but also the degree to which the systems can tailor interactions to individual learner needs, adapt to contextual cues, and maintain pedagogical coherence over time. As such, understanding the spectrum of chatbot capabilities is crucial for evaluating their educational potential and for aligning their use with instructional goals and ethical considerations.

By rigorously compiling the state of the art in the literature on AI-enabled learner support, this review seeks to offer an evidence-based account of the current landscape in both research and practice. In doing so, it examines not only areas of promise but also blind spots and specific domains, including approaches to the preservation of pedagogical integrity and data privacy.

Methods

Research Method

This study employed a systematic review to provide an overview of the literature on learner support at the institutional level that leverages AI. Systematic reviews are widely recognized as a preferred methodology in ODDE, particularly in light of the rapid growth in publications in the field, especially concerning AI (Bond et al., 2024; Zawacki-Richter et al., 2025). Their methodological rigor lies in the use of replicable and reproducible procedures, implemented through a series of predefined steps (Buntins et al., 2023; Gough et al., 2017; Zawacki-Richter et al., 2025), enabling the results to inform evidence-based discussions. The following sections present a detailed account of the

procedure, which is fully documented in the review protocol published on the Open Science Framework (OSF) platform (see: https://osf.io/cpwdz).

Data curation and criteria setting

In the initial phase of the study, the Population, Intervention, Comparison, and Outcome (PICO) (Page et al., 2021) framework was established to define the research scope, as outlined in the introduction. Based on this framework, the initial search string was developed and iteratively refined by incorporating alternative synonyms for "learner support" and adjusting the scope of AI applications (Table 1).

Table 1

Final search string

Topic	Search string
Study focus	("student support" OR "learner support" OR "student service" OR "learner service" OR "support intervention" OR "academic services")
	AND
Intervention	("artificial intelligence" OR "AI-driven" OR "AI-powered" OR "generative AI" OR intelligent OR virtual OR smart OR "recommender system" OR "recommendation system" OR personalized OR adaptive OR automated OR chatbot OR "reinforcement learning" OR "learning analytics" OR predictive)
	AND
Education level	("higher education" OR university OR college OR post-secondary OR "tertiary education" OR "vocational education")

The exclusion and inclusion criteria were set (Table 2) based on the research questions. Despite AI's relatively long history and review studies on particular AI applications for particular purposes (e.g., Labadze et al., 2023; Larrabee Sønderlund, 2019), there is no overview of the various uses of AI in student support in higher education. Therefore, no time restrictions were applied, ensuring the inclusion of all relevant studies. Although analytics and algorithm development studies were included, studies that use synthetic data were excluded as they do not include real student data and remain purely technical for the research questions this study seeks to answer.

Only peer-reviewed journal articles and papers in conference proceedings were included, while books, editorials, and other forms of grey literature were excluded to ensure a consistent methodological comparability across peer-reviewed studies in an exponentially increasing trend of paper production on AI.

Table 2

Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Time	No limit	-
Education setting	Higher education	Early childhood education, K-12, non-formal education, informal education
Intervention	AI-powered, non-human mediated support (including analytics, predictive models)	No AI-powered support OR support is not the purpose for the use or development of AI
	Applicable by the institution and not subject-domain specific.	Support is a secondary topic in the application.
Document type	Journals, conference proceedings	Books, dissertations, booklets, reports, editorials, commentaries, letters
Language	English	Not in English

Both learner support and AI have blurred lines as they are multi-faceted phenomena. If the intervention was not applicable at an institutional level but course-specific on the micro level of learning and teaching, these studies were excluded.

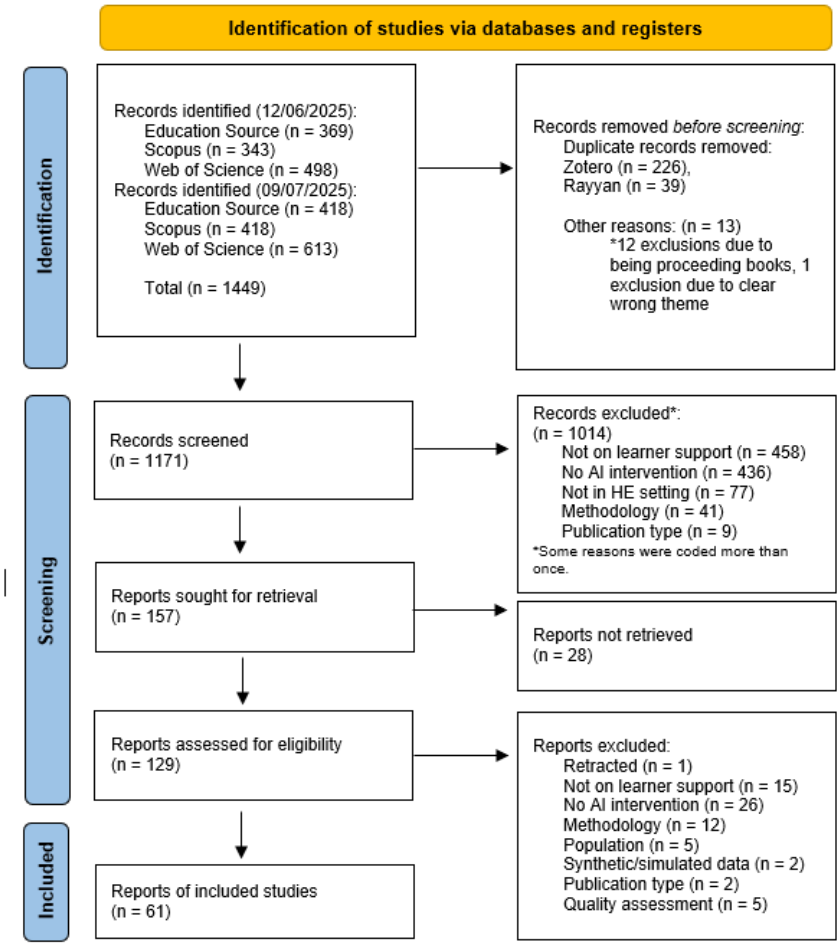
Screening and coding process

The final search was conducted on June 12, 2025, in three databases: Education Source, Scopus, and Web of Science (WoS). They were selected to ensure wide coverage of education research. Together, these databases are widely recognized as providing broad and complementary indexing of peer-reviewed literature relevant to higher education and educational technology.

On July 9, the additional search term “predictive” was introduced, and the search was repeated across the same databases. Language and document-type filters were applied in all three databases, and to ensure consistency with Scopus and WoS, the Education Source search was restricted to “peer-reviewed” publications only. As the PRISMA flow chart displays in Figure 2, out of 1449 articles identified, the phases of duplicate cleaning, title and abstract screening, and full text screening were followed, finally forming a corpus of 61 included studies.

Figure 2

PRISMA flow diagram (adapted from Page et al., 2021)



The initial search results were imported into the reference management system Zotero for preliminary duplicate removal. For title and abstract screening, 1,212 references were uploaded to the systematic review software Rayyan. Additional duplicates were identified using the software's automated detection feature, set to a 0.90 resemblance threshold, which found and merged 39 further duplicates.

Two rounds of blind title and abstract screening were conducted by the first and second authors, with reconciliation meetings to resolve discrepancies. This process yielded a Cohen's kappa (Cohen, 1960) value of 0.80, indicating substantial interrater reliability. For full-text screening, the same blind screening process was applied, necessitating two rounds for reconciliation and resulting in excellent agreement between two reviewers (Round 1: $\kappa = 0.70$; Round 2: $\kappa = 1.00$).

Data extraction and coding were conducted using EPPI-Reviewer Web (https://eppi.ioe.ac.uk/EPPIReviewer-Web/home), a dedicated tool for systematic reviews. The review focused on two complex, multi-dimensional phenomena: learner support and AI. Each topic required careful categorization due to its diverse and layered nature.

Given the scope of the systematic review, which involved analyzing dozens of studies with varying purposes, methodologies, and theoretical frameworks, it was essential to adopt a robust and

reliable approach to data interpretation. The coding process was deductive, shaped by two relevant theoretical models to ensure coherence and depth in the synthesis. This approach ensured that the extracted data would be both meaningful and easily transferable.

For the learner support interventions, the ISSPM by Floyd and Casey-Powell (2004) was adopted. This model includes five stages of institutional support aligned with the student learning journey, from initial enrollment to graduation (see Section 1). The sub-codes were adapted where necessary to describe the study's content.

For the AI dimension, the typology proposed by Zawacki-Richter et al. (2019) was used. This framework categorizes AI applications in education into four broad areas (see Figure 1), with subcodes adapted to suit the context of the reviewed studies. AI terminology often varies significantly across disciplines, making a unified classification challenging; the use of this typology helped to standardize the analysis despite such variability.

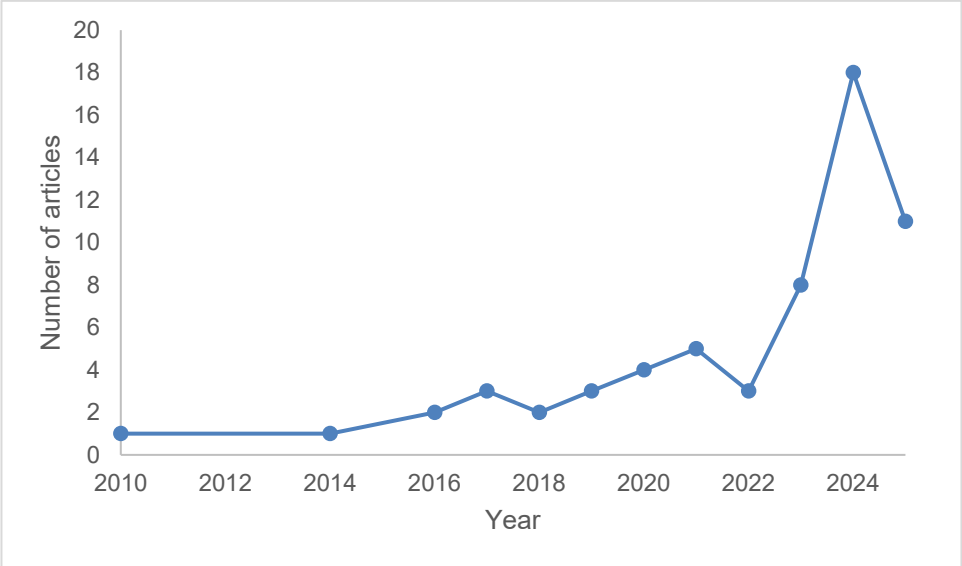
Study and data characteristics of the corpus

Disciplinary distribution of the studies in the corpus, categorized according to the OECD Field of Science and Technology (FOS) classification, shows a predominance of works in the Natural Sciences (n = 29), followed by the Social Sciences (n = 15). A smaller number of contributions came from Engineering & Technology (n = 9), Medical & Health Sciences (n = 1), Humanities & Arts (n = 1), and other fields (n = 5), which are administrative support divisions of universities, with two studies not being traced back to the disciplinary affiliation. Within the Social Sciences, most studies were classified under Educational Sciences (n = 10). In the Natural Sciences, the largest share fell under Computer and Information Sciences (n = 27), followed by Mathematics (n = 4) and other Natural Sciences (n = 1).

Although no time limit was set, the earliest publication in the corpus dates back to 2010 only and follows a steep curve in the last three years (Figure 3).

Figure 3

*Publication trend according to years (N = 61) *



To encompass a wide range of perspectives, the corpus included studies that either implemented AI-driven learner support or developed AI models for such support. Methodologically, these were coded as either Technical/Algorithm-focused (n = 20) or Intervention/Educational (n = 38), with three studies spanning both categories. If the study technically engaged in the development process, these studies were coded as technical papers. However, the studies that describe the intervention model with a practical focus were coded under intervention studies. The details are given in Table 3.

Table 3

Detailed description of technical / algorithm-focused studies

Data source	n
Institutional records	16
LMS Data	6
Participants from institutions	5
Public open dataset	4
Commercial products	2
Not specified	3

The technical or methodological papers dominantly source data from institutional records. The four studies that did not specify the data source, specified that they used institutional historical data, but as the records were not collected for the study and details are unclear, they were coded as N/S.

Consent statements for data sources were reported in only six studies (27.3%) (e.g., Acang et al., 2023; Ajoodha et al., 2020; Balderas et al., 2023; Bhave et al., 2024; Channarukul et al., 2017; Gray & Perkins, 2019; Hill et al., 2025). This code did not apply to two studies since they used a public data set.

Table 4

Methodological approach of Intervention / Educational Studies

Sub-subcategory	n
Intervention / Model description	13
Qualitative	3
Quantitative	14
Mixed method	12

Almost one third (n = 13) of the Intervention/Educational studies category (n = 41) came from intervention/model description papers. The empirical papers were predominantly quantitative or mixed-methods studies. Of the quantitative studies, two tested the model and were consequently coded under both the *intervention* and *quantitative* categories. Only three studies employed a qualitative design.

Across the corpus, the primary population group was students (n = 55), representing the dominant focus of AI-driven learner support research. Smaller shares focused on faculty members (n = 7) and administrative staff (n = 5). Four studies did not specify a participant group as they described the developed product without user data. The “Other” category (n = 2) encompassed studies that used developers as their data source (e.g., Garcia et al., 2023; Gonçalves et al., 2022).

Finally, the distribution of the corpus by educational setting was heavily weighted toward conventional campus-based institutions (n = 41, 67.2%), followed by open and distance education institutions (n = 13, 21.3%) and dual-mode universities (n = 3, 4.9%). In four cases (6.6%), the institutional type could not be identified.

Limitations

Every study is inevitably shaped, thus limited by the methodological and contextual choices it makes. This research is limited to English-language publications, which inherently restricts the representation of scholarly work produced in other languages. As a result, insights from non-English speaking contexts, particularly from countries with robust traditions in open and distance education, may be underrepresented. Given that learner support is mostly handled as a meso-level, namely, an administrative matter, related research might frequently be published in national or regional journals in the home language.

The exclusion of dissertations and other forms of grey literature may have introduced publication bias, as studies reporting non-significant, exploratory, or critical findings are less likely to appear in peer-reviewed journals, a potential limit to the comprehensiveness of the corpus.

Moreover, the scope of the study is confined to three academic databases. While each of these platforms is recognized for its credibility and wide coverage, the selection inevitably introduces some degree of bias. In parallel to that, more critical voices are more likely to prefer publishing in books, edited volumes, and blogs. Relevant literature published in less prominent or region-specific journals as well as other publication areas are excluded.

Finally, no formal quality assessment of the included studies was conducted. Nevertheless, studies lacking a clearly described methodology were excluded during full-text screening to ensure a minimum level of methodological transparency.

Results

RQ 1: What is the scope of research on AI use in higher education for learner support?

In line with the ISSPM model's five phases of learner support, this study examined how AI technologies were applied across these phases. The findings indicate that AI use was most prominent in the Learner Support Phase, followed by the Measurement, Learner Intake, and Learner Intervention phases. No AI-driven interventions were identified in the Learner Transition Phase. For an overview and meaningful interpretation in the following sections, Table 5 displays the cross-tabulation of the AI intervention with five learner support phases.

Table 5

Co-occurences of AI-driven intervention and support phases

AI intervention/Support Phase	Intake	Intervention	Support	Transition	Measurement
Profiling & prediction	1	-	9	-	20
Assessment & evaluation	-	-	1	-	-
Adaptive systems & personalization	12	14	14	-	-
Intelligent tutoring systems	-	-	1	-	-

Table 5 depicts the weighted use of AI-driven interventions in the Learner Support and Measurement Phases. As one tool might be used in the same paper for more than one service, Table 6 shows the total uses of the subcategory of AI typology.

Table 6

Frequency of AI applications for learner support

Code	Count
Adaptive systems & personalization	41
Profiling & prediction	30
Intelligent tutoring systems	1
Assessment & evaluation	1

As Table 6 shows, the most frequent uses are in adaptive systems & personalization, closely followed by profiling and predictions. In the following subsections, we examine how each subfield of the AI typology was applied across the corpus, following the learner's journey from the Intake phase to the Measurement phase.

Adaptive systems & personalization (AS&P) in learner support phases

AS&P is the most used type in learner support phases, all in the first three phases (Table 7).

Table 7

Co-occurrences of AS&P uses in learner support phases

Codes	Intake	Intervention	Support	Transition	Measurement
Chatbots-LLM-NLP	7	11	8	-	-
Recommending personalized content/program	3	1	4	-	-
Monitor & guidance	1	1	3	-	-
Robots	-	1	-	-	-
VR, Simulations (AI-based)	1	-	1	-	-

As the table presents, the most frequently used AI technology is chatbot, followed by recommender systems, and a few examples of monitoring, robots and VR.

The Learner Intake Phase benefited from AS&P, via chatbots for homepage support (n = 3; e.g., Abdul Razak, 2024; Prasertsri et al., 2024; Sereni, 2024), admissions/registrations (n = 2; e.g. Pansombut & Kirimasthong, 2025; Thongda & Songpan, 2023), and financial aid information (n = 1; e.g., Nurshatayeva et al., 2021).

Orientation is one of the key services in the Learner Intake Phase, and this includes physical and cognitive orientation of the learner in the system as well. For campus orientation, Garcia et al. (2024) initiated a virtual tour connected to the metaverse system of FEU Institute of Technology in the Philippines during the pandemic. Sereni (2024) demonstrated an effective example of an integrated international support service structure (advisory, social programs, communications) in Australia. They utilized a self-guided campus tour via a scripted chatbot to make the tours scalable and accessible. During orientation week, students log into the "Fantastic Race" app with their Uni ID, join a randomized team and complete location-based challenges. The app itself was assembled with Power Apps (interface), Power Automate (flows to SharePoint), and Excel (scores). Drawing on Copilot, the

team generates the step-by-step wiring instructions that made these pieces talk to each other smoothly. Data is secured in SharePoint; rankings update automatically; access is Uni-login only. Over 600 students have participated since its Semester 2, 2023 launch, with very positive feedback. Another chatbot was utilized as a cognitive agent, called "BARICA" at the University of Zagreb in Croatia to provide information in English on faculty location, working hours of the personnel, schedules, particularly aiming first year students (Schatten et al., 2021).

In this respect, the initial information provision is the homepage of a university. As the initial point of contact between the institution and students, the website can sometimes be confusing for newcomers or may lack essential information. For cognitive orientation, chatbots (n = 4; e.g. Abdul et al., 2024; Nurshatayeva et al., 2021; Sereni, 2024), recommendation systems (n = 1; Acang et al., 2023) and AI-based VR (n = 1; Garcia et al., 2023) were used to provide orientation and guidance via the institutional website.

Similarly, Prasertsri et al. (2024) implemented a frequently asked questions answered by a chatbot via LINE OA to avoid answering repetitive questions without contacting the university departments in Thailand. The users have to use the LINE application with their accounts. Two components: (1) a Dialogflow-based NLP chatbot; (2) LINE "Rich menu" shortcuts for direct access to services. Users add the faculty's LINE OA, receive a welcome message, and can either chat (free-text) or tap Rich-menu shortcuts to reach external resources (faculty website pages, social media, downloadable forms, Google Maps, complaint portal). Dialogflow provides intent modeling, language understanding, memory of key terms, and continual improvement via machine learning; the chatbot supports bi-directional conversation and responds with text, images, and links. Another chatbot was developed by Abdul et al. (2024) using NLP and ANN trained on JSON dataset with 80 % accuracy to increase the accessibility to information by students with immediate responses, thus avoiding workload on support.

A further service in Intake phase was financial aid. It is a key factor contributing to the 14% summer melt rate among U.S. college-intending high school graduates, alongside other overwhelming pre-enrollment tasks such as financial aid paperwork, placement tests, and course registration (Nurshatayeva et al., 2021). Without adequate family support or sufficient financial resources, these processes become even more challenging. Nurshatayeva et al. (2021) evaluated the "PeeDee" chatbot at East Carolina University, which sent tailored reminders, provided information, and answered questions in real time to all ECU-intending students. While the overall outreach increased student loan acceptance by 4 %, further analysis showed that first-generation students, who began with low initial completion rates for pre-matriculation tasks, experienced more gains (See RQ3 for more details).

A course registration recommender model was developed to "[…] profile the strengths of the education providers and the possible programs that students will endure" (Acang et al., 2023, p. 2). Thus, students were aimed to be placed the learners on the right course.

In the **Learner Intervention** phase, institutions primarily deployed chatbots as help-desk tools for administrative queries. Development of chatbots centered on applications and models aimed at reducing support-staff workload, indicating a need from technology to cater students at scale. Examples include the LA-K Telegram chatbot (Cahyana et al., 2024); a K-NN–based classification chatbot introduced as an alternative to WhatsApp exchanges between students and desk support (Hardi et al., 2022); a chatbot built with the ChatterBot library to reduce response delays in online learner-support systems (Ndunagu et al., 2025); an LLM-based chatbot to help academic advisors for extracting frequently asked questions (Wangwiwattana & Jantarick, 2024)

Öncü and Süral (2024), for instance, examined user perceptions of Anadolu University's Virtual Assistant using a sequential explanatory mixed-methods design. The keyword-based assistant, a rule-based chatbot, matched incoming queries to likely answers about administrative questions such as enrollment, tuition, course information, and related student-affairs topics. Although participants expected greater functionality (e.g., full answers rather than redirects), overall ratings for usability, information quality/security, and design were positive. Learners also expressed a desire for stronger academic-support features.

A further study, exploring GenAI for learner support, compared Anadolu University's rule-based assistant with GenAI chatbots (ChatGPT-3.5 and Bing Copilot) across four scenarios (Öncü et al., 2025). When provided with FAQs and their answers, ChatGPT-3.5 correctly matched 95.48% of previously unanswered learner questions, but its accuracy dropped to 54.24% when asked to generate full responses directly from that dataset; by contrast, Copilot, drawing on the open web without the FAQ corpus, answered learner questions correctly 64.97% of the time. The authors recommended hybrid arrangements: keep a vetted, rule-based layer for authoritative answers, use GenAI for intent recognition/matching, and escalate to human experts when needed. Priccilia and Girsang (2024) reached a similar conclusion with a GPT-2–based chatbot: for narrow, templated student-service tasks, a generative model isn't the best choice. It's more defensible when the scope is broader, and accuracy can be improved by exploring additional data-augmentation strategies.

A different intervention utilized the hybrid structure on student success strategies via a low-tool-required open-source Robotic Study Companion (RSC). (Baksh et al., 2024). The RSC provides subject-specific interaction without performance scoring, adjusts to each learner with customized study plans, and can process spoken commands through an open-source Python speech-recognition library and the GPT-3 API.

In the **Learner Support Phase**, several studies applied chatbots for academic advising (n = 3; e.g., Lucien et al., 2024; Lytridis & Tsinakos, 2025; Sweidan et al., 2021), instructional assistance, and tutoring. Advising chatbots were designed to answer course planning queries and recommend schedules aligned with institutional policies, showing promise in streamlining decision-making for students with varied academic standing (Sweidan et al., 2021). Other work explored socially aware recommendations that integrate peer network information into course scheduling, reducing search time and supporting students' preference to enroll alongside friends (Channarukul et al., 2017).

AI-based course assistants have also been deployed to provide continuous learning support in online environments. In a study of a commercial system in the United States, such assistants were associated with higher GPA and gains in motivation and self-efficacy (Hanshaw et al., 2024). Another related research examined adaptive courseware that diagnoses learner progress and personalizes instructional pathways, showing improved pass rates and modest learning gains (Ford et al., 2017). Beyond individual support, adaptive prompts in collaborative learning environments improved task efficiency and learning outcomes compared to fixed guidance, suggesting benefits of real-time, targeted support during peer interaction (Karakostas & Demetriadis, 2014).

A small set of studies addressed mental well-being (n = 3; Alves-Noreña et al., 2025; Balderas et al., 2023; Bhave et al., 2024). These interventions include conversational agents to monitor emotional states (Alves-Noreña et al., 2025); initial support through chatbots (Balderas et al., 2023) and identifying risk behaviors and providing a point of contact between students and counseling services, particularly when staff availability is limited (Bhave et al., 2024).

Prediction & Profiling (P&P) in learner support phases

Learner support and measurement phases utilized the P&P most (Table 8), which is logical as these are the phases that allow institutions to leverage secondary data and LMS data to measure and run profiling and prediction models.

Table 8

Co-occurrences of P&P uses in learner support phases

Codes	Learner Intake	Learner Intervention	Learner Support	Learner Transition	Measurement
Drop-out & retention	1		2		11
Student performance			5		7
Course scheduling			1		
student profiles, characteristics			1		2
EWS			2		2

The Learner Intake Phase one study applied prediction and personalization (P&P) to support early detection of at-risk students. Ajoodha et al. (2020) trained a classifier to estimate student risk profiles at entry using pre-tertiary academic records and personal background data, enabling proactive support for expanding first-year cohorts in South Africa.

In the Learner Support Phase, P&P was widely used for retention and performance monitoring through learning analytics (e.g., Herodotou et al., 2020; Roestenburg et al., 2024). Herodotou et al. (2020) used predictive analytics in a randomized trial to target motivational outreach to at-risk students, improving retention by allocating intervention efforts to those most likely to benefit. Similar applications linked early warning systems to instructional support and tutoring (Khan et al., 2025), or delivered digital nudges that increased persistence among vulnerable first-year students (Rodríguez et al., 2022). Khan et al. (2025) reported five cases from Saudi universities where a predictive analytics platform enabled early retention interventions by providing instructional support and tutoring. Similarly, Rodríguez et al. (2022) implemented a nudging system for first-year students at risk of failure or dropout, which significantly improved retention by enhancing students' sense of guidance and support.

Several studies also highlighted the role of analytics in informing educators. At the Open University, weekly risk predictions guided instructors toward students unlikely to submit the next assignment (Herodotou et al., 2019), while Honson et al. (2024) reported that early alerts reduced barriers to timely instructor outreach. Analytics also supported academic counselling through tools providing transparent program and grade information, improving decision-making about course loads (Henriquez et al., 2024). Extensions to library services showed that integrating resource-usage data into institutional dashboards correlated with higher academic performance (Jantti & Heath, 2016).

One study applied P&P to well-being. Thesen (2025) clustered self-reported data from U.S. medical students to identify anxiety and depression risk groups, showing the potential of unsupervised models for directing students to support services.

Measurement, unsurprisingly, was the most common phase for P&P use. Research across phases focused on predicting retention and dropout (e.g., Ajoodha et al., 2020; Herodotou et al., 2020). Villano et al. (2018) found that student inactivity was the strongest predictor of withdrawal and that early alert systems were most accurate during the first year. Additional models have used machine

learning to identify factors affecting persistence in specific disciplinary contexts—such as medicine (Hammoudi, 2023), computer science (Li et al., 2024), and non-traditional student populations (Seo et al., 2024)—often achieving high accuracy in predicting degree completion timelines (Ndou et al., 2021).

Assessment & Evaluation (A&E) in learner support phases

Only one A&E study fell within the Learner Support phase, focusing on academic advising and performance-relevant feedback (Khumalo et al., 2023). They investigated the AutoScholar Advisor System (Auto-Ad) for student advising and giving feedback on how they perform in a South African university. The Auto-Ad is "designed to automate and optimize […] higher education process toward increased graduation rate" (p. 212). It used reflective statistics and student performance analytics to "nudge" learners toward better outcomes. It tracked semester performance, if the students would be eligible for a pass with 'good', and graduation class trajectories, providing tailored advice, actionable goals, and integration with academic advising. By showing current performance alongside areas for development, Auto-Ad encouraged self-directed learning and constructive engagement with support services, shifting attention from underperformance to achievable progress.

Despite the prominence of automated assessment in educational research, it appeared only marginally in the learner-support-focused corpus. This likely reflects that AI-based assessment is more often explored in course-specific or individual work contexts rather than within the broader learner support framework.

Intelligent tutoring systems (ITS) in learner support phases

There was also only one ITS study within the Learner Support phase. Lu (2019) focused on content-relevant automated feedback, examining the use of the automated writing evaluation system JUKU Pigai.org in English teaching in China through a mixed-methods design involving pre–post tests with 114 students and qualitative analysis of 198 students and 30 teachers. The system was valued for immediate feedback on grammar, organization, revision, and content, but students still preferred teacher input due to occasional inaccuracies and the need for a human touch.

Despite the extensive literature on ITS in ODDE, the lack of studies identified in this systematic review could be due to the high domain specificity of ITS. The transferability to larger domains is not easy (Karakostas & Demetriadis, 2014).

RQ 2: What practical experiences and constraints are documented in the literature?

Across the corpus, 55 studies reported positive experiences or implications, while 23 studies identified some constraints to be taken into consideration. These categories are not mutually exclusive, as some studies reported both positive experiences and constraints. Four studies did not specify any positive or negative experiences (Table 9).

Table 9

Practical experiences with AI-enabled support

Codes	Frequency			
	P&P	A&E	AS & P	ITS
Quick/Effective solution	3		10	2
Accessible and continuous support			8	
Tailored support	9	1	13	1
User friendly			5	

Codes	Frequency		
Less workload for support providers	1	6	1
Accurate information access	1	7	
Better learning quality	4	2	1
Early intervention	13	1	
Cost-effective		4	
Intrinsic motivation		1	

Tailored support was the most mentioned benefit (n = 24; e.g., Abdul et al., 2024; Becerra et al., 2024; Karakostas & Demetriadis, 2014; Öncü et al., 2025). It was followed by early intervention (n = 14) as well as a quick and effective solution (n = 14). With timely predictive analysis, institutions can increase retention by early interventions such as identifying students at risk with their risk levels (Villano et al., 2018); sending motivational nudges (Herodotou et al., 2020) or reminders as a warning for assignment submissions (Rodriguez et al., 2022).

Accessible and continuous support was also mentioned strongly across the corpus, which was mainly the result of facilitating chatbots. Overcoming the working hour barriers as well as faculty/staff's slow response times (Thongda & Songpan, 2023), learners could use chatbots for quick answers to frequently asked questions (e.g., Öncü et al., 2025; Prasertsri et al., 2024; Wangwiwattana & Jantarick, 2024) or to ask for help in an emergency (e.g., Balderas et al., 2023).

Timely intervention is a key factor of proactive support, and institutions seek to find a solution to help students on time and get in touch with them on the particular issue they are seeking an answer for. We know proactive student support is cost-effective for the institution as compared to reactive systems (Boyle et al., 2010; Budash et al., 2017). However, for institutions operating at scale, a human-powered mass-service structure also requires costly and complex structuring (e.g., Öncü et al., 2025). Therefore, the corpus shared the experience that a chatbot service that is available for 24/7 and answering questions accurately increased the satisfaction (Ndunagu et al., 2025; Pansombut & Kirimasthong, 2025; Karakostas & Demetriadis, 2014); created less workload for teachers (e.g. Herodotou et al., 2019; Honson et al., 2024; Lu, 2019) and for student support officers (e.g. Hardi et al., 2022; Öncü et al., 2025; Prasertsri et al., 2024).

Studies also reported the constraints as well as the inherent limitations of the current AI-enabled supports (Table 10).

Table 10

Constraints about AI-enabled support

Codes	P&P	A&E	AS & P	ITS
Data security			4	
Limited human interaction			4	1
Technical constraints	3		9	1
Need for more factor integration	4	1	2	1
Emotional distress	1			
Technology acceptance	1			

Current use of chatbots in the corpus was predominantly discussed in relation to technical constraints, including the continuous need for model maintenance and updates (e.g., Alves-Noreña et al., 2025; Balderas et al., 2023), reliance on accurate or exact language input (e.g., Lucien & Park,

2024; Lytridis & Tsinakos, 2025), inability to recognize nuanced inquiries (e.g., Pansombut & Kirimasthong, 2025; Wangwiwattana & Jantarick, 2024), and, in one case, inability to detect implicit, non-verbalized emotions (e.g., Alves-Noreña et al., 2025). Among the three studies utilizing LLMs, two (Öncü et al., 2025; Priccilia & Girsang, 2024) reported that LLM-based chatbots generated better answers overall; however, a generative chatbot was found to be non-contextual and unhelpful when domain-specific responses were required, which is a limitation attributed to the lack of integration of LLMs with institutional systems and data sources (Wangwiwattana & Jantarick, 2024).

Predictive models face constraints, mislabeling of at-risk students due to success depending on engagement in only certain activities or time periods (Honson et al., 2024); limited transferability of models as ongoing data renewal alters training across contexts (Gray & Perkins, 2019); and the risk of speculative conclusions from dashboards, even when data is validated with attention to extraction timestamps, as visualizations provide only snapshots (Chong et al., 2020).

The need for more data input was another item. While having a predictive model on using data on how frequently a student accessed a learning management system, the model would perform better if study duration was also captured (e.g., Honson et al., 2024).

Data security and privacy was also, though not frequently, a reported constraint. The importance of proper management of emotional data privacy and confidentiality is essential (Alves-Noreña et al., 2025) as especially chatbots might handle sensitive information (Ndunagu et al., 2025; Ula et al, 2023). Therefore, transparency in data processing, sharing, and storage are recorded as the essential elements (Ndunagu et al., 2025).

The absence or reduction of human interaction was identified as a further drawback (e.g., Hanshaw et al., 2024; Ula et al., 2023), with recommendations that AI interventions serve as a complement rather than a replacement (Hanshaw et al., 2024). Avoiding overreliance on AI and preserving human connection were emphasized as essential (Lu, 2019; Ndunagu et al., 2025; Ula et al., 2023). Similarly, Öncü et al. (2024) found that learners experienced a diminished sense of human engagement.

RQ 3: To what extent do current AI-driven learner support services address the diverse needs of learners?

Across the corpus, AI applications largely approached “the learner” as a uniform category. Very few studies engaged with differentiated needs such as those of neurodivergent learners, minority groups, or socio-economically marginalized students. Only two studies explicitly referenced subgroups (Candrlić et al., 2025; Nurshhatayeva et al., 2021), and only one proposed a support model centred on a disadvantaged student group (Candrlić et al., 2025).

Candrlić et al. (2025) developed a process model for support services, grounded on the framework of the United Nations Convention on the Rights of Persons with Disabilities (UN CRPD, 2006, as cited in Candrlić et al., 2025). The model maps the main activities of the Academic Accommodations Counselling unit, such as counselling, training, and reporting, through data flow diagrams. It was developed through interviews with experienced AAC staff who informed the design based on longstanding practice. The authors propose integrating AI throughout the process, including conversational tools to explain rights and services, machine learning to recommend accommodations, accessible content tools such as text-to-speech and captioning, automated scheduling for meetings and assistance, and sentiment analysis to capture user feedback.

Nurshatayeva et al. (2021) assessed a chatbot intervention for all new students at East Carolina University using institutional data. Although average use did not differ by first-generation status, the most intensive engagement came from first-generation students, who also showed lower baseline completion of pre-enrollment tasks. For this group, the intervention increased loan uptake by about eight percentage points and raised both enrollment and course registration by roughly three points compared to controls. The overall positive effects were therefore concentrated among first-generation students, consistent with evidence that summer melt disproportionately affects students with fewer resources.

Beyond these cases, the corpus did not address neurodivergent learners or minority ethnic, linguistic, or low-income student populations. This absence highlights concerns about inclusivity in AIEd, including the risk of systems optimized around majority experiences and therefore disadvantaging others (Dixon-Román et al., 2020; Selwyn, 2024). Designs premised on an imagined "typical" learner may overlook important sociocultural variation, while meaningful inclusion requires more granular data on disability status, language backgrounds, and other relevant characteristics.

However, this result wrestles with a further issue of "marginal privacy" (Skinner-Thompson, 2020). Prinsloo et al. (2022), in reference to Skinner-Thompson (2020), discuss in detail how neutral privacy protections often fail to safeguard individuals in intersectional and marginalized demographic groups. In educational contexts, highly granular data aimed at personalized support may inadvertently enable intensified monitoring or reproduce structural inequities. However, ignoring what are often called 'attributes' (demographic data) in favour of 'behaviours' (attendance, paying attention, etc.) in order to nudge learners towards behavioural change, also reproduces structural inequities, by placing sole responsibility for success on individual students (Whitman, 2020; Decuypere & Hartong, 2023).

Therefore, balancing the envisioned potential of AI-enabled support with safeguards against the entrenched biases of data-driven monitoring is essential to avoid deepening existing inequities. Protecting marginalized students from harm (via marginal privacy or stronger safeguards) requires, for instance, ensuring such attributes are not easily identifiable or inferable. Methods for designing differently (Toupin, 2024), including co-designing with heterogeneous groups have been developed (Frauenberger et al., 2019), but were not used in the papers reviewed here.

Conclusion and Implications

This study examined the facilitation of AI-enabled learner support in ODDE, applying two complementary conceptual frameworks to structure the analysis. The AI typology (Zawacki-Richter et al., 2019) provided a functional context for classifying applications, focusing on what the tools do rather than their shifting brand names or technical architectures. The ISSPM (Floyd & Casey-Powell, 2004) situated these interventions along the learner's institutional journey, namely, from intake and intervention, through ongoing support, to the measurement of outcomes such as retention and programme evaluation.

In this study, we excluded individual practice-based investigations that indirectly support students by employing AI technologies in classroom contexts to enhance learning experiences. We recognize, however, that a growing body of research highlights how students are independently using AI agents to check their writing, revise language, or even generate texts. The initial fast, albeit not always rigorous, publications (Weidlich et al., 2025), emphasizing how AI could rapidly assist teachers and learners, have been challenged by more critical research over recent decades (e.g., Selwyn et al., 2025; Williamson & Eynon, 2020). Unlike earlier technological waves that typically

required structured, institutional adoption, the enthusiasm surrounding AI today is fueled by its direct availability to individuals outside formal educational systems. This shift has drawn more attention to scholarly discussions beyond the more narrow focus on interactive tutoring systems' direct effect on individual students, and toward the broader debates encompassing issues of post-truth, data colonization, and the ethical as well as societal implications of AI (Bozkurt, 2023; Bozkurt et al., 2024; Williamson et al., 2023).

In this context, it is unsurprising that relatively few publications appeared in the learner support intervention. In this review, we did not see the same rapid response trend - especially to generative AI advancements - in learner support interventions, except for the increase in the chatbot interventions. The studies also tended to adopt a cautious rather than overtly promotional stance toward AI adoption. The overly enthusiastic public-domain discourses surrounding AI were notably absent from this body of research.

Despite their potential, these interventions also raise important questions about data privacy and long-term ethical stewardship. The contractual relationship between learners and institutions is rarely transparent about how support-related data will be used beyond the period of active study. Consent, in practice, often extends over years as historical data fuels predictive models. As Prinsloo et al. (2022) and Taylor et al. (2016) caution, clustering learners by shared attributes—even for seemingly helpful purposes such as social-based course scheduling—can produce aggregate-level risks, reinforcing inequities or making sensitive traits indirectly identifiable. Institutions must therefore treat privacy not as an individual checkbox but as a systemic ethical responsibility.

Equity and inclusiveness also remain underdeveloped. Most interventions target a generalized, homogeneous student body. Specific provisions for marginalized groups, neurodivergent learners, or linguistic minorities are almost absent. Mental well-being and mentoring, peer-support and community building received minimal attention. The Learner Transition phase is untouched by AI-enabled support in the reviewed corpus, and the recent hype around large language models has not yet translated into deployment for learner support. We see this absence, in the short term, as a safeguard against premature or poorly integrated adoption. In the long term, however, it should not result in certain groups being excluded from the development of support interventions.

These gaps matter because the benefits of technological innovation often carry rebound effects: Efficiency can lead to over-automation, personalization can enable over-surveillance, and scaling up can amplify systemic biases. To realize AI's full potential for learner support in ODDE, future work should:

- extend interventions across the full ISSPM spectrum, especially in transition and mentoring stages;
- embed equity and inclusivity as core design principles rather than afterthoughts;
- strengthen consent practices and transparency over long-term data use; and
- develop hybrid models where AI augments, rather than displaces, human support relationships.

This review provides a first structured synthesis of AI-enabled learner support in higher education and identifies key limitations in current research and practice. AI-driven systems predominantly conceptualize learners as homogeneous data points, with limited attention to diversity and inclusion, while institutional applications remain largely confined to prediction-based approaches. These patterns reveal a disconnect between institutional AI implementations and the complex needs of

learners. The findings highlight the need for more learner-centered, inclusive, and ethically grounded approaches, and offer a foundation for future research and practice in this emerging field.

Our findings are consistent with broader critiques in the AI and disability literature. As with other ICT opportunities, AI is frequently presented as a promising means of enhancing accessibility, autonomy, and educational participation for students with disabilities (Fichten, 2000; Fichten et al., 2014; Martiniello et al., 2020). However, evidence suggests that these benefits remain unevenly distributed (Lillywhite & Wolbring, 2020). AI-powered tools can facilitate learning through personalized support, assistive technologies, conversational agents, and adaptive services that respond to individual needs. However, as our synthesis demonstrates as well, AI applications are based on a uniform understanding of students and employ no diverse student experiences. Concurringly with a number of studies (e.g., Trewin et al., 2019), AI applications in education frequently rely on datasets and assumptions derived from normative learner populations about learner behaviour and performance, potentially disadvantaging students whose learning trajectories, interaction patterns, or support requirements fall outside dominant expectations. Consequently, those who stand to benefit most from AI-supported learning may also be the most vulnerable to exclusion, misclassification, or ineffective support. Their call for human-centred and inclusive AI development directly reinforces our conclusion that learner support systems should more actively account for learner diversity.

The broader pattern we identified in our review, whereby learners are often positioned as objects of analysis rather than as participants in the design and evaluation of AI-supported services, concurs with what Lillywhite and Wolbring (2020) highlighted in their study. The disabled people are largely absent from AI-related academic, media, and public discourses as active stakeholders, knowledge producers, and contributors to governance discussions. Instead, they are predominantly portrayed as users or recipients of AI-enabled technologies despite a long-standing co-design call (Fichter et al., 2000; Fichten et al., 2019; Martiniello et al., 2020; Trewin et al., 2019).

Accordingly, the findings of this review have several practical implications for educators and institutional decision-makers. First, institutions should critically reflect on the assumptions embedded in AI-driven learner support systems, particularly the tendency to conceptualize learners as homogeneous groups. Second, AI-based support tools should be complemented by human-centered support structures to ensure that diverse learner needs (especially those of underrepresented or disadvantaged groups) are adequately addressed. Third, transparency in data use and algorithmic decision-making processes is essential to maintain trust and align with ethical standards in learner support. Furthermore, current implementations, which predominantly focus on prediction-based applications such as early warning systems, should be expanded toward more holistic and pedagogically grounded support approaches. Finally, establishing co-design culture with all stakeholders is crucial. Thus, academic and non-academic support points will be able to critically evaluate and effectively integrate AI-supported learner services into their practice.

Across the research reviewed in this paper, AI was summoned for help as a powerful tool to enhance responsiveness, scalability, and personalization in ODDE learner support. However, if these studies do not pay deliberate attention to ethics, equity, and the human dimension, these systems risk reinforcing the very challenges they aim to solve.

Declarations

Conflicts of Interest Statement

The authors declare no conflicts of interest related to the research, authorship, or publication of this article.

References

The articles that are included in the corpus are indicated with an asterisk (*).

Abdul Razak, M.S, Manoj Kumar, M.V, Nirmala, C. R., Naseer, R., Prashanth, B. S., & Sneha, H. R. (2024). Enhancing Student Support with AI: A College Assistance Chatbot Using NLP and ANN. *2nd IEEE International Conference on Networks, Multimedia and Information Technology, NMITCON 2024*, 1–8. https://doi.org/10.1109/NMITCON62075.2024.10699279

*Acang, J. P. A., Nisperos, S. F., Pagtaconan, W. C.R, & Dela Cruz, R. L. (2023). Generating Program Recommendations to Aid in the Decision-Making of Students in Higher Education Using Non-Parametric Supervised Machine Learning. 1–5. https://doi.org/10.1109/HNICEM60674.2023.10589265

*Ajoodha, R, Jadhav, A., & Dukhan, S. (2020). *Forecasting Learner Attrition for Student Success at a South African University. 41st South African Institute of Computer Scientists and Information Technologists, SAICSIT 2020,* 19–28. https://doi.org/10.1145/3410886.3410973

*Alves-Noreña, A. C., Rodríguez-Conde, M. J., Hernández-Ramos, J. P., & Castro-Salgado, J. W. (2025). Technology and Emotions: AI-Driven Software Prototyping for the Analysis of Emotional States and Early Detection of Risky Behaviors in University Students. *Education Sciences*, *15*(3), 1–23. https://doi.org/10.3390/educsci15030350

Arnhold, N. & Bassett, R. M. (2021). Steering tertiary education: Toward resilient systems that deliver for all. Washington DC: World Bank.

Arun, K., Srinagesh, A., & Ganga, P. (2019). A Multi-Model And Ai-Based Collegebot Management System (Aicms) For Professional Engineering Colleges. *International Journal of Innovative Technology and Exploring Engineering*, *8*(9), 2910–2914. https://doi.org/10.35940/ijitee.I8818.078919

Bai, J. Y. H., Zawacki-Richter, O., Bozkurt, A., Lee, K., Fanguy, M., Cefa Sari, B., & Marin, V. I. (2022, September). Automated Essay Scoring (AES) Systems: Opportunities and Challenges for Open and Distance Education. *Tenth Pan-Commonwealth Forum on Open Learning*. Tenth Pan-Commonwealth Forum on Open Learning. https://doi.org/10.56059/pcf10.8339

*Baksh, F., Zorec, M. B., & Kruusamäe, K. (2024). Open-Source Robotic Study Companion with Multimodal Human–Robot Interaction to Improve the Learning Experience of University Students. *Applied Sciences (Switzerland), 14*(13), 1–21. https://doi.org/10.3390/app14135644

*Balderas, A., García-Mena, R. F., Huerta, M., Mora, N., & Dodero, J. M. (2023). Chatbot for communicating with university students in emergency situation. *Heliyon, 9*(9), 1–12. https://doi.org/10.1016/j.heliyon.2023.e19517

*Becerra Á, Mohseni Z, Sanz J, & Cobos R. (2024). *A Generative AI-Based Personalized Guidance Tool for Enhancing the Feedback to MOOC Learners*. 1–8. https://doi.org/10.1109/EDUCON60312.2024.10578809

*Bhave U, Narendra M M, Bhadresh D J, Suhas J A, & Ajit R R. (2024). MindWell Solace: Your Mental Health Companion. 4th IEEE Asian Conference on Innovation in Technology, ASIANCON 2024, pp. 1–4. https://doi.org/10.1109/ASIANCON62057.2024.10837922

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*(1), 4. https://doi.org/10.1186/s41239-023-00436-z

Boyle, F., Kwon, J., Ross, C., & Simpson, O. (2010). Student–student mentoring for retention and engagement in distance education. *Open Learning: The Journal of Open, Distance and e-Learning, 25*(2), 115–130. https://doi.org/10.1080/02680511003787370

Bozkurt, A. (2023). Postdigital Artificial Intelligence. In P. Jandrić (Ed.), *Encyclopedia of Postdigital Science and Education* (pp. 1–5). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35469-4_2-2

Bozkurt, A., & Sharma, R. C. (2020). Emergency remote teaching in a time of global crisis due to CoronaVirus pandemic. *Asian Journal of Distance Education, 15*(1), i-vi. https://doi.org/10.5281/zenodo.3778083

Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., Dron, J., Stracke, C. M., Singh, L., Crompton, H., Koutropoulos, A., Terentev, E., Pazurek, A., Nichols, M., Sidorkin, A. M., Costello, E., Watson, S., Mulligan, D., Honeychurch, S., … Asino, T. I. (2024). The Manifesto for Teaching and Learning in a Time of Generative AI: A Critical Collective Stance to Better Navigate the Future. *Open Praxis, 16*(4), 487–513. https://doi.org/10.55982/openpraxis.16.4.777

Brindley, J. (2014). Learner support in online distance education: Essential and evolving. In O. Zawacki-Richter & A. Terry (Eds.) *Online distance education. Towards a research agenda* (pp. 287–310). AU Press.

Brindley, J. E., & Paul, R. (1996). Lessons from distance education for the university of the future. In *Supporting the learner in open and distance learning* (pp. 43–55). Pitman Publishing.

Brindley, J., C. Walti, and O. Zawacki-Richter (2004). *Learner support in open, distance and online learning environments*. BIS Oldenburg.

Budash, D., & Shaw, M. (2017). Persistence in an Online Master's Degree Program: Perceptions of Students and Faculty. *Online Journal of Distance Learning Administration, 20*(3), 1–19.

Buntins, K., Bedenlier, S., Marín, V., Händel, M., & Bond, M. (2023). Methodische Ansätze zu Evidenzsynthesen in der Bildungstechnologie: Eine tertiäre Übersichtsarbeit. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, 54*, 167–191. https://doi.org/10.21240/mpaed/54/2023.12.20.X

Burke, A. (2019). Student retention models in higher education - a literature review. *College and University, 94*(2), 12–21.

*Cahyana, C, Gozali, A. A., Mandasari, R. I. M, Ramadhan, C. R., & Harnum, K. P. (2024). *Enhancing University Student Services Through the LA-K Chatbot: A Telegram-Based Solution*. 4th International Conference on Electronic and Electrical Engineering and Intelligent System, ICE3IS 2024, pp. 247–252. https://doi.org/10.1109/ICE3IS62977.2024.10775658

Calvert, C. E. (2014). Developing a model and applications for probabilities of student success: A case study of predictive analytics. Open Learning: *The Journal of Open, Distance and e-Learning, 29*(2), 160–173. https://doi.org/10.1080/02680513.2014.931805

*Candrlic, S., Nuzdic, S., & Katic, M. A. (2025). *A Process Model for Academic Student Support Services—A Croatian Case*. 1–8. https://doi.org/10.1109/ACIE64499.2025.00007

Cefa, B. (2025). *Learner support as communication: From service culture to (eco)systems in open, distance, and digital education (ODDE)*. EdArXiv. https://doi.org/10.35542/osf.io/dhwns_v1

Cefa, B., Macgilchrist, F., ElGamal, H., Bai, J. Y. H., Zawacki-Richter, O., & Loglo, F. S. (2025). Responses to the Initial Hype: ChatGPT Supporting Teaching, Learning, and Scholarship*? Open Praxis, 17*(2), 227–250. https://doi.org/10.55982/openpraxis.17.2.872

Cefa, B., & Zawacki-Richter, O. (2025). Mapping the dropout phenomenon in open, distance, and digital education (ODDE): A conceptual analysis. *European Journal of Open, Distance and E-Learning*, *27*(2), 1–15. https://doi.org/10.65043/eurodl.153

*Channarukul, S., Saejiem, N., Bhumichitr, K., Jiamthapthaksin, R., Nicklamai, V., & Terdvikran, K.. (2017). Social-aware automated course planner: An integrated recommender system for university registration system. 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 545–548. https://doi.org/10.1109/ECTICon.2017.8096295

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318. https://doi.org/10.1504/IJTEL.2012.051815

*Chong, S., Lee, Y. H., & Tang, Y. W. (2020). Data Analytics and Visualization to Support the Adult Learner in Higher Education. 4th International Conference on E-Society, E-Education and E-Technology, ICSET 2020, pp. 126–131. https://doi.org/10.1145/3421682.3421698

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Daniel, J., Kanwar, A., & Uvalić-Trumbić, S. (2009). Breaking Higher Education's Iron Triangle: Access, Cost, and Quality. *Change: The Magazine of Higher Learning*, *41*(2), 30–35. https://doi.org/10.3200/CHNG.41.2.30-35

Decuypere, M., & Hartong, S. (2023). Edunudge. *Learning, Media and Technology*, *48*(1), 138-152. https://doi.org/10.1080/17439884.2022.2086261

DiPierro, A. (2025, February 4). *Cal State unveils artificial intelligence tools for students*. EdSource. https://edsource.org/2025/cal-state-unveils-artificial-intelligence-tools-for-students/726205

Dixon-Román, E., Nichols, T. P., & Nyame-Mensah, A. (2020). The racializing forces of/in AI educational technologies. *Learning, Media and Technology*, *45*(3), 236-250. https://doi.org/10.1080/17439884.2020.1667825

Dollinger, M., Cox, S., Eaton, R., Vanderlelie, J., & Ridsdale, S. (2020). Investigating the usage and perceptions of third-party online learning support services for diverse students. *Journal of Interactive Media in Education, 1*, 1–9. https://doi.org/10.5334/jime.555

Duval, E. (2011). Attention please!: Learning analytics for visualization and recommendation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 9–17. https://doi.org/10.1145/2090116.2090118

Fichten, C. S., Asuncion, J. V., Barile, M., Fossey, M., & Simone, C. D. (2000). Access to educational and instructional computer technologies for post-secondary students with disabilities: Lessons from three empirical studies. *Journal of Educational Media, 25*(3), 179–201. https://doi.org/10.1080/1358165000250303

Fichten, C. S., Asuncion, J. V., Barile, M., Ferraro, V., & Wolforth, J. (2019). Accessibility of e-Learning and computer and information technologies for students with visual impairments in postsecondary education. *Journal of Visual Impairment & Blindness*, *103*(9), 543–557. https://doi.org/10.1177/0145482X0910300905

Floyd, D. L., & Casey-Powell, D. (2004). New roles for student support services in distance learning. *New Directions for Community Colleges*, *2004*(128), 55–64. https://doi.org/10.1002/cc.175

*Ford, C., McNally, D., & Ford, K.. (2017). Using design-based research in higher education innovation. *Online Learning Journal, 21*(3), 50–67. https://doi.org/10.24059/olj.v21i3.1232

Frauenberger, C., Spiel, K., Scheepmaker, L., & Posch, I. (2019). Nurturing Constructive Disagreement - Agonistic Design with Neurodiverse Children. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland UK.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning, *TechTrends, 59*(1), 64-71. https://doi.org/10.1007/s11528-014-0822-x

*Gonçalves, G. S., Ribeiro, T. D. S., Teixeira, J. E. V., & Costa, B. K. (2022). The deployment of chatbot to improve customer service in higher education institutions during COVID-19. *International Journal of Innovation, 10*(1), 178–203. https://doi.org/10.5585/iji.v10i1.20652

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews*, (2nd ed.). Los Angeles: SAGE.

*Gray, C. & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, 22–32. https://doi.org/10.1016/j.compedu.2018.12.006

*Hammoudi, H. D, Abdel-Salam, A. S. G, Bensaid, A., Soltani, A., Alsarraj, L., Dalli. R., & Malki, A. (2023). Use of machine learning to assess factors affecting progression, retention, and graduation in first-year health professions students in Qatar: A longitudinal study. *BMC Medical Education, 23*(1), 1–11. https://doi.org/10.1186/s12909-023-04887-w

Han, S., & Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses. *Computers & Education, 179*, 1–22. https://doi.org/10.1016/j.compedu.2021.104395

*Hanshaw, G., Vance, J., & Brewer Craig. (2024). Exploring the Effectiveness of AI Course Assistants on the Student Learning Experience. *Open Praxis*, *16*(4), 627–644. https://doi.org/10.55982/openpraxis.16.4.719

*Hardi, R., Pee, A N.C, Abdullah, M. H. L. B, Pitogo, V. A, Pribadi, A. S., & Rusdi, J. F. (2022). *Academic Smart Chatbot to Support Emerging Artificial Intelligence Conversation*. 194–199. https://doi.org/10.1109/ICSINTESA56431.2022.10041632

*Henríquez, V., Guerra, J., & Scheihing, E. (2024). The impact of an academic counselling learning analytics tool: Evidence from 3 years of use. *British Journal of Educational Technology*, *55*(5), 1884–1899. https://doi.org/10.1111/bjet.13474

*Herodotou, C., Hlosta, M., Boroowa, A., Rienties, B., Zdrahal, Z, & Mangafa, C. (2019). Empowering online teachers through predictive learning analytics. *British Journal of Educational Technology*, *50*(6), 3064–3079. https://doi.org/10.1111/bjet.12853

*Herodotou, C., Naydenova, G., Boroowa, A., Gilmour, A., & Rienties, B. (2020). How can predictive learning analytics and motivational interventions increase student retention and enhance administrative support in distance education? *Journal of Learning Analytics*, *7*(2), 72–83. https://doi.org/10.18608/JLA.2020.72.4

*Hill, A., Goo, K, & Agarwal, P. (2025). Recommending the right academic programs: An interest mining approach using BERTopic. *Data Mining and Knowledge Discovery, 39*(3), 1–40. https://doi.org/10.1007/s10618-024-01087-y

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emer-gency remote teaching and online learning. *EDUCAUSE Review*. https://er.edu-cause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

*Honson, V., Vu, T., Tran, T. P., & Tejada Estay, W. (2024). Using learning analytics to alleviate course and student support administrative load for large classes: A case study. *Journal of Work-Applied Management*, *16*(2), 303–315. https://doi.org/10.1108/JWAM-11-2023-0121

Huang, S.-P. (2018). Effects of Using Artificial Intelligence Teaching System for Environmental Education on Environmental Knowledge and Attitude. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(7). https://doi.org/10.29333/ejmste/91248

Huang, J., & Chen, Z. (2016). The research and design of web-based intelligent tutoring system. *International Journal of Multimedia and Ubiquitous Engineering, 11*(6), 337–348. https://doi.org/10.14257/ijmue.2016.11.6.30.

Jantti, M. & Heath, J. (2016). What role for libraries in learning analytics? Performance Measurement and Metrics, 17*(2), 203–210. https://doi.org/10.1108/PMM-04-2016-0020

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics, 1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Jensen, L. X., Buhl, A., Sharma, A., & Bearman, M. (2025). Generative AI and higher education: A review of claims from the first months of ChatGPT. *Higher Education*, *89*(4), 1145–1161. https://doi.org/10.1007/s10734-024-01265-3

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education*, *87*, 204–217. https://doi.org/10.1016/j.compedu.2015.07.002

Joksimović, S., Kovanović, V., & Dawson, S. (2019). The Journey of Learning Analytics. *HERDSA Review of Higher Education*, *6*, 37–63.

*Karakostas, & Demetriadis, S. (2014). Adaptive vs. Fixed Domain Support in the Context of Scripted Collaborative Learning. *Educational Technology & Society*, *17*(1), 206–217.

*Khan, M. A., Rehman, A., Shah, A. A., Abbas, S., Alharbi, M., Ahmad, M., & Ghazal, T. M. (2025). Navigating the future of higher education in Saudi Arabia: Implementing AI, machine learning, and big data for sustainable university development. *Discover Sustainability*, *6*(1). https://doi.org/10.1007/s43621-025-01388-2

*Khumalo, S., Rawatlal, R., Nnadozie, V., Mahadew, A., Mpungose, C. B., & Mazibuko, P. (2023). Technology-mediated advising for student success: Exploring selfmediated academic support for undergraduate students using AutoScholar Advisor System. *Perspectives in Education, 41*(2), 211–232. https://doi.org/10.38140/pie.v41i2.7088

Labadze, L., Grigolia, M. & Machaidze, L. Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technologies and Higher Education, 20*, 56 (2023). https://doi.org/10.1186/s41239-023-00426-1

Larrabee Sønderlund, A., Hughes, E. and Smith, J. (2019), The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technologies, 50*, 2594-2618. https://doi.org/10.1111/bjet.12720

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review (Online)*. https://www.proquest.com/scholarly-journals/penetrating-fog-analytics-learning-education/docview/3224594060/se-2?accountid=12968

*Li J, Chu C, & Whittaker T. (2024, October). *Identifying Interpretable Features Impacting Nontraditional Undergraduate Computer Science Student Retention*. 2024 IEEE Frontiers in Education Conference (FIE), 1–9. https://doi.org/10.1109/FIE61694.2024.10892997

Lillywhite, A., & Wolbring, G. (2020). Coverage of Artificial Intelligence and Machine Learning within Academic Literature, Canadian Newspapers, and Twitter Tweets: The Case of Disabled People. *Societies, 10*(1), 23. https://doi.org/10.3390/soc10010023

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

*Lu, X. (2019). An Empirical Study on the Artificial Intelligence Writing Evaluation System in China CET. *Big Data*, *7*(2), 121–129. https://doi.org/10.1089/big.2018.0151

*Lucien, R. & Park, S. (2024). Design and Development of an Advising Chatbot as a Student Support Intervention in a University System. *TechTrends: Linking Research & Practice to Improve Learning*, *68*(1), 79–90. https://doi.org/10.1007/s11528-023-00898-y

*Lytridis C & Tsinakos A. (2025). AIMT Agent: An Artificial Intelligence-Based Academic Support System. *Information*, *16*(4), 1–15. https://doi.org/10.3390/info16040275

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Martiniello, N., Asuncion, J., Fichten, C., Jorgensen, M., Havel, A., Harvison, M., Legault, A., Lussier, A., & Vo, C. (2020). Artificial intelligence for students in postsecondary education: A world of opportunity. *AI* Matters, 6(3), 17–29. https://doi.org/10.1145/3446243.3446250

Moore, M. G., & Kearsley, G. (2012). *Distance education: A systematic view of online learning* (3rd.). Wadsworth Cengage Learning.

*Ndou, N., Ajoodha, R., & Jadhav, A. (2021). A case study to enhance student support initiatives through forecasting student success in higher-education. *Advances in Science, Technology and Engineering Systems*, *6*(1), 230–241. https://doi.org/10.25046/aj060126

*Ndunagu, J. N., Ezeanya, C. U., Onuorah B O, Onyeakazi J C, & Ukwandu E. (2025). A Chatbot Student Support System in Open and Distance Learning Institutions. *Computers*, *14*(3), 1–17. https://doi.org/10.3390/computers14030096

*Nurshatayeva, A., Page, L. C., White, C. C., & Gehlbach, H. (2021). Are Artificially Intelligent Conversational Chatbots Uniformly Effective in Reducing Summer Melt? Evidence from a Randomized Controlled Trial. *Research in Higher Education*, *62*(3), 392–402. https://doi.org/10.1007/s11162-021-09633-z

*Öncü, S. E., Gevher Merve, & Erdoğdu Erdem. (2025). Exploring the potential use of generative AI for learner support in ODL at scale. *Journal of Educational Technology & Online Learning* (JETOL), *8*(1), 80–99. https://doi.org/10.31681/jetol.1559442

*Öncü Sefa Emre & Süral İrfan. (2024). Leveraging AI for Enhanced Support: Satisfaction Levels of Users Utilizing Virtual Assistant in Open Education. *Asian Journal of Distance Education*, *19*(1), 200–220.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinnrdd, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *71*. https://doi.org/10.1136/bmj.n71

*Pansombut. S. & Kirimasthong, K. (2025). *Customer Service Transformation for Admission in Mae Fah Luang University Using Chatbots Technology*. 378–383. https://doi.org/10.1109/ECTIDAMTNCON64748.2025.10962037

*Prasertsri, S., Yotchimplee, C., Dangdee, P., Khadsee, J., Boonserchchoo, P., Juengjarone, P., & Patmanee, J. (2024). *Automated Academic Service Questions Answering System with Chatbot via LINE OA*. 9th International STEM Education Conference, iSTEM-Ed 2024, Cha-am, Hua Hin. https://doi.org/10.1109/iSTEM-Ed62750.2024.10663129

*Priccilia, S. & Girsang, A. S. (2024). Indonesian generative chatbot model for student services using GPT. *International Journal of Informatics and Communication Technology*, *13*(1), 50–56. https://doi.org/10.11591/ijict.v13i1.pp50-56

Prinsloo, P., Slade, S., & Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology*, *53*(4), 876–893. https://doi.org/10.1111/bjet.13216

Rodríguez, M. E., Guerrero-Roldán, A. E., Baneres, D., & Karadeniz, A.. (2022). An Intelligent Nudging System to Guide Online Learners. *International Review of Research in Open & Distributed Learning*, *23*(1), 41–62.

*Roestenburg, J., Kruger, C. J., Nel, M., & van Rensburg, Z. J. (2024). Exploring factors influencing academic literacy—A data-driven perspective. *South African Journal Of Information Management, 26*(1), 1–10. https://doi.org/10.4102/sajim.v26i1.1729

Rotar, O. (2022a). A missing theoretical element of online higher education student attrition, retention, and progress: A systematic literature review. *SN Social Sciences*, *2*(12), 278. https://doi.org/10.1007/s43545-022-00550-1

Rotar, O. (2022b). Online student support: A framework for embedding support interventions into the online learning cycle. *Research and Practice in Technology Enhanced Learning*, *17*(1), 2. https://doi.org/10.1186/s41039-021-00178-4

Rumble, G. (2000). Student support in distance education in the 21st century: Learning from service management. *Distance Education, 21*(2), 216–235. https://doi.org/10.1080/0158791000210202

*Schatten, M., Euric, B. O., & Peharda T. (17-20 2021). *A cognitive agent for university student support*. 1–6. https://doi.org/10.1109/TEMSCON-EUR52034.2021.9488627

Selwyn, N. (2024). On the Limits of Artificial Intelligence (AI) in Education. *Nordisk Tidsskrift for Pedagogikk Og Kritikk*, *10*(1), 3–14.

Selwyn, N., Ljungqvist, M., & Sonesson, A. (2025). When the prompting stops: Exploring teachers' work around the educational frailties of generative AI tools. *Learning, Media and Technology*, 1–14. https://doi.org/10.1080/17439884.2025.2537959

*Seo, E. Y., Yang, J., Lee, J. E., & So, G. (2024). Predictive modelling of student dropout risk: Practical insights from a South Korean distance university. *Heliyon*, *10*(11), 1–17. https://doi.org/10.1016/j.heliyon.2024.e30960

*Sereni, A. (2024). The Crucial Role of International Student Support Experts: Creative Initiatives Empowering International Students at the University of Adelaide. *Journal of the Australian and New Zealand Student Services Association*, *32*(2), 183–196. https://doi.org/10.30688/janzssa.2024-2-09

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Simpson, O. (2002). *Supporting students in online, open and distance learning* (2nd ed.). Routledge. https://doi.org/10.4324/9780203417003

Simpson, O. (2004). The impact on retention of interventions to support distance learning students. *Open Learning: The Journal of Open, Distance and e-Learning*, *19*(1), 79–95. https://doi.org/10.1080/0268051042000177863

Simpson, O. (2013). Student retention in distance education: Are we failing our students? *Open Learning*, *28*(2), 105–119.

Simpson, O. (2022). *Supporting Students in Online, Open and Distance Learning* (2nd ed.). Routledge. https://doi.org/10.4324/9780203417003

Simpson, O. (2023). Student Support in Online Learning—We Need to Talk About Money. *The International Review of Research in Open and Distributed Learning*, *24*(4), 102–118. https://doi.org/10.19173/irrodl.v24i4.7241

Skinner-Thompson, S. (2020). *Privacy at the Margins* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316850350

Suta, P., Lan, X., Wu, B., Mongkolnam, P., & Chan, J. H. (2020). An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research, 9*(4), 502-510. https://doi.org/10.18178/ijmerr.9.4.502-510

*Sweidan, S. Z., Abu Laban, S. S., Njood. A., & Darabkh, K. A. (2021). SIAAA-C: A student interactive assistant android application with chatbot during COVID-19 pandemic. *Computer Applications in Engineering Education*, *29*(6), 1718–1742. https://doi.org/10.1002/cae.22419

Tait, A. (2000). Planning Student Support for Open and Distance Learning. *Open Learning: The Journal of Open, Distance and e-Learning*, *15*(3), 287–299. https://doi.org/10.1080/713688410

Taylor, L., Floridi, L., van der Sloot, B. (Eds.). (2016). *Group Privacy: New challenges of data technologies* (Vol. *126*). Springer.

Tempelaar, D., Rienties, B., & Nguyen, Q. (2021). The Contribution of Dispositional Learning Analytics to Precision Education. *Educational Technology & Society, 24*(1), 109-122.

*Thesen, T., Marrero, W. J., Konopasky, A. J., Duncan, M. S., & Blackmon, K. E. (2025). Towards precision well-being in medical education. *Medical Teacher, 47*(4), 630–634. https://doi.org/10.1080/0142159X.2024.2357279

*Thongda, N. & Songpan, W. (2023). Enhancing Education Process Through Intelligent Association Mining (IAM) System Conversations: Chatbot Services in Educational Settings. WSSE '23: *Proceedings of the 2023 5th World Symposium on Software Engineering,* 261–266. https://doi.org/10.1145/3631991.3632034

Thorpe, M. (2003). Collaborative online learning: Transforming learner support and course design. In *Rethinking learner support in distance education: Change and continuity in an international context (Eds. A. Tait, & R. Mills)* (1st ed., pp. 198–211). RoutledgeFalmer.

Toupin, S. (2024). Shaping feminist artificial intelligence. *New Media & Society*, *26*(1), 580-595. https://doi.org/10.1177/14614448221150776

Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., Hebert, D., Lyckowski, N., & Manser, E. (2019). Considerations for AI fairness for people with disabilities. *AI Matters, 5*(3), 40–63. https://doi.org/10.1145/3362077.3362086

*Ula, M., Hardi, R., & Hipiny, I. (2023). An Improved Structure for Academic Information Services through AI Chatbots. *Journal of Engineering Science and Technology Review, 16*(5), 164–173. https://doi.org/10.25103/jestr.165.20

*Villano, R., Harrison, S., Lynch, G., & Chen, G. (2018). Linking early alert systems and student retention: A survival analysis approach. *Higher Education*, *76*(5), 903–920. https://doi.org/10.1007/s10734-018-0249-y

Walsh, C., Mital, A., Ratcliff, M., Yap, A., & Jamaleddine, Z. (2020). A public-private partnership to transform online education through high levels of academic student support. *Australasian Journal of Educational Technology*, *36*(5), 30–45. https://doi.org/10.14742/ajet.6107

*Wangwiwattana, C. & Jantarick, W. (2024). *Building Intelligent Academic Service Agents: FAQ Extraction and Chatbots With Large Language Models*. 370–375. https://doi.org/10.1109/InCIT63192.2024.10810553

Weidlich, J., Gasevic, D., Drachsler, H., & Kirschner, P. (2025). ChatGPT in education: An effect in search of a cause. PsyArXiv. https://doi.org/10.31234/osf.io/t6uzy_v2

Welham, D. (2008). AI in training (1980-2000): Foundation for the future or misplaced optimism? *British Journal of Educational Technology*, *39*(2), 287–303.

Whitman, M. (2020). “We called that a behavior”: The making of institutional data. *Big Data & Society*, *7*(1), 1-13. https://doi.org/10.1177/2053951720932200

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, *45*(3), 223-235. https://doi.org/10.1080/17439884.2020.1798995

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., & Anderson, T. (2014). *Online Distance Education: Towards a Research Agenda*. Athabasca University Press. https://doi.org/10.15215/aupress/9781927356623.01

Zawacki-Richter, O., & Bozkurt, A. (2022). Research Trends in Open, Distance, and Digital Education. In O. Zawacki-Richter & I. Jung (Eds.) *Handbook of Open, Distance and Digital Education* (pp. 1–23). Springer Nature. https://doi.org/10.1007/978-981-19-0351-9_12-1

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zawacki-Richter, O., & Naidu, S. (2016). Mapping research trends from 35 years of publications in *Distance Education*. *Distance Education*, *37*(3), 245–269. https://doi.org/10.1080/01587919.2016.1185079

Zawacki-Richter, O., & Jung, I. (2023). Shaping the Field of Open, Distance, and Digital Education. In O. Zawacki-Richter & I. Jung (Eds.) *Handbook of Open, Distance and Digital Education* (pp. 1–10). Springer Nature. https://doi.org/10.1007/978-981-19-0351-9_94-2

Appendices and Supplemental Materials

For the following list of appendices, please see: https://osf.io/gyc5z

Appendix A: Country of the first author
Appendix B: Subcategories of department affiliations
Appendix C: Research design coding overview
Appendix D: Education institution types
Appendix E: Coding list

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5691
https://doi.org/10.24059/olj.v30iSI2.5691

A Meta-Analysis of AI-Assisted Self-Regulated Learning

Zhuo Zhang
Towson University, USA
Belle Li
Purdue University, USA
Qian Xu
Iowa State University, USA
Zui Cheng
Shenzhen University, China
Richard E. West
Brigham Young University, USA

Abstract

This meta-analysis examines the effects of artificial intelligence (AI)-based pedagogical interventions on learners' self-regulated learning and its subdomains, including cognitive, metacognitive, and resource management strategies. In addition, the study investigates whether these effects vary across four groups of moderators concerning research context, intervention design, AI, and self-regulated learning. A systematic search of databases identified an initial pool of 5,741 potentially relevant studies, of which 34 met all inclusion criteria. These studies yielded a total of 91 effect sizes. Results showed a positive overall effect of AI-based pedagogical interventions on self-regulated learning (Hedges' g = 0.76). Positive effects were also observed across SRL subdomains (g = 0.48 for cognitive strategies, 0.73 for metacognitive strategies, and 0.93 for resource management strategies). Moderator analyses showed that the effects of AI-based interventions on self-regulated learning varied significantly across educational levels, AI pedagogical roles, and types of AI applications, whereas other factors such as discipline, intervention length, type of control group, sample size, SRL measure did not show significant differences. By synthesizing existing empirical evidence, this meta-analysis provides a quantitative foundation for understanding how and under what conditions AI can be effectively integrated into education to support learners' self-regulated learning.

Keywords: self-regulated learning, artificial intelligence, meta-analysis

Introduction

Self-regulated learners take an active role in planning, monitoring, managing, and evaluating their learning processes and outcomes to attain predetermined goals (Pintrich, 2000, 2004). Self-regulated learning (SRL) is therefore widely recognized as a key competency associated with academic achievement, motivation, and satisfaction across various educational contexts, including K-12 education (Dent & Koenka, 2016), higher education (Broadbent & Poon, 2015; Cheng et al., 2025), and continuing education (Hemmler & Ifenthaler, 2024).

Given its importance across multiple dimensions of learning outcomes, a substantial body of research has examined interventions designed to strengthen learners' SRL capacity, particularly through the integration of educational technology. Among these technologies, artificial intelligence (AI) has emerged as an increasingly prominent approach. In recent years, researchers have shown growing interest in incorporating AI technologies, such as large language models (LLMs), chatbots, intelligent tutoring systems, and learning analytics dashboards, into pedagogical interventions aimed at supporting learners' SRL (e.g., Fleur et al., 2023; Liu et al., 2024; McCarthy et al., 2018; Ortega-Ochoa et al., 2024).

Corresponding author: Zhuo Zhang, zhuozhang@towson.edu
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

The growing body of empirical evidence on AI-based pedagogical interventions (AI-PI) and SRL necessitates a rigorous quantitative research synthesis to address two key questions. First, do AI-PI enhance learners' SRL, and if so, to what extent? Second, what factors, such as AI's pedagogical role, AI tool type, learners' educational level, and disciplinary context, may influence the effects of AI-PI on SRL? Existing quantitative research syntheses on AI integration in education have largely focused on learning achievement or outcomes (e.g., Dai et al., 2024; Deng et al., 2025; Dong et al., 2025; Zheng et al., 2023). In contrast, the ways in which AI integration shapes how learners learn (their learning strategies and behaviors) have received far less systematic attention.

To address this gap, the present research synthesis integrates existing quantitative evidence on AI-PI and SRL. By adopting meta-analytic procedures, this study aims to provide a comprehensive assessment of the overall effects of AI-PI on SRL and to examine the conditions under which these effects vary.

Literature Review

Self-Regulated Learning (SRL) and SRL Strategies

Self-regulated learning (SRL) refers to a purposeful process in which learners set goals and actively monitor, adjust, and manage their cognition, motivation, and behaviors (Pintrich, 2000). Drawing on Weinstein et al.'s (2000) definition of learning strategies as "any thoughts, behaviors, beliefs, or emotions that facilitate the acquisition, understanding, or later transfer of new knowledge and skills" (p. 727), this study defines SRL strategies as the intentional thoughts, behaviors, beliefs, and emotional actions learners use to set goals and to regulate their cognition, motivation, and behaviors in support of learning and knowledge transfer.

SRL strategies can be further categorized by learning domain and phase. A common domain-based distinction includes metacognitive, cognitive, and resource management learning strategies (Hemmler & Ifenthaler, 2024; Pintrich et al., 1999; Theobald, 2021). According to Pintrich et al. (1999), cognitive learning strategies are strategies that help learners both remember information and comprehend more complex content, such as rehearsal, elaboration, and organization. Metacognitive strategies are strategies that support planning, monitoring, and reflecting activities, such as goal setting, task analysis, and self-evaluation. Resource management strategies are strategies that help manage and control internal and external resources such as effort, time, learning environment, and interpersonal support. Examples include effort regulation, time management, environmental structuring, and help-seeking. A widely adopted phase-based distinction is Zimmerman's (2000) cyclical three-phase model, which includes the forethought, performance, and self-reflection phases. In forethought phase, sample learning strategies include goal setting, planning, and activation of prior knowledge; in performance phase, sample learning strategies include help-seeking, time management, environmental structuring, rehearsal, and elaboration; and in the reflection phase, sample learning strategies include self-reflection and self-assessment.

Artificial Intelligence in Education

Education, like nearly all disciplines, has been affected by the emergence of generative artificial intelligence tools. While artificial intelligence tools have been studied since the 1950s with early focus on natural language processing and machine translation to generate simple sentences and paragraphs (Yu & Guo, 2023), more recent generative versions of these technologies produce output that is more interactive, more helpful, and more accurate (while deep understanding of concepts as yet is still elusive). These tools can impact education including supporting teachers in creating lesson plans (Gunbater et al., 2025), summarizing discussion comments to facilitate focused interactions (Bauer et

al., 2023), and providing more detailed feedback to students (West & Wiley, 2024). In addition, AI tools provide many opportunities to students, including support in understanding concepts through chatbot interactions (Thottoli et al., 2024), scaffolding in writing (Kim & Kim, 2022), and even making educational materials more accessible and personalized (Wiley, 2025). In addition, AI tools have been discussed as problematic for education in that it may facilitate cheating (Elom et al., 2025), limit human interactions (Alwaqdani, 2025), or impact students negatively due to bias (Warr et al., 2024).

Perhaps as would be expected, much of the discussion around the use of AI in education has lacked nuance, either espousing its great benefits or significant harms. For example, Nedungadi et al. (2024) argued that Generative AI provided "transformative power" for achieving quality education that is equitable and lifelong. Konstantinova et al. (2023) similarly argued that Generative AI has "all the potential" for solving educational challenges. In addition, much of the early research on AI in education has focused on teacher readiness or acceptance (Alwakdani, 2025; Yue et al., 2024) and AI literacy (Su et al., 2023). However, some scholars are calling for AI scholarship that delves deeper into how AI can support various kinds of learning. For example, Ouyang and Jiao (2021) argued that AI research could be organized into three paradigms: AI-directed, AI-supported, and AI-empowered learning. This final paradigm represents the "learner as leader" where learners work with AI to take agency over their learning in powerful ways. Part of this empowerment could be using AI to improve self-regulated learning, or the empowerment of the individual to best control their own learning processes.

AI-Based Pedagogical Interventions (AI-PI) and Self-Regulated Learning

In this study, AI-PI are defined as intentional implementations of AI-powered tools and systems designed to support, scaffold, or enhance students' learning processes and outcomes. Recent systematic reviews have documented the growing body of research on how various AI technologies can be integrated into educational practice to support learning (Banihashem et al., 2025; Lan & Zhou, 2025), including scaffolding students' SRL processes. AI-driven educational tools, such as intelligent tutoring systems, adaptive learning platforms, and conversational agents, are increasingly used to provide personalized guidance and feedback that support key SRL processes, including goal setting, strategic planning, monitoring, and reflection (Guan et al., 2025; Lan & Zhou, 2025). For example, chatbot tutors can prompt learners to identify resources, apply appropriate learning strategies, and monitor their understanding in real time.

Empirical studies across different educational contexts have reported improvements in students' SRL when AI tools are integrated into learning activities. In the domain of language learning, for instance, Wei (2023) showed that college-level English learners who engaged with an AI-powered learning platform exhibited significantly greater gains in SRL strategies compared to a control group receiving traditional instruction. The AI-mediated instruction in Wei's study provided interactive exercises, instant feedback, and adaptive hints, which helped students plan their study, monitor progress, and adjust strategies more effectively than their peers in the non-AI condition (Wei, 2023). Similarly, in science education, Ng et al. (2024) conducted a pilot experiment comparing a generative AI-based chatbot to a rule-based chatbot for supporting high school students' SRL during a three-week online course. Their findings indicated that the AI-enhanced chatbot ("SRLbot," powered by ChatGPT) led to higher levels of student engagement, more frequent help-seeking and self-monitoring behaviors, and improved science learning outcomes, relative to the traditional rule-based tutor (Ng et al., 2024). Students using the generative AI tutor reported that its personalized recommendations and adaptive responses helped them maintain regular study habits and reduce learning anxiety, thereby boosting motivation and proactive SRL behaviors (Ng et al., 2024). These examples align with a broader trend in the literature: well-designed AI interventions (e.g. intelligent tutors, recommender systems, and adaptive feedback agents) can effectively mimic the one-on-one support of a human coach, prompting learners to utilize SRL strategies more consistently and confidently (Chang & Sun, 2024).

Traditional and AI-Based Pedagogical Interventions for Self-Regulated Learning

Before AI was used to support pedagogical interventions, SRL was commonly developed through teacher-led strategy training, reflective journals, peer support, and metacognitive prompts. Meta-analyses show consistent benefits: university-level SRL programs improve academic performance and metacognitive strategies (Theobald, 2021), while K-12 interventions yield even larger effects on achievement (Elhusseini et al., 2022). Gains are strongest when instruction emphasizes metacognition and includes feedback from teachers or peers (Jansen et al., 2019).

Both traditional and AI-PI aim to enhance students' self-regulated learning, but they differ in how this support is delivered. Traditional interventions rely on human interaction and the expertise of teachers or mentors to guide learners. This human involvement allows for rich, nuanced feedback and emotional support, such as encouragement and empathy, that technology alone may not fully replicate. In contrast, AI-based systems are well suited for scalability and real-time adaptation. For example, intelligent tutoring systems can continuously track learner performance and immediately tailor practice questions or hints, whereas human instructors face practical limits in providing instant, individualized feedback in classroom settings (Lan & Zhou, 2025).

However, there are also cautions to consider. Some scholars argue that an over-reliance on AI guidance could risk diminishing students' own agency if not carefully managed, for instance, if learners simply follow algorithmic prompts without internalizing the self-regulatory reasoning behind them (Molenaar, 2022). The ideal path forward, as suggested by Molenaar (2022), is to combine the strengths of both approaches. That is "hybrid human-AI regulation" where control over the learning process dynamically shifts between the learner, the human teacher, and the AI as needed, with the aim of transferring responsibility to the learner.

Existing Reviews and Research Questions

Although the use of AI to support SRL is still an emerging area, a substantial body of empirical research has already accumulated. As a result, numerous recent systematic reviews have examined AI-supported SRL from different angles, including reviews with a broad scope (Banihashem et al., 2025) and those focused on specific contexts such as higher education (Lan & Zhou, 2025), language learning (e.g., Chang & Sun, 2024; Zhang et al., 2025), generative AI (Sardi et al., 2025; Qi et al., 2025), AI chatbots (Wu et al., 2025), AI-generated scaffolds (Qian et al., 2025), and data streams and learning analytics (de Mooij et al., 2025). Despite this growing literature, research synthesis that quantifies the effect of AI integration on SRL remains limited. Existing meta-analyses have been predominantly used to evaluate AI tools or interventions on learning outcomes instead of learning strategies (e.g., Dai et al., 2024; Deng et al., 2025; Zheng et al., 2023). To our knowledge, the only meta-analysis directly related to this topic is the one by Yin et al. (2025), which examined the effects of pedagogical agents (a specific type of AI-enhanced learning assistant) on overall SRL. Beyond the work of Yin et al. (2025), our meta-analysis provides a deeper and more comprehensive examination by investigating how different forms of AI tools, such as intelligent tutoring systems, conversational agents, adaptive learning platforms, and learning analytics dashboards, influence the effects of interventions on overall SRL as well as specific SRL subdomains.

The following three research questions guided our investigation.

1. What is the overall effect of AI-based pedagogical interventions (AI-PI) on self-regulated learning (SRL) strategies and how does the effect vary across studies?
2. What are the effects of AI-PI on subdomains (e.g., metacognitive, cognitive, and resource management strategies) of SRL? How do the effects vary across studies?

3. What are the conditions (e.g., educational level, discipline, intervention length, type of AI applications, AI's pedagogical role, SRL measure) that moderate the effect of AI-based interventions on SRL strategies?

Method

Study Identification

Potentially relevant studies were identified through searches in ERIC, Education Source, PsycINFO, Web of Science, Scopus, and ProQuest Dissertations and Theses Global. As documented in Appendix B, our search string included three categories: SRL, AI, and research design. The Boolean operator "and" connected the three categories, and "or" was used within each category. Search terms were developed by reviewing and adapting published search terms for SRL (Broadbent & Poon, 2015; Cheng et al., 2025), AI (Ouyang et al., 2022; Zawacki-Richter et al., 2019), and research design (Deng et al., 2025).

Within the SRL category, we included both general terms for SRL and terms for specific strategies, such as time management and help-seeking, to capture studies that examined individual strategies without explicitly referring to the broader SRL construct. Within the AI category, we expanded the terms to encompass specific models, tools, and applications, accommodating variations in how researchers described AI-based interventions. The research-design block focused retrieval on experimental and quasi-experimental studies, consistent with our aim of estimating intervention effects on SRL. Its inclusion was informed by previous meta-analyses of technology-based interventions (e.g., Deng et al., 2025) that used design terms in their search strategies, as well as by the consideration that experimental designs provide a relatively strong basis for causal inference (e.g., Shadish et al., 2002). To reduce reliance on design terminology appearing in titles or abstracts, this block was searched across broader searchable fields where supported. SRL and AI terms were searched in titles, abstracts, and keywords (where supported by the database) because these central concepts were expected to be reported in those fields.

The search was conducted on December 29, 2024. No initial date limit was applied to ensure a comprehensive inclusion of relevant studies. The complete database-specific search strategies are provided in Appendix B, including field codes, Boolean operators, quotation marks, and truncation symbols. ERIC, Education Source, and PsycINFO were searched simultaneously through the EBSCOhost interface using a common search string. The search terms and Boolean structure were consistent across platforms, while field restrictions were adapted to each platform's search interface. Definitions of the field codes are provided in the appendix notes.

Inclusion and Exclusion Criteria

Inclusion Criteria

(1) Were reported in English
(2) Were reported in journal articles, conference papers, or dissertations.
(3) Employed experimental or quasi-experimental designs and at least one experimental group employed AI-based intervention (e.g., AI-driven feedback, interaction with a conversational agent, adaptive prompts and recommendations)
(4) Studies included a control group that either received no intervention (passive control) or used an alternative traditional pedagogical approach that relied on a teacher, peers, a non-intelligent learning tool or system, textbooks, or checklists (active control)

(5) Included SRL, a SRL subdomain (metacognitive, cognitive, or resource management strategies), or at least one SRL strategy (e.g., help-seeking, self-evaluation) as an outcome variable

(6) Reported sufficient quantitative data (e.g., means and standard deviations, t-statistic, F-statistic) for computing effect size.

Exclusion Criteria

(1) Employed a one-group pre-posttest design, which poses a potential threat to internal validity and limits causal inference

(2) Examined the effect of an AI-PI on multiple SRL strategies but reported quantitative results only for strategies that reached statistical significance. This selective reporting could introduce bias into the evidence base.

Screening

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021) for screening identified reports. The PRISMA flow diagram was used to illustrate the screening process, and Rayyan (Ouzzani et al., 2016), a digital systematic review platform, was used to manage and facilitate screening.

As shown in Figure 1, our search identified 5,741 records, which included 1,323 duplicates. After removing duplicates, 4,418 records remained for the initial round of screening. Given the large number of records, the first and second authors jointly screened an initial set of 50 records to pilot the process and to ensure consistent application of the inclusion and exclusion criteria. After piloting, the two authors divided the remaining records into two sets and independently screened titles, keywords, and abstracts based on the established criteria. Both reviewers adopted a conservative decision rule: records were retained for full-text assessment whenever the available title, keywords, and abstract did not provide sufficient information to determine eligibility.

A total of 216 reports were retained for full-text review. These reports were divided into two groups. The first and fourth authors reviewed reports published in 2024 and 2025, which resulted in 130 records. The second and third authors reviewed reports published in or before 2023, which resulted in 86 records. Each report was reviewed by two authors to ensure reliability. Within each group, the corresponding two authors independently reviewed all full text reports, and any disagreements were resolved through discussion.

After full-text screening, 34 reports met all eligibility criteria and were included for coding and analysis. These reports consisted of 30 journal articles and four conference papers.

Citation Searching

Backward citation searching was conducted using the reference lists of the 34 reports included in the meta-analysis, yielding 2,176 records. Before screening, 251 duplicates or already included records were removed; this count included five reports already included and coded in the meta-analysis. The remaining records were compared with the 4,418 records screened during the initial search, identifying 13 additional previously screened records. In total, 264 records were removed, leaving 1,912 new records for title and abstract screening.

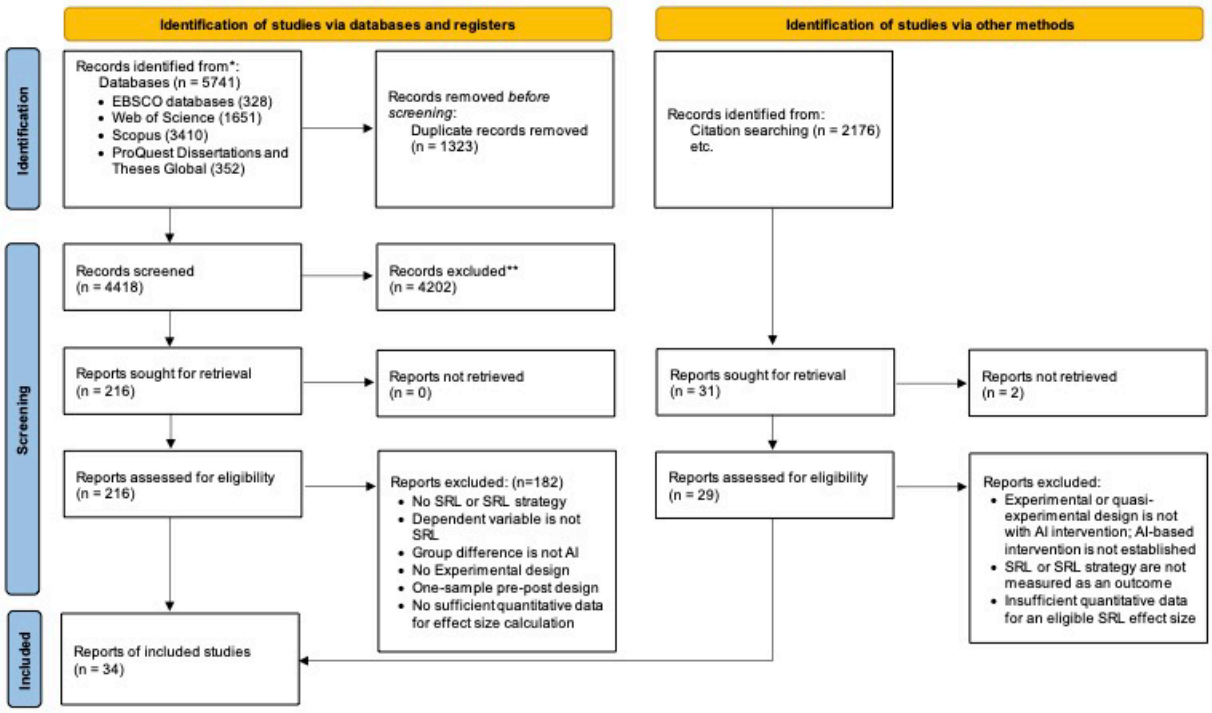
One researcher served as the primary reviewer for title and abstract screening. This researcher had completed screening training and participated in the initial screening round; therefore, no additional pilot screening exercise was conducted. Records were screened against the established inclusion and

exclusion criteria. Of the 1,912 records screened, 1,881 were excluded, and 31 reports were sought for full-text review.

Two dissertations could not be retrieved, leaving 29 reports for full-text assessment by two researchers. The primary reviewer assessed all available reports and provisionally identified two for inclusion. The second researcher served as a co-reviewer, after which the researchers discussed their assessments and reached consensus. Ultimately, all 29 retrieved reports were excluded: 20 did not evaluate an established AI-based intervention using an experimental or quasi-experimental design (e.g., Arnold & Pistilli, 2012); seven did not measure SRL or an SRL strategy as an outcome (e.g., Arguedas et al., 2018); and two provided insufficient quantitative data to calculate an eligible SRL effect size (e.g., Zheng et al., 2019). Thus, backward citation searching yielded no additional reports for inclusion in the meta-analysis.

Figure 1

PRISMA Flow Diagram



Quality Appraisal

The methodological quality of the 34 included articles was appraised using six criteria adapted from Cheng et al. (2025): baseline control, external validity, internal validity, examination of statistical assumptions, construct validity, and power or sample adequacy. The appraisal built on existing study-characteristic codes for pretest administration (pretest versus no pretest) and research design (experimental versus quasi-experimental). Two researchers independently reviewed these codes alongside the full texts and coded the additional information required for the six appraisal criteria. Each criterion was scored dichotomously (1 = criterion met; 0 = criterion not met or not reported), and the ratings were summed to produce a descriptive quality score ranging from 0 to 6.

Before reconciliation, the researchers agreed on 182 of 204 ratings (89.2%). Twenty-two discrepancies occurred across 17 articles, primarily involving baseline control and interpretation of the appraisal criteria. The researchers revisited the relevant full texts, discussed discrepancies using consistent decision rules, and reached consensus on all final ratings. Agreement statistics were calculated from the independent ratings before reconciliation.

The consensus quality scores ranged from 1 to 4 (M = 2.76, SD = 0.85). Three articles scored 1, eight scored 2, seventeen scored 3, and six scored 4.

No additional exclusion threshold was imposed based on quality-appraisal scores. This approach is consistent with examining methodological limitations within the synthesis rather than excluding studies solely based on quality judgments (Glass et al., 1981). Appraisal findings informed the interpretation of the evidence and its limitations.

Coding, Categorization, and Moderators

The codebook consists of five categories of variables: report characteristics (e.g., publication year, report type), research context (e.g., educational level, discipline), research design (e.g., type of experiment, type of control group, intervention length, sample size, pre-test inclusion), AI (e.g., AI applications, pedagogical role), and SRL (e.g., SRL strategies, SRL measure). The 34 retained reports were divided into two parts. The first and second authors coded 28 reports, and the third and fourth authors coded six reports. Each report was coded by two authors to establish interrater reliability. Within each group, the corresponding two authors independently coded all full-text reports. Percentage agreement was calculated separately for each reviewer pair across all coding fields listed in Appendix C. Disagreements between the independently completed coding forms were highlighted and counted, with each discrepant field counted as one disagreement. Percentage agreement was calculated as one minus the number of disagreements divided by the total number of coding fields compared, multiplied by 100. Agreement before discussion was 95.24% for the first and second authors, who coded 28 reports, and 93.33% for the third and fourth authors, who coded six reports. All disagreements were resolved through discussion to produce the final consensus dataset.

The coded textual information was analyzed and categorized below and used for moderator analyses. The codes and their descriptions were also organized into a codebook table in Appendix C.

Research Context Moderators

Three moderators were related to the research context. These were discipline (education, language, engineering, science), whether the study focused on STEM (Science, Technology, Engineering, and Mathematics) or non-STEM, and educational level (elementary, secondary, and higher education).

Research Design Moderators

Five moderators were related to research design. These were type of experiment (experimental or quasi-experimental), intervention length (less than 1 week, 1 to 5 weeks, 6 to 10 weeks, 11 to 16 weeks), sample size (1 to 50, 51 to 100, 101 to 150, and more than 150), whether a pretest was included (yes or no), and type of control group (passive or active control). To guide coding of passive and active control, we first separated studies that provided no instructional support in the control group (passive control) from those that provided some form of instructional support (active control). Passive control referred to studies where the control group did not receive any instructional support from AI, teachers, peers, or other tools. Active control referred to studies in which both groups received support, but the experimental group received AI-based support while the control group received a non-AI alternative.

For example, studies comparing AI-generated feedback with teacher or system feedback were coded as active control because both groups received feedback. However, if only the experimental group received AI-generated feedback and the control group received no feedback, the study was coded as passive control. A more detailed classification of active-control conditions was considered, but subdividing the 17 studies with active controls resulted in sparse categories. We therefore retained the broader active–passive distinction for moderator analysis.

SRL Moderators

Two moderators are associated with SRL strategies. They are SRL phase and SRL measure. Aligned with Zimmerman's three phase cyclic model of SRL (Zimmerman 2000), SRL strategies are categorized into forethought, performance, and reflection phase. In addition, we categorized SRL measures based on when they were administered (Rovers et al., 2019; Schellings 2011; Veenman 2005). Measures collected in real time during a learning task (e.g., trace/log data, observations) were classified as online measures. Measures administered before or after a learning task (e.g., self-report questionnaires) were classified as offline measures.

AI Moderators

Three moderators were associated with AI. These were AI applications, whether the application used Generative AI (yes or no), and the pedagogical role of AI.

For the moderator of AI applications, four categories were used. The first category was conversational agent (CA) (e.g., chatbot, pedagogical agent). These systems interact with learners through natural language and may appear as chatbots, avatars, or named virtual companions such as "Bubi" (Liang et al., 2024) or "Replika" (Ortega-Ochoa et al., 2024). The second category was LLM-supported systems (e.g., ChatGPT, GPT, BERT), which involve the direct use of foundation models as tools, such as for writing or text generation, without a specialized persona (e.g., Chen et al., 2025; Zheng et al., 2021). The third category was intelligent tutoring or adaptive learning systems, which adjust content, difficulty, or pacing based on learner performance using rule-based or machine learning approaches (e.g., Chu et al., 2023). The fourth category was learning analytic systems (e.g., dashboards or visualization tools) that collect and display learning data to support awareness and reflection without directly guiding task completion (e.g., Liao et al., 2024).

For the moderator of the pedagogical role of AI, four roles were identified. The first was feedback or assessment, where AI evaluates student work and provides information on correctness or quality. The second was monitoring or analytics, where AI tracks or visualizes learning behaviors to increase metacognitive awareness without offering evaluative comments (e.g., Liao et al., 2024). The third was scaffolding or facilitation (e.g., prompts, hints, recommendations), where AI guides learning processes by offering strategies or structured support that help students progress in tasks (e.g., Chen et al., 2025). The fourth was mixed, which included systems that combine multiple pedagogical functions and cannot be categorized under a single role (e.g., Liang et al., 2024).

Effect Size

In this meta-analysis, Hedges' g (Hedges, 1981) was used as the measure of the effect of AI-PI on SRL. Hedges' g was calculated as the standardized mean difference between the experimental and control group, with a small-sample bias correction applied. When pre and posttest data were both reported, the effect size was computed using a pretest-adjusted difference approach, which accounts for initial group differences and provides a more accurate estimate of the effect. The pre-posttest correlation was assumed to be .50 when not reported. For studies reporting test statistics (e.g., t, F, or

χ^2) instead of means and standard deviations, Hedges' g was derived from those statistics using established conversion formulas.

Handling Dependence Among Effect Sizes

Because many primary studies (e.g., Fleur et al., 2023; Lim et al., 2023) reported multiple effect sizes for different SRL strategies, these effect sizes from the same study were statistically dependent. Ignoring such dependency can result in underestimated standard errors and inflated Type I error rates (Hedges et al., 2010). To address this issue, the present study employed the robust variance estimation (RVE, Hedges et al., 2010) and implemented in the robumeta package (Fisher & Tipton, 2015). The RVE method extends random-effects meta-regression to accommodate multiple, correlated effect sizes from the same study. It does so by adjusting the weighting and estimating robust standard errors that remain valid even when the exact covariance structure is unknown. A within-study correlation of $\rho = .80$ was assumed among effect sizes (Tanner-Smith & Tipton, 2014), reflecting a moderate to high degree of dependence typically observed in educational and psychological research. The small-sample correction recommended by Tipton (2015) was applied to improve the accuracy of statistical inference given the modest number of studies included in the synthesis.

Data Analysis

Analyses associated with research questions were conducted in R Studio using random-effects models estimated with the RVE method implemented in the robumeta package (Fisher and Tipton, 2015). Random-effects models were used to account for potential variability in true effect sizes across studies, recognizing that the impact of AI-PI on SRL may differ across moderators. As suggested by the RVE method (Hedges et al., 2010), Heterogeneity (variation in the true effect sizes, Borenstein et al., 2021) was measured using the between-study variance (τ^2) and the proportion of variance attributable to heterogeneity (I^2).

The first research question examined the overall effect of AI-PI on SRL as a global construct. A random-effects RVE model was estimated to obtain the Hedges' g across all included studies. The standard error and the 95% confidence interval were reported to indicate the precision of the estimate. Heterogeneity was assessed using τ^2 and I^2 to evaluate whether meaningful variability existed across studies.

The second research question examined whether the effects of AI-PI differed across SRL subdomains (metacognitive, cognitive, and resource management strategies). Three separate random-effects RVE meta-analyses were conducted to estimate the overall Hedges' g for each subdomain.

For the third research question, moderator analyses were also conducted using RVE-based meta-regression models. Each moderator was examined in a separate meta-regression model. In each model, the intercept represented the estimated mean effect size for the reference category, and regression coefficients represented differences between the reference category and other levels of the moderator. Statistical significance of these coefficients was evaluated using t tests to assess whether the effects of AI-PI differed across moderator levels. To obtain all possible pairwise comparisons among moderator levels, the reference category was systematically rotated, and a series of models was estimated for each moderator, with a different level specified as the reference in each model. This procedure yielded a complete set of pairwise coefficient comparisons and allowed for direct examination of differences across all categories. A moderator was considered significant if at least one pair of comparison yielded a statistically significant t test.

Publication bias was examined using the funnel plot, Egger's regression test (Egger et al., 1997), and the trim-and-fill method (Duval & Tweedie, 2000). Analyses were conducted at the study-level to

preserve the assumption of independence underlying funnel plot interpretation. Specifically, multiple effect sizes within studies were aggregated into a single study-level effect size using inverse variance weighting. Publication bias analyses were conducted in R Studio using the metafor package (Viechtbauer, 2010).

Results

A total of 91 independent and dependent effect sizes were retrieved from 34 included studies. Descriptive information about the 34 included studies were shown in Table 2. The 34 studies included in this meta-analysis were published primarily in recent years, with half appearing in 2024 or 2025 (accessible online in 2024) and most of the remainder published between 2021 and 2023. The majority were journal articles, with a small number of conference papers and no dissertations meeting the inclusion criteria. Most studies were conducted in higher education settings, with fewer focusing on secondary or elementary education. The studies covered multiple disciplinary domains, most commonly language learning and science, followed by education and engineering. Both experimental and quasi-experimental designs were common, with slightly more quasi-experimental studies. Active and passive control groups were used at similar rates. Intervention length varied widely, from less than one week to up to 16 weeks, with short-term interventions being most frequent.

Table 2

Descriptive Information for Included Studies (n = 34)

Variables	Numbers
Year of Publication	
2024 & 2025	17
2021 - 2023	10
2020 or before	7
Type of Publication	
Journal article	30
Conference paper	4
Dissertation	0
Educational level	
Higher Education	25
Secondary Education	5
Elementary	4
Discipline	
Engineering	1
Science	9
Education	3
Language	13
Research Design	
Experimental	15
Quasi-experimental	19
Type of Control group	
Active control	17
Passive control	18
Intervention Length (weeks)	
<1	15
1 - 5	4
6 - 10	6
11-16	8

Note 1. Totals do not sum to 34 for some variables because of missing study-level information

Note 2. Totals exceed 34 for control group type because one study included both active and passive controls

RQ1: Overall Effect of AI-Based Interventions on SRL and Heterogeneity

As shown in Table 3, the overall mean effect size of AI-PI on SRL across 34 independent studies (91 effect sizes) was statistically significant (Hedges' g = 0.76, SE = 0.12, p < .001). This indicates a moderate-to-large positive effect (Sawilowsky, 2009) of the AI-PI on SRL.

Between-study variation in effect sizes (heterogeneity) was substantial (I^2 = 88.34%, τ^2 = 0.36), suggesting considerable variability in true effects across studies. The forest plot in Appendix A provides a visual summary of between-study heterogeneity. Given this level of heterogeneity, subsequent moderator analyses were necessary to examine potential study level factors (e.g., educational level, discipline, type of AI application, and AI pedagogical role) that may help explain differences in observed effects.

Table 3

Summary Effect of AI-PI on SRL Overall

Effect Size	Estimate	SE	t(df)	95% CI	p
Overall Hedges' g	0.76	0.12	6.41 (32.7)	[0.52, 1.01]	***

Note. *** p < .001

RQ2 Effects of AI-Based Pedagogical Interventions (AI-PI) on Subdomains of SRL and Heterogeneity

The second research question in this meta-analysis focuses on the effects of AI-PI on the subdomains of SRL including cognitive, metacognitive, and resource management strategies. As shown in Table 4, across the three SRL subdomains, AI-PI demonstrated consistently positive effects, though the magnitude of effects varied.

For metacognitive strategies, the analysis included 20 studies and 38 effect sizes. The pooled effect was large (g = 0.73, p < .001). For cognitive strategies, the model synthesized 9 studies and 15 effect sizes. The pooled effect was moderate (g = 0.48, p < .05). For resource-management strategies, the analysis included 11 studies and 21 effect sizes. The pooled effect was large (g = 0.93, p < .05).

Taken together, the results show that AI-PI had a positive and meaningful impact on all three SRL subdomains. The strongest effects emerged for resource management and metacognitive regulation, while cognitive strategy use showed more moderate improvements.

Table 4

Effects of AI-PI on Subdomains of SRL Strategies

SRL Subdomain	Hedges' g	SE	95% CI	t	df	p
Metacognitive	0.73	0.14	[0.44, 1.02]	5.21	8.00	***
Cognitive	0.48	0.20	[0.04, 0.90]	2.39	7.96	*
Resource Management	0.93	0.30	[0.24, 1.62]	3.02	9.90	*

Note. * p < .05, ** p < .01, *** p < .001

RQ3: Moderator Analyses

Moderators Associated with SRL Strategies

SRL Strategies by Learning Phase. The reflection phase yielded the largest mean effect size (g = 1.02, SE = 0.23), followed by the performance phase (g = 0.86). The forethought phase and the mixed phase both showed moderate effect sizes. Results from the meta-regression models indicated that pairwise t tests comparing regression coefficients across learning phases were not statistically significant ($p > .05$), suggesting that learning phase did not function as a significant moderator of SRL outcomes.

SRL Measure. Although offline measures yielded a larger effect size (g = 0.85) than online measures (g = 0.62), the t-test comparing regression coefficients for online and offline measures were not statistically significant ($p > .05$), indicating that SRL measure did not significantly moderate intervention effects**.**

Table 5

SRL Phase and Measure

Moderator & Subgroups	K (ES)	Mean	SE
SRL Strategies by learning phase			
Forethought Phase	8	0.72	0.33
Performance Phase	45	0.86	0.21
Reflection Phase	9	1.02	0.23
mixed	13	0.73	0.20
SRL Measure			
Online Measure (e.g., log, trace data)	31	0.62	0.12
Offline Measure (e.g., self-report survey)	59	0.85	0.18

Note. SE = standard error, ES = effect size

Moderators Associated with AI

Based on the results from pairwise t tests of regression coefficients across moderator levels, both the type of AI application and the pedagogical role of AI significantly moderated the effects of AI-PI on SRL, whereas the use of generative AI versus non-generative AI did not.

Type of AI application. Intelligent tutoring or adaptive learning systems yielded the largest mean effect size (g = 1.20), followed by conversational agents (g = 0.88), LLM-empowered systems (g = 0.56), and learning analytic systems (g = 0.37). Pairwise comparisons indicated that conversational agent-based interventions were significantly more effective in supporting SRL than learning analytic systems.

Pedagogical role of AI. Feedback or assessment showed the largest mean effect size (g = 1.11), followed by scaffolding or facilitation (g = 0.80), mixed roles (g = 0.71), and monitoring or analytics (g = 0.32). Pairwise comparisons indicated significantly stronger effects for scaffolding or facilitation roles than for monitoring or analytic roles.

Generative AI versus non-generative AI. The t-test comparing regression coefficients for generative AI and non-generative AI were not statistically significant ($p > .05$), indicating using generative AI or not is not a significant moderator.

Table 6

AI Applications and Pedagogical Role of AI

Moderator & Subgroups	K (ES)	Mean	SE
Type of AI			
Conversational Agent (e.g., chatbot, pedagogical agent)	23	0.88	0.15
Intelligent tutoring/adaptive learning system	23	1.20	0.38
Learning analytic system (e.g., learning analytic dashboard)	12	0.37	0.10
LLM-empowered system (e.g., GPT, BERT)	33	0.56	0.19
Pedagogical role of AI			
Feedback/Assessment	18	1.11	0.51
Monitoring/Analytics	6	0.32	0.09
Scaffolding/Facilitation	38	0.80	0.16
Mixed	29	0.71	0.16
Generative AI or not			
Yes	35	0.59	0.14
No	56	0.88	0.17

Note. SE = standard error, ES = effect size

Moderators Associated with Research Context

The moderator analysis showed that educational level functioned as a significant moderator of the effects of AI-PI on SRL, whereas discipline and STEM versus non-STEM affiliation did not.

Educational Level. Effect sizes differed across stages. Higher education showed the largest mean effect size (g = 0.83), followed closely by elementary education (g = 0.81), while secondary education demonstrated a smaller effect (g = 0.41). The elementary-secondary comparison yielded an estimated difference of g = 0.40, which was marginally significant. Higher-secondary and higher-elementary comparisons did not meet the .05 threshold. These findings should be interpreted cautiously, particularly given the borderline elementary-secondary result and the low degrees of freedom for the higher-elementary comparison.

Discipline. The moderator was not significant. Although effect sizes varied descriptively across subject areas, with language learning (g = 0.97), education (g = 0.76), science (g = 0.73), and engineering (g = 0.59), pairwise comparisons did not support significant differences across disciplines.

STEM versus non-STEM. This factor did not significantly moderate the effects of AI-PI on SRL, given that mean effect sizes were comparable between non-STEM disciplines (g = 0.82) and STEM disciplines (g = 0.66).

Table 7

Educational Level and Discipline

Moderator & Subgroups	K (ES)	Mean	SE
Educational Level			
Elementary Education	7	0.81	0.15
Secondary Education	8	0.41	0.08
Higher Education	76	0.83	0.16
Discipline			
Education	5	0.76	0.44
Engineering	6	0.59	0.00
Language	31	0.97	0.28

Moderator & Subgroups	K (ES)	Mean	SE
Science	17	0.73	0.15
STEM or not			
Yes	29	0.66	0.13
No	45	0.82	0.20

Note. SE = standard error, ES = effect size

Moderators Associated with Intervention Design

The moderator analyses examining experimental design, type of control group, pretest inclusion, intervention length, and sample size indicated that none of these study design characteristics functioned as significant moderators of the effects of AI-PI on SRL.

Although these moderators were not statistically significant, descriptive patterns in effect sizes were observed. Effect sizes were comparable for experimental (g = 0.77) and quasi-experimental designs (g = 0.76), as well as for studies using active versus passive control groups. Studies that included pretest measures showed larger mean effect sizes (g = 0.94) than those without pretests (g = 0.59), although this difference was not supported by pairwise comparisons.

For intervention length, shorter interventions (less than one week; g = 0.84) and longer length interventions (11 to 16 weeks; g = 0.85) yielded larger mean effect sizes than moderate length interventions (6 to 10 weeks; g = 0.62), but no pairwise differences were statistically significant. Similarly, sample size did not significantly moderate effects, although larger effect sizes were observed in studies with small to medium samples compared with studies with larger samples.

Table 8

Research Design, Intervention Length, sample size

Moderator & Subgroups	K (ES)	Mean	SE
Experimental or Quasi-experimental			
Experimental	31	.77	.21
Quasi-experimental	60	.76	.14
Type of Control Group			
Active control	44	.73	.16
Passive control	47	.80	.18
Accounted for Pre-test or not			
Yes	40	.94	.21
No	51	.59	.11
Intervention length			
<1week	45	.84	.24
1 – 5 weeks	10	.70	.11
6 – 10 weeks	14	.62	.16
11 – 16 weeks	21	.85	.22
Sample Size			
1 - 50	24	.93	.31
51 - 100	41	.62	.11
101 - 150	14	1.28	.42
> 150	12	.37	.09

Note. SE = standard error, ES = effect size

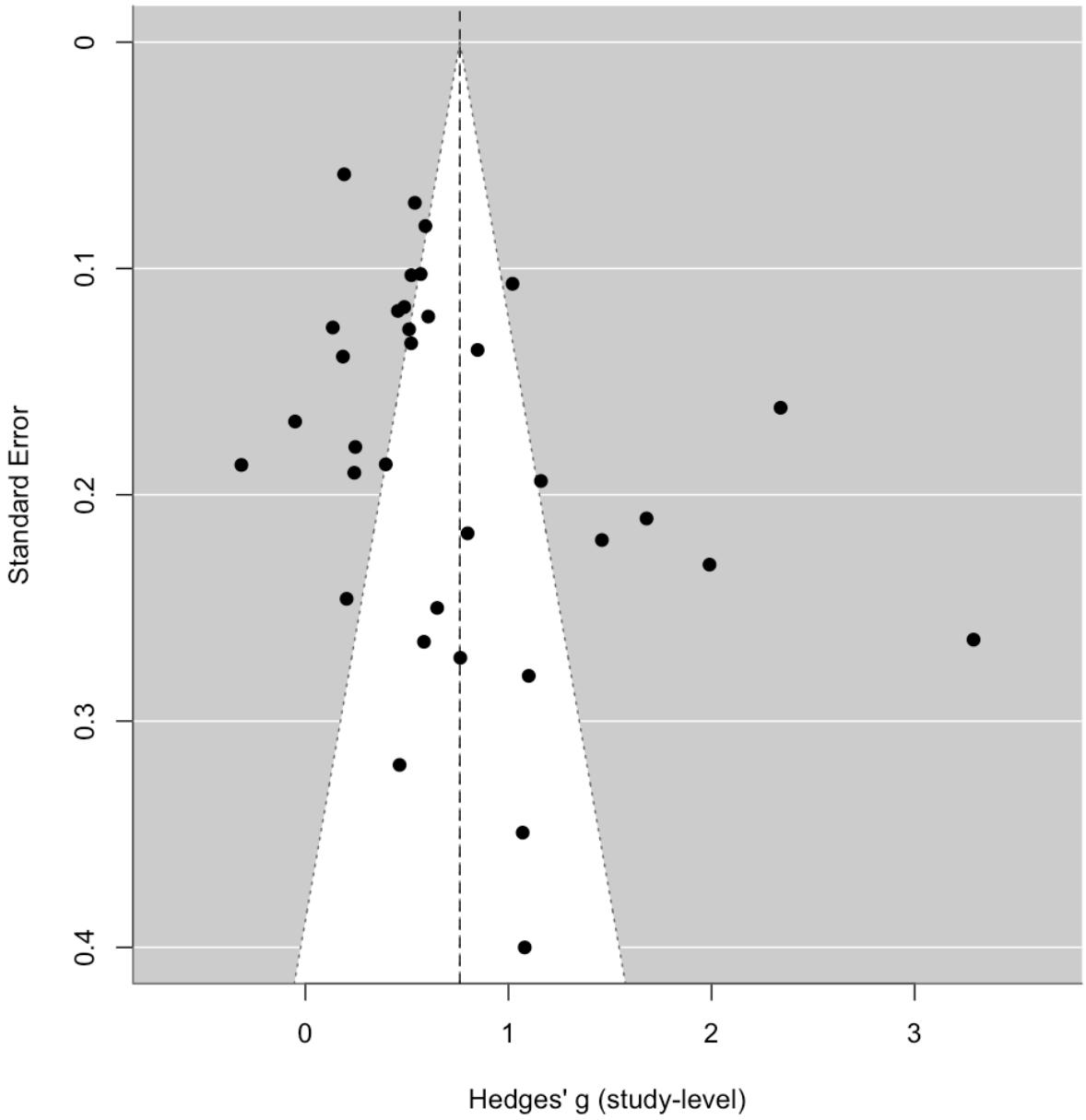
Detailed results for all pairwise moderator comparisons, including estimated differences, robust standard errors, t statistics, degrees of freedom, and unadjusted p-values, are provided in the four tables in Appendix D.

Publication Bias

The funnel plot (shown in Figure 2) suggested some asymmetry, with studies with larger standard errors tending to report larger positive effects. Egger's regression test did not provide statistically significant evidence of funnel plot asymmetry ($p > .05$). Trim and fill analysis did not identify any missing studies and yielded a pooled effect size that was comparable to the original estimate, indicating that potential publication bias was unlikely to substantially influence the overall findings. Taken together, these results suggest that although small study effects cannot be entirely ruled out, there is no strong evidence that publication bias meaningfully threatens the validity of the observed effects.

Figure 2

Funnel Plot



Discussion

This meta-analysis examined the effects of AI-PI on learners' SRL and its subdomains, including cognitive, metacognitive, and resource management strategies. The study also investigated whether these effects varied across four groups of moderators related to research context, intervention design, AI characteristics, and SRL measurement. The discussion interprets the main findings in

relation to prior research, provides implications for practice and research, acknowledges limitations and strengths, and outlines directions for future work.

RQ 1 & 2 Effects of AI-based interventions on SRL and SRL Subdomains

The first part of our meta-analysis showed that AI-PI led to a significant improvement in learners' use of SRL strategies, with a medium to large overall effect (g = .76). This finding is consistent with a previous meta-analysis by Yin et al. (2025), which reported a moderate positive effect of pedagogical agents on SRL, as well as qualitative evidence from recent systematic reviews highlighting the potential of AI for supporting SRL (e.g., Banihashem et al., 2025). At a more detailed level, our analyses of SRL subdomains revealed that AI-PI had significant and positive effects across all three subdomains of SRL (cognitive, metacognitive, and resource management strategies). Although all effects were positive, their magnitudes differed. Resource management showed the strongest effect, followed by metacognitive strategies, while the effect on cognitive strategies was comparatively smaller.

Several factors may help tentatively explain this pattern. First, many AI systems in the included studies were intentionally designed to support learners' resource-management behaviors such as help-seeking (e.g., Chen et al., 2025; Chou et al., 2018). For example, Chen et al. (2025) developed an online learning platform that guided students in diagnosing problems, seeking help, evaluating the quality of that help, and integrating feedback from ChatGPT. Because these designs provide frequent opportunities to practice specific strategies while interacting with AI tools, learners may experience more immediate and consistent benefits in these areas. Second, the ways learners interact with AI systems naturally align with key metacognitive processes such as planning, monitoring, and evaluating their own learning. For instance, tools like learning-analytics dashboards help learners monitor their progress (Liao et al., 2024), and AI-generated prompts can encourage reflection or reconsideration of strategies (e.g., Essel et al., 2024). These forms of support may reinforce the processes within the metacognitive cycle and may explain why metacognitive outcomes responded strongly to AI-PI. Third, cognitive strategies such as rehearsal, elaboration, and organization depend heavily on learners' prior knowledge, personal ways of thinking, and learning beliefs (Pintrich, 2000). These deeper and internal processes often take longer to develop and may not change substantially during short-term interventions. This may help explain the relatively smaller effect observed in the cognitive strategy domain. These explanations, however, remain tentative: they are hypotheses derived from the characteristics of the included interventions and SRL theory, and the present meta-analysis did not directly test these mechanisms.

These results have meaningful implications for educators, instructional designers, and AI system developers. First, integrating AI-PI into instruction may be especially helpful for supporting students' management of both internal resources (e.g., effort, time) and external resources (e.g., learning environments, peers, teachers). Second, higher-order thinking skills such as goal setting, self-monitoring, and self-evaluation appear to benefit more from AI-based support than basic cognitive strategies like rehearsal or elaboration. Third, when the goal is to improve cognitive strategies specifically, AI systems should be intentionally designed to provide direct, content-rich scaffolds for these skills (for example, prompts that encourage elaboration or organization). Pairing AI support with teacher guidance may be another good strategy to increase the likelihood of improving students' use of cognitive strategies.

RQ 3 Moderators

Differences Across Educational Levels

Estimated effects were descriptively largest in higher education (g = 0.83), followed closely by elementary education (g = 0.81), and were smaller in secondary education (g = 0.41). However, only the elementary-secondary comparison met the .05 threshold, and this result was borderline (p = .04996). This result contrasts with findings from another meta-analysis by Yin et al. (2025), which found that educational level was not a significant moderating variable in the relationship between pedagogical agents and SRL. One possible explanation for this discrepancy lies in the scope of AI-PI examined. While Yin et al. focused specifically on pedagogical agents, which typically embedded characters that facilitate learning (Yin et al., 2025; Nguyen, 2023), our study included a broader range of AI-based tools such as adaptive learning systems, personalized learning platforms, conversational agents, and learning analytics dashboard. Differences in the range of AI applications examined may contribute to the differing patterns across reviews, although this explanation was not directly tested.

Furthermore, our dataset shows an uneven distribution of studies across educational levels: 25 studies were conducted in higher education contexts, compared to only four in elementary education and five in secondary education. The uneven distribution of studies limits the precision of comparisons involving elementary and secondary education and constrains the generalizability of the observed pattern. This finding aligns with the conclusions of a systematic review by Banihashem et al. (2025), who observed that AI-SRL research has predominantly focused on higher education, with minimal attention paid to K-12 settings. In their review, 74% of studies were conducted in higher-education contexts, while only 20 studies focused on middle or high school, and just one study addressed primary education. Such an unbalanced research landscape leaves the true effects of AI interventions on SRL in K-12 education uncertain, highlighting the need for more inclusive research across educational levels.

Moderators Associated with AI (AI applications, GenAI or not, AI pedagogical Role)

The exploratory moderator comparisons suggest that specific AI applications and pedagogical roles may be relevant to intervention effectiveness. Differences were detected between conversational agents and learning analytic systems and between scaffolding/facilitation and monitoring/analytics, whereas no statistically significant difference was detected between generative and non-generative AI. However, this pattern does not establish that pedagogical function is more consequential than generative AI status. The findings instead identify specific contrasts that warrant further investigation.

With respect to AI application type, intelligent tutoring and adaptive learning systems had the largest estimated mean effect on SRL, followed by conversational agents such as chatbots and pedagogical agents. Learning analytics dashboards and LLM-supported systems had smaller, though still positive, estimates. A tentative explanation for this descriptive pattern is that these technologies support regulation in different ways. Intelligent tutoring and adaptive systems provide ongoing guidance and adaptive prompts, while conversational agents support SRL through interactive dialogue. In contrast, learning analytics systems primarily present descriptive information that learners must interpret and act upon. The observed difference between conversational agents and learning analytics tools is consistent with the possibility that interactive guidance offers benefits beyond making learning processes visible. This interpretation relates to the work of Yin et al. (2025), whose meta-analysis of pedagogical agents found stronger effects for more interactive and supportive agents. Our findings raise the possibility that a similar pattern may extend to other AI applications. However, this interpretation remains tentative: the only AI application comparison that met the unadjusted .05 threshold was conversational agents versus learning analytic systems. This specific contrast does not establish a general hierarchy of effectiveness across AI applications, and the proposed explanation was not directly tested in the present meta-analysis.

The pedagogical role of AI further clarified these findings. AI interventions that focused on feedback, assessment, or scaffolding had the largest estimated mean effects on SRL, followed by

systems with mixed roles, whereas monitoring or analytics-only tools produced the smallest gains. One tentative explanation for this descriptive pattern is that explicit guidance may help learners translate information about their learning into subsequent actions. Feedback and scaffolding roles directly support core regulatory processes, while monitoring tools place most of the interpretive and planning burden on learners. This interpretation aligns with Yin et al. (2025), who found that pedagogical agents framed as virtual assistants were more effective for SRL than mentor-type or peer-like agents. The scaffolding/facilitation–monitoring/analytics contrast suggests a possible benefit of providing structured guidance beyond displaying learning information. However, the remaining comparisons did not meet the .05 threshold, and these exploratory findings do not establish that feedback, scaffolding, or coaching roles are generally superior.

No statistically significant difference was detected between generative and non-generative AI interventions, although both categories had positive estimated mean effects. This finding does not establish equivalent effectiveness or demonstrate that generative capabilities are unimportant for SRL. Differences in intervention design and the limited evidence base may have constrained our ability to detect differences between these categories.

Generative models may expand the possibilities for supporting SRL by enabling richer dialogue, more flexible explanations, and context-sensitive assistance. Clear pedagogical roles and intentional scaffolding may help translate these capabilities into SRL benefits, although the present meta-analysis did not directly test this possibility. As a tentative design implication, educators and developers could consider how generative AI supports specific instructional functions, such as feedback, scaffolding, or monitoring. Future research should examine whether and how particular generative features interact with these pedagogical roles to support learners' SRL strategies.

Moderators Associated with SRL Phase and Measure

Our moderator analyses indicated that the effects of AI-PI on SRL did not differ significantly across SRL phases or by the type of SRL measure used. With respect to SRL phases, including forethought, performance, reflection, and mixed phases, the absence of significant differences may be partly attributable to the unequal distribution of effect sizes across categories. For example, the performance phase contributed 45 effect sizes, whereas the mixed, forethought, and reflection phases contributed 13, 8, and 9 effect sizes, respectively. Such imbalances can reduce statistical power and limit the ability to detect meaningful differences across SRL phases.

For SRL measures, both online measures (e.g., trace/log data, observations) and offline measures (e.g., self-report questionnaires) showed comparable effects, indicating that measurement type did not significantly influence the impact of AI-PI on SRL. These findings suggest that AI-PI have relatively consistent effects across different SRL phases. While current research often highlights the advantages of log or trace data over self-report questionnaires for assessing SRL (Du et al., 2023), our results indicate that both online and offline measures are appropriate for evaluating the effects of AI-PI on SRL

Moderators Associated with Intervention Design

Unexpectedly, none of the intervention design characteristics examined in this study, including experimental design, sample size, intervention length, pretest inclusion, and type of control group, emerged as significant moderators of the effects of AI based pedagogical interventions on SRL. This finding differs from Yin et al. (2025), who reported that intervention length significantly moderated the effects of pedagogical agents on SRL. One possible explanation for this discrepancy is that Yin et al.'s meta-analysis focused on a single category of AI, whereas the present study synthesized a much broader range of AI tools and systems. It is possible that intervention length plays a more important role for

certain types of AI than for others, and that combining diverse AI applications within a single analysis reduced the ability to detect moderation effects.

Another methodological factor that may have contributed to the absence of significant moderators is the use of robust variance estimation. RVE is well suited for handling dependent effect sizes and provides more accurate standard error estimates by accounting for within study dependence and applying small sample corrections (Tanner-Smith et al., 2016). However, these adjustments also make moderator tests more conservative. As a result, even when some design characteristics were associated with meaningful differences in estimated effects, those differences were less likely to reach statistical significance. This may help explain why these intervention design variables did not emerge as statistically significant moderators in the present study.

From a practical perspective, we did not detect statistically significant pairwise differences by experimental design, intervention length, sample size, pretest inclusion, or control-group type. These findings do not establish equivalent effectiveness or robustness across study designs and implementation conditions. The limited and uneven evidence across categories may have constrained our ability to detect differences.

In particular, the absence of a statistically significant sample-size comparison does not demonstrate that intervention effectiveness is independent of study size or that benefits generalize equally to small- and large-scale implementations. Similarly, nonsignificant comparisons involving pretest inclusion and control-group type do not establish stability across design approaches. Future studies should examine these conditions using rigorous designs and sufficiently large samples.

At the institutional policy level, these findings support considering SRL outcomes alongside academic achievement when evaluating AI-based educational tools. Schools and universities could establish procurement and evaluation criteria that examine how a tool supports learners' planning, monitoring, and reflection, and provide professional development for educators to integrate these functions into instruction.

Strengths, Limitations, and Future Research

A key strength of our meta-analysis is that we examined the effects of AI-PI not only on overall SRL but also on its major subdomains including metacognitive, cognitive, and resource management strategies. Future meta-analyses may consider an even more fine-grained level of analysis by moving from subdomains to specific SRL strategies. For example, researchers could directly evaluate the effects of AI-PI on strategies such as goal setting, help seeking, and time management. Conducting analyses at this level, however, would require a substantially larger pool of studies and effect sizes. In addition, whereas many existing AI-related meta-analyses focus primarily on generative AI (e.g., Zhu et al., 2025) or on a single tool such as ChatGPT (Deng et al., 2025), the present study synthesized evidence across a diverse range of AI systems, including intelligent tutoring systems, conversational agents, LLM-empowered systems, and learning analytics tools. This intentional inclusion provides a more comprehensive picture of how different forms of AI tools could be integrated into instructions to support SRL. Despite the broader field coverage, the research-design block may have reduced sensitivity when eligible studies used different terminology or omitted design terms from searchable fields.

This broad scope also shaped the search and screening process. Diverse SRL and AI terminology and multiple databases were used to reduce the risk of missing relevant studies, yielding 4,418 unique records. However, only 34 studies met the focused eligibility criteria, which required experimental or quasi-experimental evaluations of AI-PI and sufficient quantitative information to calculate effect sizes. Many records did not meet these requirements, and some quantitative reports

lacked usable data. Future reviews could refine search strategies against known eligible studies to improve precision while preserving coverage. The exclusion of reports with insufficient quantitative information remains a limitation.

One limitation of this meta-analysis is the relatively small number of studies that met the inclusion criteria (n = 34), which may limit the generalizability of the findings. We applied small-sample corrections in the robust variance estimation models (Tipton, 2015) and used RVE to retain the available effect sizes (k = 91) while accounting for within-study dependence (Hedges et al., 2010). However, these approaches do not overcome the limited evidence within individual moderator categories. For example, only 17 studies included active controls, and further subdivision by specific comparison condition yielded sparse categories. We therefore retained the broader active-passive classification, although combining heterogeneous active-control conditions limits interpretation of that comparison. Future syntheses with a larger evidence base could examine specific control types. Besides, multiple unadjusted pairwise comparisons increased the potential for Type I error, warranting cautious interpretation.

During the screening stage, following joint calibration of 50 records, title and abstract records were screened by one reviewer each. Although full-text reports were independently assessed by two reviewers, the original decisions were not fully preserved during conflict resolution, preventing calculation of agreement before discussion.

Another limitation of the existing literature on AI-PI for SRL is the lack of clear differentiation among specific AI affordances, such as personalization or adaptation, automation, and real time assistance. Many existing studies (e.g., Chen, 2024; Liu et al., 2024) examine AI interventions at the tool level by comparing learning activities completed with or without AI support. While such designs are useful for establishing whether a given AI tool can support SRL, they provide limited insight into which specific feature or affordance of the tool are responsible for observed effects. As a result, the mechanisms through which AI supports learners' SRL often remain unclear. Future research would benefit from designs that isolate and systematically vary individual AI affordances. For example, a study could compare two versions of the same AI system, one that provides personalized prompts based on learners' performance and another that delivers the same prompts in a non-personalized manner. Similarly, researchers could manipulate the timing of AI-support by comparing real-time assistance with delayed feedback while holding other features constant. Such designs would allow researchers to more directly examine how a specific AI affordance contribute to the development of SRL.

Another direction for future research involves expanding the focus of meta-analyses in the field of AI in education. Most existing meta-analyses examine how AI-PI influence academic achievement (e.g., Dong et al., 2025; Zheng, et al., 2023). However, meta-analyses evaluating the impact of AI on learners' behaviors, beliefs, or motivations are still very limited, which are areas that are central to understanding how students engage with and benefit from AI-supported learning environments. More meta-analyses in these areas are needed.

Conclusion

In this paper, we report a meta-analysis of 34 empirical studies on AI-PI's effects on SRL, exploring how these effects vary across studies, subdomains, and moderating variables. We found that, overall, AI-PI showed positive effects on SRL and its subdomains, including cognitive, metacognitive, and resource management strategies. In addition, the effectiveness of these interventions depended less on the presence of AI itself and more on how AI was designed and used instructionally. Differences in effects were most strongly associated with educational level, the pedagogical roles assumed by AI systems, and the types of AI applications implemented. In contrast, factors such as discipline, intervention length, study design characteristics, and measurement approaches did not consistently

influence outcomes. Together, these findings suggest that AI can support SRL across diverse instructional contexts while underscoring the central role of pedagogical design in shaping its impact.

A key contribution of this study is in understanding strategies for integrating AI into education in improving SRL, and in what pedagogical ways AI can best support the different phases and domains of SRL. With the widespread application of AI tools in education, many researchers have argued there is a critical need for improved teacher development in how to use these tools effectively (e.g., Pham & Sampson, 2022). However, much of the discussion on teacher training about GenAI has focused on what these tools can do, and how to use them for cognitive support. Our study found that there is substantial value also in AI providing SRL support, and we recommend that training for educators and students focus more on the why and how of AI by showing how these tools can be used effectively to support people in owning their own learning through self-regulation and metacognition. Doing so can help education move from an AI-directed paradigm towards more AI-empowered learning (Ouyang & Jiao, 2021).

Conflict of Interest

The authors declare that there are no conflicts of interest.

AI Use Statement

The authors used ChatGPT to improve the readability and language of some sentences within the manuscript. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content.

Data Availability

The data are available from the corresponding author upon reasonable request.

References

Studies included in the meta-analysis are marked with an asterisk*

Alwaqdani, M. (2025). Investigating teachers' perceptions of artificial intelligence tools in education: potential and difficulties. Education and Information Technologies, 30(3), 2737-2755. https://doi.org/10.1007/s10639-024-12903-9

*Al-Abdullatif, A. M., Al-Dokhny, A. A., & Drwish, A. M. (2023). Implementing the Bashayer chatbot in Saudi higher education: measuring the influence on students' motivation and learning strategies. *Frontiers in psychology, 14*, 1129070. https://doi.org/10.3389/fpsyg.2023.1129070

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330666

Arguedas, M., Xhafa, F., Casillas, L., Daradoumis, T., Peña, A., & Caballé, S. (2018). A model for providing emotion awareness and feedback using fuzzy logic in online learning. *Soft Computing, 22*(3), 963–977. https://doi.org/10.1007/s00500-016-2399-0

Banihashem, S. K., Bond, M., Bergdahl, N., Khosravi, H., & Noroozi, O. (2025). A systematic mapping review at the intersection of artificial intelligence and self-regulated learning. *International Journal of Educational Technology in Higher Education, 22*(1), 50. https://doi.org/10.1186/s41239-025-00548-8

Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., ... & Fischer, F. (2023). Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda. *British Journal of Educational Technology*, *54*(5), 1222-1245. https://doi.org/10.1111/bjet.13336

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis.* John Wiley & Sons.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The internet and higher education*, *27*, 1-13. https://doi.org/10.1016/j.iheduc.2015.04.007

Chang, W. L., & Sun, J. C. Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System, 126*, 103484. https://doi.org/10.1016/j.system.2024.103484

*Chen, A., Xiang, M., Zhou, J., Jia, J., Shang, J., Li, X., ... & Fan, Y. (2025). Unpacking help-seeking process through multimodal learning analytics: A comparative study of ChatGPT vs Human expert. *Computers & Education, 226*, 105198. https://doi.org/10.1016/j.compedu.2024.105198

*Chen, M. R. A. (2024). The AI chatbot interaction for semantic learning: A collaborative note-taking approach with EFL students. *Language Learning & Technology, 28*(1), 1-25. https://hdl.handle.net/10125/73586

Cheng, Z., Zhang, Z., Xu, Q., Maeda, Y., & Gu, P. (2025). A meta-analysis addressing the relationship between self-regulated learning strategies and academic performance in online higher education. *Journal of Computing in Higher Education*, *37*(1), 195-224. https://doi.org/10.1007/s12528-023-09390-1

*Chu, S. T., Hwang, G. J., Chien, S. Y., & Chang, S. C. (2023). Incorporating teacher intelligence into digital games: An expert system-guided self-regulated learning approach to promoting EFL students' performance in digital gaming contexts. *British Journal of Educational Technology, 54*(2), 534-553. https://doi.org/10.1111/bjet.13260

*Chou, C. Y., Lai, K. R., Chao, P. Y., Tseng, S. F., & Liao, T. Y. (2018). A negotiation-based adaptive learning system for regulating help-seeking behaviors. *Computers & Education, 126,* 115-128. https://doi.org/10.1016/j.compedu.2018.07.010

Dai, C. P., Ke, F., Pan, Y., Moon, J., & Liu, Z. (2024). Effects of artificial intelligence-powered virtual agents on learning outcomes in computer-based simulations: A meta-analysis. *Educational Psychology Review, 36*(1), 31. https://doi.org/10.1007/s10648-024-09855-4

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational psychology review, 28*(3), 425-474. https://doi.org/10.1007/s10648-015-9320-8

Dong, L., Tang, X., & Wang, X. (2025). Examining the effect of artificial intelligence in relation to students' academic achievement in classroom: A meta-analysis. *Computers and Education: Artificial Intelligence*, 100400. https://doi.org/10.1016/j.caeai.2025.100400

Du, J., Hew, K. F., & Liu, L. (2023). What can online traces tell us about students' self-regulated learning? A systematic review of online trace data analysis. *Computers & Education, 201*, 104828. https://doi.org/10.1016/j.compedu.2023.104828

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

*Dever, D. A., Sonnenfeld, N. A., Wiedbusch, M. D., & Azevedo, R. (2022). Pedagogical agent support and its relationship to learners' self-regulated learning strategy use with an intelligent tutoring system. In *International conference on artificial intelligence in education* (pp. 332-343). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_27

de Mooij, S., Lämsä, J., Lim, L., Aksela, O., Athavale, S., Bistolfi, I., ... & Molenaar, I. (2025). A Systematic Review of Self-Regulated Learning through Integration of Multimodal Data and Artificial Intelligence. *Educational Psychology Review, 37*(2), 54. https://doi.org/10.1007/s10648-025-10028-0

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Elom, C. O., Ayanwale, M. A., Ukeje, I. O., Offiah, G. A., Umoke, C. C., & Ogbonnaya, C. E. (2025). Does AI Knowledge Encourage Cheating? Investigating Student Perceptions, Ethical Engagement, and Academic Integrity in the Digital Age. *International Journal of Learning, Teaching and Educational Research*, *24*(4), 708-729. https://doi.org/10.26803/ijlter.24.4.33

*Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based

conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6,* 100198. https://doi.org/10.1016/j.caeai.2023.100198

Elhusseini, S. A., Tischner, C. M., Aspiranti, K. B., & Fedewa, A. L. (2022). A quantitative review of the effects of self-regulation interventions on primary and secondary student academic achievement. Metacognition and Learning, 17(3), 1117–1139. https://doi.org/10.1007/s11409-022-09311-0

*Fleur, D. S., Marshall, M., Pieters, M., Brouwer, N., Oomens, G., Konstantinidis, A., ... & van Vliet, E. A. (2023). IguideME: Supporting Self-Regulated Learning and Academic Achievement with Personalized Peer-Comparison Feedback in Higher Education. *Journal of Learning Analytics, 10*(2), 100-114. https://doi.org/10.18608/jla.2023.7853

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. arXiv preprint arXiv:1503.02220. https://doi.org/10.48550/arXiv.1503.02220

Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning: A systematic review of the literature. *Education and Information Technologies, 30*, 4493–4518. https://doi.org/10.1007/s10639-024-12881-y

Gunbatar, S. A., Sirin, G. T., Ilkyaz, O. C., & Mutlu, Y. (2025). Exploring the artificial intelligence interaction profiles of participants with different levels of teaching experience for lesson planning in the context of acids and bases. *Chemistry Education Research and Practice*, *26*(4), 977-995. https://doi.org/10.1039/D5RP00064E

*Gombert, S., Fink, A., Giorgashvili, T., Jivet, I., Di Mitri, D., Yau, J., ... & Drachsler, H. (2024). From the automated assessment of student essay content to highly informative feedback: A case study. *International Journal of Artificial Intelligence in Education, 34*(4), 1378-1416. https://doi.org/10.1007/s40593-023-00387-6

*Giorgashvili, T., Jivet, I., Artelt, C., Biedermann, D., Bengs, D., Goldhammer, F., ... & Drachsler, H. (2024). Exploring Learners' Self-reflection and Intended Actions After Consulting Learning Analytics Dashboards in an Authentic Learning Setting. In: Ferreira Mello, R., Rummel, N., Jivet, I., Pishtari, G., Ruipérez Valiente, J.A. (eds) *Technology Enhanced Learning for Inclusive and Equitable Quality Education*. EC-TEL 2024. Lecture Notes in Computer Science, vol 15159. Springer, Cham. https://doi.org/10.1007/978-3-031-72315-5_10

Glass, G. V., McGaw, B., Smith, M. L. (1981). *Meta-Analysis in Social Research*. United Kingdom: SAGE Publications.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods, 1*(1), 39-65. https://doi.org/10.1002/jrsm.5

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hemmler, Y. M., & Ifenthaler, D. (2024). Self-regulated learning strategies in continuing education: A systematic review and meta-analysis. *Educational Research Review, 45*, 100629. https://doi.org/10.1016/j.edurev.2024.100629

Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review, 28*, 100292. https://doi.org/10.1016/j.edurev.2019.100292

*Karaoğlan Yılmaz, F. G., Olpak, Y. Z., & Yılmaz, R. (2018). The effect of the metacognitive support via pedagogical agent on self-regulation skills. *Journal of Educational Computing Research, 56*(2), 159-180. https://doi.org/10.1177/0735633117707696

*Kautzmann, T. R., & Jaques, P. A. (2019). Effects of adaptive training on metacognitive knowledge monitoring ability in computer-based learning. *Computers & Education, 129*, 92-105. https://doi.org/10.1016/j.compedu.2018.10.017

Kim, N. J., & Kim, M. K. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in education, 7*, 755914. Frontiers Media SA. https://doi.org/10.3389/feduc.2022.755914

Konstantivo, L. V., Vorozhikhin, V. V., Petrov, A. M., Titova, E. S., & Shtykhno, D. A. (2023). Generative artificial intelligence in education: discussions and forecasts. *Open education*, *27*(2), 36-48. https://doi.org/10.21686/1818-4243-2023-2-36-48

Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *Science of Learning, 10*(1), 21. https://doi.org/10.1038/s41539-025-00319-0

*Liao, X., Zhang, X., Wang, Z., & Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: An experimental study. *British Journal of Educational Technology, 55*(3), 1253-1276. https://doi.org/10.1111/bjet.13424

*Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., ... & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior, 139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

*Lim, L., Bannert, M., van der Graaf, J., Fan, Y., Rakovic, M., Singh, S., ... & Gašević, D. (2024). How do students learn with real-time personalized scaffolds?. *British Journal of Educational Technology, 55*(4), 1309-1327. https://doi.org/10.1111/bjet.13414

*Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2024). Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 1-25. https://doi.org/10.1080/09588221.2024.2389923

*Liang, H. Y., Hwang, G. J., Hsu, T. Y., & Yeh, J. Y. (2024). Effect of an AI-based chatbot on students' learning performance in alternate reality game-based museum learning. *British Journal of Educational Technology, 55*(5), 2315-2338. https://doi.org/10.1111/bjet.13448

*Long, Y., & Aleven, V. (2013). Active learners: Redesigning an intelligent tutoring system to support self-regulated learning. In *European Conference on Technology Enhanced Learning* (pp. 490-495). Berlin, Heidelberg: Springer Berlin Heidelberg.

Molenaar, I. (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

*McCarthy, K. S., Likens, A. D., Johnson, A. M., Guerrero, T. A., & McNamara, D. S. (2018). Metacognitive overload!: Positive and negative effects of metacognitive prompts in an intelligent tutoring system. *International Journal of Artificial Intelligence in Education, 28*(3), 420-438. https://doi.org/10.1007/s40593-018-0164-5

Nedungadi, P., Tang, K. Y., & Raman, R. (2024). The transformative power of generative artificial intelligence for achieving the sustainable development goal of quality education. *Sustainability*, *16*(22), 9779. https://doi.org/10.3390/su16229779

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology, 55*(4), 1328-1353. https://doi.org/10.1111/bjet.13454

*Ortega-Ochoa, E., Pérez, J. Q., Arguedas, M., Daradoumis, T., & Puig, J. M. M. (2024). The effectiveness of empathic chatbot feedback for developing computer competencies, motivation, self-regulation, and metacognitive reasoning in online higher education. *Internet of Things, 25,* 101101. https://doi.org/10.1016/j.iot.2024.101101

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011–2020. *Education and Information Technologies, 27*, 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, *5*(1), 210.

https://doi. org/10.1186/s13643-016-0384-4

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

* Pan, M., Lai, C., & Guo, K. (2025). Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. *The Internet and Higher Education, 65*, 100991. https://doi.org/10.1016/j.iheduc.2024.100991

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, ´ A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pham, S. T., & Sampson, P. M. (2022). The development of artificial intelligence in education: A review in context. *Journal of Computer Assisted Learning, 38*(5), 1408–1421. https://doi.org/10.1111/jcal.12687

Pintrich, P. R. (2004). Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review 16,* 385-407. https://doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International journal of educational research*, *31*(6), 459-470. https://doi.org/10.1016/S0883-0355(99)00015-4

Qian, K., Liu, S., Li, T., Raković, M., Li, X., Guan, R., ... & Gašević, D. (2025). Towards reliable generative AI-driven scaffolding: Reducing hallucinations and enhancing quality in self-regulated learning support. *Computers & Education, 240,* 105448. https://doi.org/10.1016/j.compedu.2025.105448

Qi, X. I. A., Liu, Q., Tlili, A., & Thomas, K. F. (2025). A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions. *Computers & Education, 240*, 105465. https://doi.org/10.1016/j.compedu.2025.105465

*Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology, 14*, 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merri¨enboer, J. J. G. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning, 14*(1), 1–19. https://doi.org/10.1007/s11409-019-09188-6

*Ramachandran, A., Litoiu, A., & Scassellati, B. (2016). Shaping productive help-seeking behavior during robot-child tutoring interactions. In *Proceedings of the 11th ACM/IEEE international Conference on human robot interaction* (pp. 247–254). Piscataway, NJ: IEEE Press.

Sardi, J., Darmansyah, Candra, O., Yuliana, D. F., Habibullah, Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy (iJEP), 15*(1), 94–108. https://doi.org/10.3991/ijep.v15i1.53379

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of modern applied statistical methods, 8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Schellings, G. L. M. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning, 6*(2), 91–109. https://doi.org/10.1007/s11409-011-9069-5

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton, Mifflin and Company

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, *4*, 100124. https://doi.org/10.1016/j.caeai.2023.100124

*Serrano, M. Á., Vidal-Abarca, E., & Ferrer, A. (2018). Teaching self-regulation strategies via an intelligent tutoring system (TuinLECweb): Effects for low-skilled comprehenders. Journal of Computer Assisted Learning, 34(5), 515-525. https://doi.org/10.1111/jcal.12256

*Shi, L. (2025). The integration of advanced AI-enabled emotion detection and adaptive learning systems for improved emotional regulation. Journal of Educational Computing Research, 63(1), 173-201. https://doi.org/10.1177/07356331241296890

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2(1), 85-112. https://doi.org/10.1007/s40865-016-0026-5

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research synthesis methods, 5*(1), 13-30. https://doi.org/10.1002/jrsm.1091

*Tang, L., Shen, K., Le, H., Shen, Y., Tan, S., Zhao, Y., ... & Fan, Y. (2024). Facilitating learners' self-assessment during formative writing tasks using writing analytics toolkit. *Journal of Computer Assisted Learning, 40*(6), 2822-2839. https://doi.org/10.1111/jcal.13036

*Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education, 60*(1), e12811. https://doi.org/10.1111/ejed.12811

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological methods, 20*(3), 375. https://doi.org/10.1037/met0000011

Thottoli, M. M., Alruqaishi, B. H., & Soosaimanickam, A. (2024). Robo academic advisor: Can chatbots and artificial intelligence replace human interaction?. *Contemporary Educational Technology*, *16*(1), ep485. https://doi.org/10.30935/cedtech/13948

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology, 66*, Article e101976. https://doi.org/10.1016/j.cedpsych.2021.101976

*Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. Computers & Education, 215, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 77–99). Münster: Waxmann

Viechtbauer, W. (2010). Conducting meta-analyses in R with the met-afor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Wang, H. H., & Wang, C. H. A. (2024). Teaching design students machine learning to enhance motivation for learning computational thinking skills. Acta Psychologica, 251, 104619. https://doi.org/10.1016/j.actpsy.2024.104619

Warr, M., Oster, N. J., & Isaac, R. (2024). Implicit bias in large language models: Experimental proof and implications for education. *Journal of research on technology in education*, 1-24. https://doi.org/10.1080/15391523.2024.2395295

*Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. Frontiers in Psychology, 14, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Wiley, D. (2025). Democratizing AI participation in AI in education. https://opencontent.org/blog/archives/7821

Wu, X. Y., Radloff, J. D., Yeter, I. H., Wang, L., & Chiu, T. K. (2025). Designing artificial intelligence chatbots for self-regulated learning from a systematic review based on Habermas's three interests. Interactive Learning Environments, 1-24. https://doi.org/10.1080/10494820.2025.2563086

*Wu, T. T., Lee, H. Y., Li, P. H., Huang, C. N., & Huang, Y. M. (2024). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. Journal of Educational Computing Research, 61(8), 1539-1567. https://doi.org/10.1177/07356331231191125.

Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-Regulation interventions with a focus on learning strategies. In M. Boekaerts, P. R. Printrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 727–747). San Diego, CA, US: Academic Press. https://doi.org/10.1016/b978-012109890-2/50051-2

West, R. E., & Wiley, D. (2024). The 5Cs of openness: Articulating an open education infrastructure (OER). In R. E. West & H. Leary (Eds.), Foundations of learning and instructional design technology: Historical roots & current trends. EdTech Books. https://doi.org/10.59668/473

Yin, X., Ye, J., Yu, S., Luo, S., Nan, M., Zhao, G., ... & Zheng, L. (2025). A Meta-analysis of Pedagogical Agents promoting Self-Regulated Learning. In 2025 7th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 413-417). IEEE. https://doi.org/10.1109/CSTE64638.2025.11091971

*Yıldız, T. A. (2023). The impact of ChatGPT on language learners' motivation. Journal of Teacher Education and Lifelong Learning, 5(2), 582-597. https://doi.org/10.51535/tell.1314355

Yu, H., & Guo, Y. (2023, June). Generative artificial intelligence empowers educational reform: current status, issues, and prospects. In *Frontiers in Education* (Vol. 8, p. 1183162). Frontiers Media SA. https://doi.org/10.3389/feduc.2023.1183162

Yue, M., Jong, M. S. Y., & Ng, D. T. K. (2024). Understanding K–12 teachers' technological pedagogical content knowledge readiness and attitudes toward artificial intelligence education. *Education and information technologies*, *29*(15), 19505-19536. https://doi.org/10.1007/s10639-024-12621-2

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International Journal of Educational Technology in Higher Education, 16(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. *In Handbook of self-regulation* (pp. 13-39). Academic press.

Zhang, Z., Zhou, S., & Bai, B. (2025). A Systematic Scoping Review of Self-Regulated Learning With AI in Language Education. *International Journal of Applied Linguistics*. 1-15. https://doi.org/10.1111/ijal.70030

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. F. (2023). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments, 31*(9), 5650-5664. https://doi.org/10.1080/10494820.2021.2015693

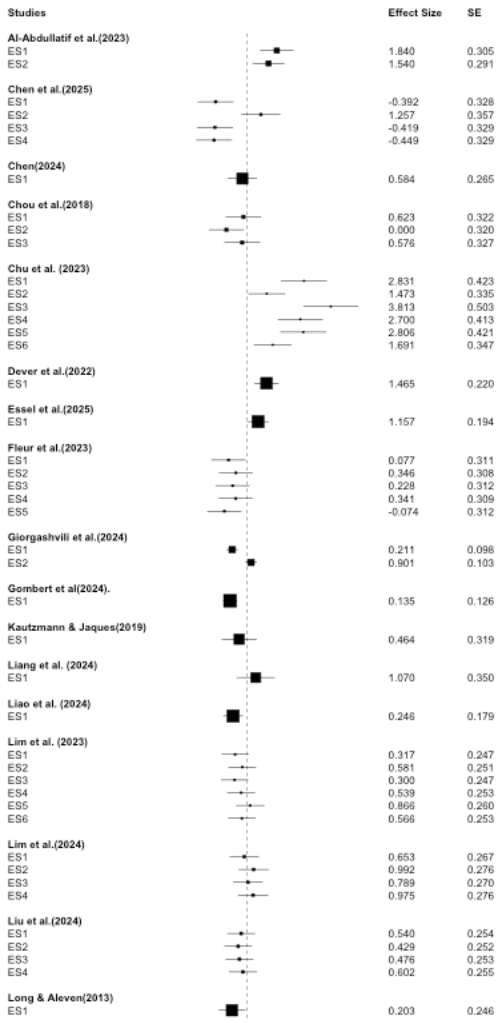
*Zheng, L., Zhong, L., Niu, J., Long, M., & Zhao, J. (2021). Effects of personalized intervention on collaborative knowledge building, group performance, socially shared metacognitive regulation, and cognitive load in computer-supported collaborative learning. *Educational Technology & Society, 24*(3), 174-193. https://www.jstor.org/stable/27032864

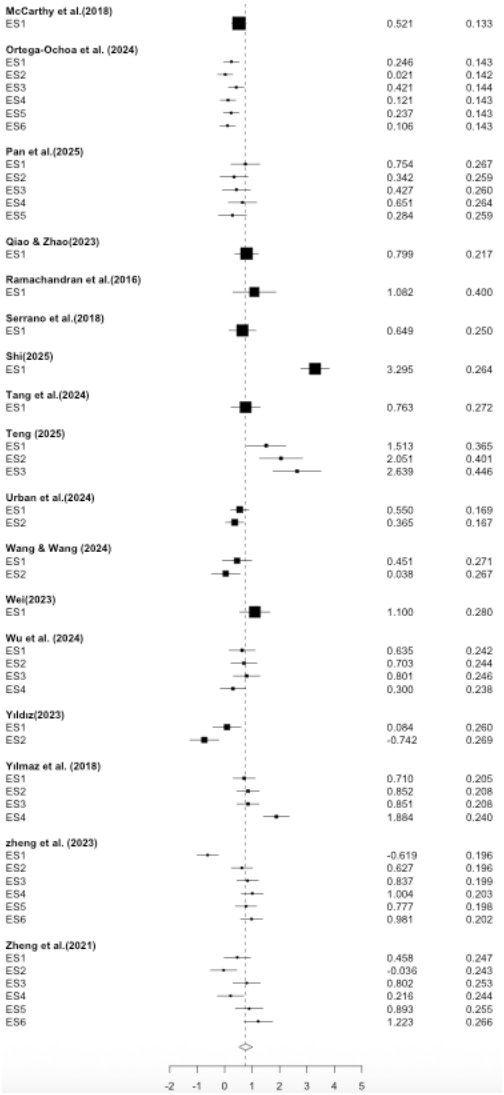
*Zheng, L., Long, M., Chen, B., & Fan, Y. (2023). Promoting knowledge elaboration, socially shared regulation, and group performance in collaborative learning: An automated assessment and feedback approach based on knowledge graphs. *International Journal of Educational Technology in Higher Education, 20*(1), 46. https://doi.org/10.1186/s41239-023-00415-4

Zheng, L., Li, X., Zhang, X., & Sun, W. (2019). The effects of group metacognitive scaffolding on group metacognitive behaviors, group performance, and cognitive load in computer-supported collaborative learning. *The Internet and Higher Education, 42*, 13–24. https://doi.org/10.1016/j.iheduc.2019.03.002

Zhu, Y., Liu, Q., & Zhao, L. (2025). Exploring the impact of generative artificial intelligence on students' learning outcomes: A meta-analysis. *Education and Information Technologies*, 1-29. https://doi.org/10.1007/s10639-025-13420-z

Appendix A. (Forest Plot of Studies 1-17 (Left) and Studies 18-34 (Right))





Appendix B. Search Strings for Different Databases

Databases	Search String
ERIC, Education Source, PsycINFO (EBSCOhost)	AB (“self regulat*” OR “metacog*” OR “learning regul*” OR “goal setting” OR “self evaluat*” OR “self assess*” OR “self reflect*” OR “rehearsal” OR “critical thinking” OR “elaboration” OR “effort regulat*” OR “time management” OR “peer learning” OR “help seeking” OR “environment structur*”) AND AB (“artificial intelligence” OR “AI” OR “large language model” OR “LLM” OR “machine learning” OR “deep learning” OR “expert system” OR “natural language processing” OR “NLP” OR “chatgpt” OR “intelligent agent” OR “intelligent assistant” OR “intelligent tutor” OR “intelligent tutoring system” OR “real time analytic*” OR “adaptive learning system*” OR “automatic feedback” OR “chatbot*” OR “conversational agent”) AND (“experiment*” OR “quasi experiment*” OR “intervention” OR “randomized controlled trial*” OR “randomised controlled trial*” OR "RCT" OR “control group” OR “pre test” OR “post test”)
Scopus	TITLE-ABS-KEY ("self regulat*" OR "metacog*" OR "learning regul*" OR "goal setting" OR "self evaluat*" OR "self assess*" OR "self reflect*" OR "rehearsal" OR "critical thinking" OR "elaboration" OR "effort regulat*" OR "time management" OR "peer learning" OR "help seeking" OR "environment structur*") AND TITLE-ABS-KEY ("artificial intelligence" OR "AI" OR "large language model" OR "LLM" OR "machine learning" OR "deep learning" OR "expert system" OR "natural language processing" OR "NLP" OR "chatgpt" OR "intelligent agent" OR "intelligent assistant" OR "intelligent tutor" OR "intelligent tutoring system" OR "real time analytic*" OR "adaptive learning system*" OR "automatic feedback" OR "chatbot*" OR "conversational agent") AND ALL ("experiment*" OR "quasi experiment*" OR "intervention" OR "randomized controlled trial*" OR "randomised controlled trial*" OR "RCT" OR "control group" OR "pre test" OR "post test")
Web of Science	((TS=(“self regulat*” OR “metacog*” OR “learning regul*” OR “goal setting” OR “self evaluat*” OR “self assess*” OR “self reflect*” OR “rehearsal” OR “critical thinking” OR “elaboration” OR “effort regulat*” OR “time management” OR “peer learning” OR “help seeking” OR “environment structur*”)) AND TS=(“artificial intelligence” OR “AI” OR “large language model” OR “LLM” OR “machine learning” OR “deep learning” OR “expert system” OR “natural language processing” OR “NLP” OR “chatgpt” OR “intelligent agent” OR

Databases	Search String
	“intelligent assistant” OR “intelligent tutor” OR “intelligent tutoring system” OR “real time analytic*” OR “adaptive learning system*” OR “automatic feedback” OR “chatbot*” OR “conversational agent”)) AND TS=(“experiment*” OR “quasi experiment*” OR “intervention” OR “randomized controlled trial*” OR “randomised controlled trial*” OR "RCT" OR “control group” OR “pre test” OR “post test”)
ProQuest Dissertations and Theses Global	noft(“self regulat*” OR “metacog*” OR “learning regul*” OR “goal setting” OR “self evaluat*” OR “self assess*” OR “self reflect*” OR “rehearsal” OR “critical thinking” OR “elaboration” OR “effort regulat*” OR “time management” OR “peer learning” OR “help seeking” OR “environment structur*”) AND noft(“artificial intelligence” OR “AI” OR “large language model” OR “LLM” OR “machine learning” OR “deep learning” OR “expert system” OR “natural language processing” OR “NLP” OR “chatgpt” OR “intelligent agent” OR “intelligent assistant” OR “intelligent tutor” OR “intelligent tutoring system” OR “real time analytic*” OR “adaptive learning system*” OR “automatic feedback” OR “chatbot*” OR “conversational agent”) AND (“experiment*” OR “quasi experiment*” OR “intervention” OR “randomized controlled trial*” OR “randomised controlled trial*” OR "RCT" OR “control group” OR “pre test” OR “post test”)

Note. AB = Abstract in EBSCOhost. TITLE-ABS-KEY = Title, Abstract, and Keywords in Scopus; ALL searches across Scopus’s searchable record fields and does not denote article full-text searching. TS = Topic in Web of Science covering titles, abstracts, and Indexing. NOFT = Anywhere except full text in ProQuest.

Appendix C. Codebook

Category	Variable	Coding options
Report characteristics	Publication year	Year of publication
Report characteristics	Report type	Journal articles, conference papers, dissertations, etc.
Research context	Discipline	Education; language; engineering; science
Research context	STEM affiliation	STEM; non-STEM
Research context	Educational level	Elementary; secondary; higher education
Research design	Type of experiment	Experimental; quasi-experimental
Research design	Intervention length	Less than 1 week; 1–5 weeks; 6–10 weeks; 11–16 weeks
Research design	Sample size	1–50; 51–100; 101–150; more than 150
Research design	Pretest inclusion	Yes; no
Research design	Type of control group	Passive; active
SRL	SRL phase	Forethought; performance; reflection; mixed
SRL	SRL measure	Online; offline
AI	AI application	Conversational agent; LLM-supported system; intelligent tutoring or adaptive learning system; learning analytic system
AI	Generative AI use	Yes; no
AI	Pedagogical role of AI	Feedback or assessment; monitoring or analytics; scaffolding or facilitation; mixed

Appendix D

Table D1

Research context

Comparison A − B	Difference in g	Robust SE	t	df	p
Educational level (34 studies; 91 effect sizes)					
Higher − elementary	0.0245	0.214	0.114	3.92*	.9145168
Elementary − secondary	0.4002	0.1657	2.415	6.35	.04996
Higher − secondary	0.4248	0.1763	2.410	5.75	.054360
Discipline (26 studies; 59 effect sizes)					
Engineering − education	−0.1639	0.441	−0.3713	1.99*	.746
Language − education	0.2159	0.522	0.4138	3.07*	.70623
Science − education	−0.0266	0.467	−0.0568	3.56*	.95772
Engineering − language	−0.380	0.278	−1.366	11.95	.19716
Science − language	−0.2424	0.317	−0.7638	17.29	.45530
Engineering − science	−0.1373	0.153	−0.8973	7.96	.39588
STEM affiliation (30 studies; 74 effect sizes)					
STEM − non-STEM	−0.161	0.242	−0.667	20.84	.512103

Note. Differences are A minus B; positive values indicate a larger estimated effect for category A. SE = standard error; df = small-sample-adjusted degrees of freedom. Duplicate reverse contrasts are omitted.

* df < 4: inference may be unreliable, as flagged in the R output.

Table D2

Research design

Comparison A − B	Difference in g	Robust SE	t	df	p
Study design (34 studies; 91 effect sizes)					
Quasi-experimental − experimental	−0.0161	0.251	−0.0641	30.1	.9493359
Intervention length (33 studies; 90 effect sizes)					
1–5 weeks − 11–16 weeks	−0.1560	0.251	−0.6211	5.89	.55780
6–10 weeks − 11–16 weeks	−0.2366	0.278	−0.8521	10.74	.41273
Less than 1 week − 11–16 weeks	−0.0172	0.326	−0.0528	14.47	.95858
6–10 weeks − 1–5 weeks	−0.0806	0.200	−0.403	6.51	.69956
Less than 1 week − 1–5 weeks	0.1388	0.262	0.5292	4.65	.62093
Less than 1 week − 6–10 weeks	0.2194	0.288	0.7625	9.18	.46493
Sample size (34 studies; 91 effect sizes)					
More than 150 − 101–150	−0.904	0.428	−2.115	8.75	.0645
1–50 − 101–150	−0.346	0.522	−0.663	11.45	.5205
51–100 − 101–150	−0.662	0.433	−1.528	9.52	.1589934
More than 150 − 1–50	−0.558	0.324	−1.722	9.56	.1171

Comparison A − B	Difference in g	Robust SE	t	df	p
51–100 − 1–50	−0.316	0.331	−0.953	13.49	.3571682
More than 150 − 51–100	−0.242	0.1417	−1.712	7.32	.1288452
Pretest inclusion (34 studies; 91 effect sizes)					
Yes − no	0.352	0.238	1.48	31.7	.1494463
Control group type (34 studies; 91 effect sizes)					
Passive − active	0.0777	0.240	0.324	31.7	.748432

Table D3

AI characteristics

Comparison A − B	Difference in g	Robust SE	t	df	p
Generative AI use (34 studies; 91 effect sizes)					
Yes − no	−0.292	0.223	−1.31	25.6	.2023740
AI application (34 studies; 91 effect sizes)					
ITS/AL − CA	0.321	0.406	0.791	15.01	.441295
LA − CA	−0.511	0.181	−2.825	10.79	.016815
LLM − CA	−0.323	0.242	−1.337	17.86	.197925
LA − ITS/AL	−0.832	0.387	−2.147	11.00	.0549
LLM − ITS/AL	−0.644	0.419	−1.536	14.91	.1455
LA − LLM	−0.188	0.209	−0.896	10.67	.3898
Pedagogical role (34 studies; 91 effect sizes)					
Mixed − feedback/assessment	−0.3962	0.533	−0.744	10.43	.4735
Monitoring/analytics − feedback/assessment	−0.790	0.515	−1.533	6.64	.1715
Scaffolding/facilitation − feedback/assessment	−0.3083	0.532	−0.579	9.88	.575497
Mixed − monitoring/analytics	0.3934	0.1827	2.153	5.51	.0789
Scaffolding/facilitation − monitoring/analytics	0.4812	0.1818	2.647	5.15	.044289
Mixed − scaffolding/facilitation	−0.0879	0.227	−0.387	21.29	.702648

Note. CA = conversational agent; ITS/AL = intelligent tutoring or adaptive learning system; LA = learning analytic system; LLM = LLM-supported system. Mixed denotes the mixed pedagogical-role category.

Table D4

SRL characteristics

Comparison A − B	Difference in g	Robust SE	t	df	p
SRL phase (28 studies; 75 effect sizes)					
Mixed − forethought	0.00654	0.388	0.0168	6.74	.98706
Performance − forethought	0.13723	0.350	0.3924	5.39	.709774
Reflection − forethought	0.29683	0.338	0.8781	6.59	.4108
Mixed − performance	−0.13069	0.273	−0.4785	12.88	.640306
Reflection − performance	0.160	0.296	0.538	6.39	.608519
Mixed − reflection	−0.29029	0.302	−0.9612	8.58	.36277
SRL measure (33 studies; 90 effect sizes)					
Online − offline	−0.230	0.211	−1.09	20.8	.2875361

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5742
https://doi.org/10.24059/olj.v30iSI2.5742

Online Learner Success Dimensions and Factors: A Framework Development Through Systematic Review

Florence Martin
North Carolina State University
Jered Borup
George Mason University

Abstract

In this systematic review we examined 188 articles from 2013 to 2024, to examine online learner success dimensions and factors. This review followed the DISCAR systematic review process and articles were searched in five databases, and screened using title, abstract, and full text by two researchers. Descriptive statistics, timeline, and qualitative themes were provided for learning context, journals published, research methodology, dimensions, and online learner success factors. Among the 188 articles, we analyzed publication patterns including year of publication, journals, context, research methodology, dimensions and factors related to online learner success. Focusing on four dimensions (learner, instructor, program and organization, and community) we identified a total of 30 success factors. The course and instructor related dimension was the most studied, followed by the learner dimension, and the program/organization dimension. There were 11 learner-related factors, 10 course and instructor-related factors, five program and organization related factors, and four community related factors. These factors will benefit researchers and online learning practitioners in designing and implementing online programs where learners are successful. Implications for online learners include that they actively manage their learning through readiness, self-regulation, engagement and help-seeking, while instructors and instructional designers create well-organized, accessible and engaging courses with timely communication and feedback. Administrators and communities can enhance online learner success by providing orientation, professional development, support services and providing a stable and encouraging home environment.

Keywords: online learner success, systematic review, online learning, framework development

Corresponding author: Dr. Florence Martin, fmartin3@ncsu.edu
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

Review of Literature

Online course offerings have increased over the last decade, and online learning has opened doors for students who need the flexibility to attend remotely due to full-time employment and other commitments. It has also extended the reach to students who live far from campus and those seeking specialized courses that are not commonly available locally. In many cases, online courses are less expensive than in-person courses because of lower fees and savings on commuting expenses. However, online learning is often perceived as being of lower quality, and students frequently face challenges with engagement and course completion. As a result, online courses have experienced failure rates higher than those found in in-person courses. While research has been conducted on the factors that lead to online learner success, a review of this research is needed to provide a more complete description of online learner success factors.

Bruno and Leidecker (1984) defined critical success factors as "characteristics, conditions or variables that, when properly sustained, maintained, or managed, can have a significant impact on the success of a firm competing in a particular industry" (p. 24). In the educational context, critical success factors "assist the online stakeholder groups to be guided in their operation in order to achieve the institution's vision" (Cheawijindakarn et al., 2013, p. 61). Arulkadacham et al. (2021) discussed how learner success may be interpreted differently from different perspectives, including those of the learners, instructors and organizations. The purpose of this systematic review is to synthesize the research published between 2013 and 2024 using a multidimensional lens including learner, course and instructor, program and organization and community to identify key dimensions and factors associated with online learner success.

Systematic Reviews on Success Factors

Online learner success is a frequent focus of research but often researchers fail to define *success.* One challenge to defining success is that it can be contextually dependent and based on both the explicit and implicit goals of the course (Arulkadacham et al., 2021). Like Martin et al. (2020), for the purposes of this research, we defined success broadly "to encompass achieving the desired outcomes of the course" (p. 811). Borup et al.'s (2020) review of the literature found that outcomes are frequently categorized in terms of knowledge, skills, and dispositions "(what students should *know*, be able to *do,* and what they are *becoming*)" (p. 811). While dispositional or "becoming" outcomes are often overlooked due to their subjective nature, they can be especially important. For instance, pre-service teachers taking a course on technology integration can learn the course topics and demonstrate their ability to use technology and integrate it into instruction but if they do not believe that technology integrated pedagogy can benefit student learning they are unlikely to actually apply their learning into their classrooms--the ultimate goal of the course (Ertmer et al., 2012).

Prior research has examined success factors in online, elearning and distance learning across various contexts (Barclay et al., 2018; Bhuasiri et al., 2012; Selim, 2007). Systematic reviews have investigated student success factors in online or elearning environments. Cheawjindakarn et al. (2013) reviewed 19 studies published between 2000-2012 and identified the following five categories of factors: institutional management, learning environment, instructional design, services support, and

course evaluation. Asalla et al. (2017) reviewed 54 journal articles published between 2006 and 2016 to examine critical success factors for elearning in higher education and identified 17 key critical success factors. They found that information technology (IT) infrastructure or the learning management system (LMS) was the most critical success factor, followed by e-learning tools and e-learning models.

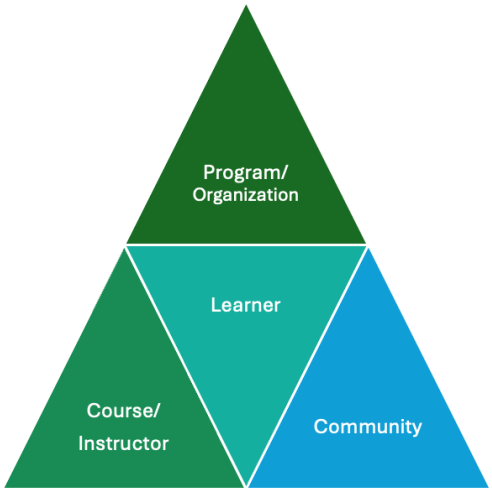
Arulkadacham et al. (2021) conducted a systematic review from 11 studies published between 2009 and 2021 and focus group study to identify predictors of student success in online courses. They identified 29 unique predictors of student success in eight categories: engagement, student characteristics, attitude towards learning and technology, academic history, psychological attributes, course structure, instructional design, and social presence. Muljana and Luo (2019) examined factors influencing student retention in online learning through a systematic review of 40 studies published between 2010 and 2018. They identified institutional support, program difficulty, promotion of a sense of belonging, facilitation of learning, course design, student behavioral characteristics, and demographic variables along with other personal variables as critical factors contributing to retention. Although a few systematic reviews exist, there is a need to examine online learner success from a multidimensional perspective to more comprehensively identify key success factors.

Dimensions of Online Learning

Building on the framework of online learning research themes proposed by Martin et al. (2020), which includes, learner, course and instructor, and organization dimensions, in this framework we extended the model by adding program and community as additional dimensions of online learning. Figure 1 presents a visual representation of the online learner success dimensions.

Figure 1

Online Learner Success Dimensions



Learner

This dimension focuses on the online learner who engages in learning within a digital setting which includes hybrid, blended, online education which is a subaspect of what Zawacki-Richter and

Jung (2023) refers to as open, distance and digital education (ODDE). In the review by Martin et al. (2020), the learner dimension encompassed learner characteristics, learner outcomes, and learner engagement. Learner characteristics are defined as "various measures of learners' psychological, behavioral nature, and attitudes toward everything related to learning" (Nakayama & Santiago, 2012, p. 1745). Specific to the online learning context, Martin et al. (2020) categorized online learner characteristics to demographic, academic, cognitive, affective, self-regulation and motivational characteristics. For instance, the same flexibility that has drawn many learners to enroll in online courses, also requires learners to exercise high levels of self-regulation and independent learning skills and habits. In fact, Roblyer et al. (2007) argued that online learners' ability to be successful in online courses "appears to depend more on motivation, self-direction, or the ability to take responsibility for individual learning" than their ability to learn the course material (p. 11).

Learner outcomes are described as clearly defined statements describing the outcomes learners should achieve after completing a course or program (Martin et al., 2020). Common learner outcomes include dropout, course completion, retention, and learner success. Online learner engagement is defined as "the productive cognitive, affective, and behavioral energy that a learner exerts interacting with others and learning materials and/or through learning activities and experiences in online learning environments (Martin & Borup, 2022). Researchers have examined learner engagement in online contexts through interaction, participation, community, collaboration, communication, involvement and presence.

Course and Instructor

This dimension focuses on both the course offered online and the instructor who teaches it. In Martin et al.'s (2020) review, the course and instructor dimension focused on instructor characteristics, course design and development, course facilitation, course assessment, course evaluation and quality assurance, and course technologies. Instructor characteristics address instructors' experience, satisfaction and roles in online courses. Course design and development refer to the processes used to develop the online courses and meet online learners' needs. Course facilitation emphasizes the effective delivery of the online course, while course assessment examines how learners are assessed in the online course. Course evaluation and quality assurance focus on assessing the overall effectiveness of the course, and course technologies highlight the various applications and tools that support online learning. Faza and Lestari's (2025) review of 121 articles found that technologies such as generative AI and collaborative platforms could improve performance, motivation, and engagement. Instructors should also adapt their support based on learner needs and abilities (Rose et al., 2015).

Program and Organization

The program and organization dimension focuses on the administrative and support structures provided by the program and/or organization that extend beyond and across courses. Learners who are new to online learning "need to learn how to learn online" (Lowes & Lin, 2015, p. 18). While course instructors can help, teaching both the course content and effective online learning strategies can prove overwhelming (Borup & Stimson, 2019). As a result, online programs have long provided orientations for online learners that set clear expectations and provide strategies for successfully meeting those standards (Scagnoli, 2001). Research has shown that such orientations can improve

online learners' readiness and confidence (Liu, 2019; Moon, 2025). Programs and organizations can also provide support to learners throughout their time in program and across courses. For instance, programs and organizations can provide centers where online learners can turn for support (Lehan et al., 2018; Lehan et al., 2020). Learners can also be provided with a mentor or success coach who proactively develops relationships with learners, provides emotional and psychological support, helps them to develop online learning skills, and advocates for their needs (Andersen & West, 2021). Similarly, institutions can facilitate peer mentoring programs (Goodrich, 2021).

The program and organization can also indirectly impact online student success by helping to ensure that courses are designed and delivered in a consistent manner—responsibilities often contracted to online program managers with varying degrees of success (Springer, 2018; Sun & Turner, 2022). They can also help to ensure that instructors and coaches are prepared for the challenges of learning online (Leary et al., 2020).

Community

The community dimension focuses on the personal and professional environments in which students live, work, and learn that can be determining factors in both their decision to enroll and success in online courses. These are the communities in which the learner belongs to independent of their enrollment in their online coursework. Actors within these communities commonly have long—even life-long—relationships with the learner (Borup et al., 2020). As a result, while they may not have an understanding of learning online, they have a more holistic understanding of the learner's needs. Parents play a particularly important role in supporting their children' s online learning (Oviatt et al., 2018). Even in higher education parental and family support and resources can be a determining factor in learners' ability to engage online (Brubacher & Silinda, 2021; Gao et al., 2021). In fact, emotional and tangible family support can positively impact middle aged (aged 50-64) and older adults' (over the age of 65) e-learning efforts (Chu, 2010). Alternatively, Samra et al. (2021) explained that family demands can negatively impact performance if online learners fail to "create boundaries with family members and set clear expectations of what they need," adding that this may require "negotiating and renegotiating support at times of greater need" (p. 95). Online learners are also more likely to complete online coursework when they receive financial support from employers (Lenio, 2021).

Purpose of the Study and Research Questions

Several factors influence the success of online learning relating to various dimensions such as learner, course, community, program/organization. Although a few reviews have been conducted in the past, there remains a need for a recent comprehensive review that focuses on learner success in online learning contexts. The purpose of the systematic review is to examine the literature on online learner success from 2013 to 2024 from a multidimensional perspective, focusing on learner, course and instructor, program and organization, and community dimensions to identify key factors associated with online learner success. The systematic review addressed the following research questions.

1. What are the publication patterns of studies examining online learner success (publication trend, journals, research methods, learning context)?

2. What are the dimensions and factors contributing to the success of online learners as identified in the existing research?

Methods

This review followed the DISCAR systematic review process (Martin, 2025), which involves (a) designing the review, (b) setting inclusion and exclusion criteria, (c) selecting data sources, search strategies, and screening criteria, (d) extracting data from primary studies to address research questions, and (e) analyzing and synthesizing data to report findings. In addition, we used the Preferred Reporting Items for Systematic Reviews and Methods (PRISMA) guidelines to conduct the systematic review on online learning success factors and dimensions.

Inclusion/exclusion criteria

Based on the focus of the review, we established the inclusion and exclusion criteria outlined in Table 1. These criteria were applied during both the search and screening phases. Criteria, such as publication date, type, and language, were set during the search process, whereas others were used during screening. Articles failing to meet any inclusion criteria were excluded. To ensure study quality, we included only peer-reviewed articles and required empirical articles to have a clearly defined "methods" section.

Table 1

Inclusion/Exclusion Criteria

Criteria	Inclusion	Exclusion
Publication date	From Jan 2013 until May 2024	Before 2013 or Published after May 2024
Publication type	Scholarly articles of original research from peer-reviewed journals	Book chapters, technical reports, dissertations, or proceedings
Article focus	Articles focused primarily on online learner success	Articles that did not have online learner success as the primary focus
Research method and results	An identifiable method and results section describing how the empirical study was conducted and the findings. Quantitative, qualitative and mixed-methods were included.	Reviews of other articles, opinion, discussion papers or conceptual articles
Language	Journal articles written in English	Other languages than English were not included

Data Sources and Search Strategies

Five databases including ProQuest Education Database, Academic Search Complete, APA PsychInfo, Education Research Complete and ERIC were used in this review. Searches were conducted in two phases: the ProQuest yielded 1224 results, and the remaining four databases yielded 403 results, for a combined total of 1627 articles.

The following search string was used to conduct the search,

(abstract("online learner success") OR title("online learner success") OR abstract("online student success") OR title("online student success"))

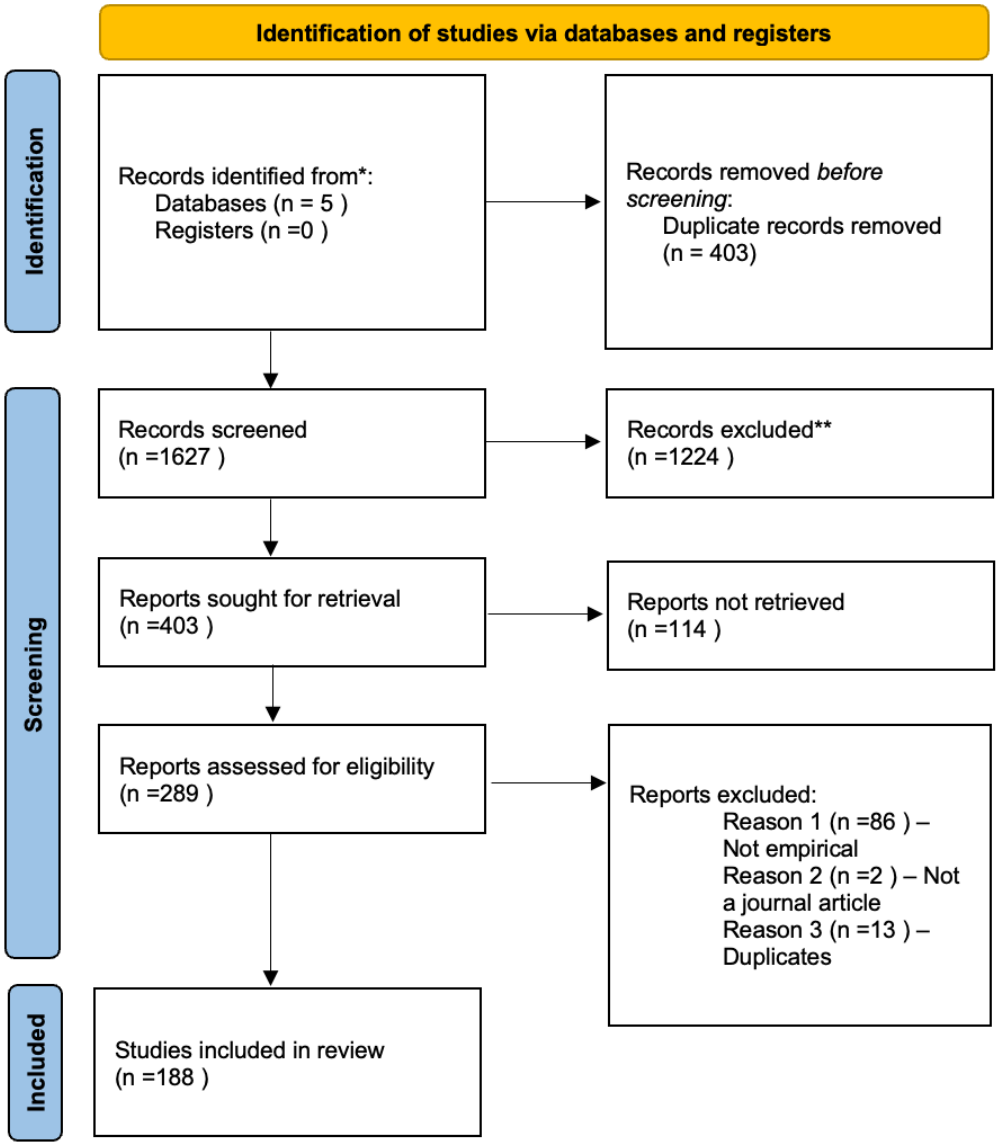
Screening Process

The review followed the PRISMA guidelines proposed by the Ottawa Methods Center for reporting items for systematic reviews and meta-analyses (Moher et al., 2009; Page et al., 2021). The PRISMA flow diagram (Figure 1) illustrates the process flow, and the number of articles reviewed at each stage of the review process, including identification, screening, eligibility, and inclusion.

After removing 403 duplicate records that were identified by two researchers by the title, the initial search resulted in 1627 articles, which records were exported as RIS file into Zotero. After title screening, 403 studies were identified as focusing on online learner success, but we were unable to obtain 114 records and 289 were screened. During abstract screening, additional studies were excluded because they were not empirical (86 studies), were not journal articles (2 studies) or were duplicates (13 studies). This resulted in 188 studies being included for full text review. During screening, articles were put into three categories, yes, maybe and no by two researchers who met periodically to discuss the screening results. The "maybe" articles were reviewed by both researchers before a decision was made.

Figure 2

Process flow using PRISMA flow diagram



Data Extraction

The research team consisted of two researchers who had expertise in educational technology, online learning, and systematic review research methodology. Both researchers were involved in screening and iterative coding using both inductive and deductive methods. Google sheets was used for both screening and coding. Coding was conducted in Google sheets by creating columns for year, journal, dimensions, factors, context, and research design. Year and journal publication data was extracted from Zotero into the codebook.

Codebook is included in Appendix A. The researchers met periodically during the process to discuss questions and resolve any discrepancies. Coding occurred in two cycles. In the first round of coding, data was extracted for context, methodology, dimension and success factors by both researchers. Success factors were inductively coded. In the second round of coding, the success factors were reviewed once again by both researchers and were grouped into 30 categories. Disagreements were refined through discussion during the synchronous meetings.

Data Analysis

Descriptive statistics, including frequencies and percentages were generated for learning context, journals published, research methodology, dimensions, and online learner success factors. In addition, a timeline chart was created to illustrate publication trend by years.

Results

In this section, we first share publication patterns. We then summarize the success factors for each dimension. This systematic review presents the findings of online learner success dimensions and factors from 188 studies. Across the four dimensions, a total of 30 success factors were identified. There were 11 learner-related factors, 10 course and instructor-related factors, five program and organization related factors, and four community related factors.

Publication Patterns

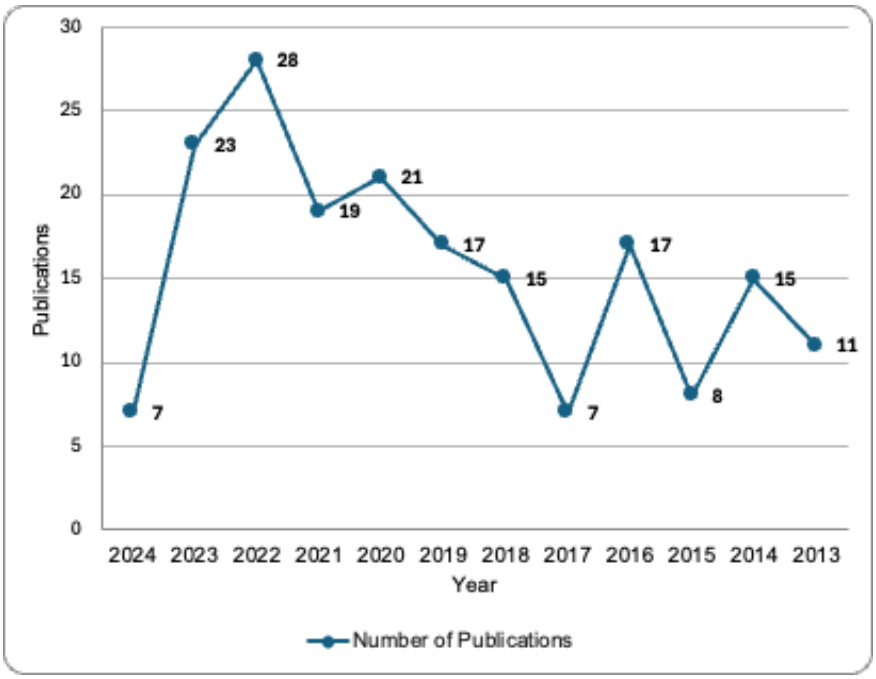
We examined the publication trends, journals of publication, research design and learning context as part of the publication patterns for studies on online learner success examined in this review.

Publication Trends

The year 2022 had the highest number with 28 publications, followed by 2023 with 23 publications. Data for 2024 did not represent a full year because the search was conducted in May. Overall, the trend shows a steady increase in research on online learner success, with brief declines in 2015 and 2017. The years 2023 and 2024 show particularly high levels of research activity, which may be attributed to the increased reliance on online learning during and following the COVID-19 pandemic.

Figure 3

Trends of Publications by Year



Journals

The top five journals that published highest number of studies on online learner success in this review were *Online Learning Journal* (n=13), *International Review of Research on Online and Distributed Learning* (n=8), *Online Journal of Distance Education Administration* (n=7), *Journal of Educators Online* (n=7), and *American Journal of Distance Education* (n=5).

Research Design

Of the studies reviewed, 135 were quantitative (71.81%), 29 were qualitative (15.43%) and 24 employed mixed or multimethod designs (12.77%).

Learning Context

Of the 188 reviewed articles, 160 studies (85.11%) were conducted in the higher education context, 19 studies (10.10%) were conducted in K12 settings, and eight studies (4.26%) in continuing education. One study was conducted across both higher education and K12 settings.

Success Factors

Table 2 summarizes the frequencies of studies across each dimension. The learner and course and instructor were the most studied dimensions, followed by the program/organization when examining online learner success. There were very few studies on the community dimension.

Table 2

Frequency and Percentage of Online Learner Success Studies based on Dimensions

Dimension	Frequency	Percentage
Learner	76	40.43%
Course and Instructor	76	40.43%
Program/Organization	30	15.96%
Community	6	3.19%

Learner

Within the learner dimension, multiple success factors have been examined in research. These include learner readiness, personal and environmental factors, cognitive factors, self-regulation, time management, time on task and learner effort, motivational, self-efficacy, social and emotional factors, and technology skills and use.

Learner Readiness

Learner readiness are essential learner attributes and conditions that enable them to be successful in online courses. Readiness could be assessed through a learner needs or readiness assessment to better understand the learners' needs and requirements (Latheef, 2021; Tainsh, 2023). Students' general lack of readiness was laid bare during the Covid-19 pandemic when students had to rapidly pivot to remote learning and highlighted the importance of which can inform targeted advisement strategies and course orientations. For instance, using insights from corporate onboarding, Latheef (2021) conducted a mixed methods study that found a combination of an online learner readiness self-assessment and targeted preview materials help to set realistic expectations and primed students for success.

Personal and Environmental Factors

Understanding learners' background characteristics is important for promoting learner success. Environmental characteristics can be identified by requesting the students to complete a biographical questionnaire before or at the beginning of the course. These may include items such as location, life pressures, and how they juggle online learning with daily life, work commitments, and study environment (Du et al., 2022; Goad et al., 2021). For instance, Goad et al. (2021) research that included 821 online learners in higher education found a relationship between course completion and hours worked outside of school and organization and study environment.

Cognitive Factors

Cognitive factors reflect learners' intellectual abilities and academic preparedness that influence their capacity to succeed in the online course. These include factors such as learner's self-reported GPA, academic expectations, stage of cognitive development based on age and class standing as well as skills related to memory, comprehension, learning strategies, reading rate and recall (Guidry, 2013; Liu, 2023).

Self-Regulation

Self-regulation refers to learners' ability to manage their learning process by setting goals, monitoring progress to achieve academic success. This factor has been widely examined in online learning research. Studies have focused on goal setting, self-monitoring using performance progress tools, grit, self-directed learning, all of which help learners manage distractions, persist, and remain organized (Feldman-Maggor et al., 2022; Price et al., 2021). Primary and secondary students tend to have low self-regulation abilities and in higher education there is a self-regulation skill gap between undergraduate and graduate students (Tseng et al., 2019).

Time Management

Time management skills are critical for online learners to be successful as it influences how students plan and allocate time to complete the various tasks in the online course. This includes the use of time-management tools, and these skills influence student study habits and overall academic performance (Baker et al., 2019; Feldman-Maggor et al., 2022; McLeod, 2019). Baker et al.'s (2019) found that these skills can be developed when students are provided opportunities to practice them, leading to higher performance, but that these skills can also quickly atrophy if students stop their efforts to develop and use time management skills.

Engagement Level

Learner engagement and participation levels are important indicators of learners' current performance and course success. Engagement can be measured through indicators such as students' interaction frequency, time on task, login patterns, duration of participation, interaction with course materials, time spent on assignments and early detection of students' academic performance (Chen, 2023; Xu, 2023). Research has also worked to identify minimal engagement thresholds for success. For instance, Pazzagilia et al. (2016) large study at the high school level found that students who spent at least 1.5 hours per week on task largely passed the course and students who engaged for two or more hours per week had significantly better course outcomes than those students with less engagement.

Motivational Factors

Student motivation is a critical determinant for online learner success as it influences students' willingness to invest effort in the online course activities. Studies have examined forms of motivation such as autonomous motivation, motivational beliefs, and academic motivation (Basila, 2014; Newhouse & Cerniak, 2016).

Learner Self-Efficacy

Self-efficacy refers to a learner's belief in their ability to complete the various tasks required to succeed in online courses. In online learning, self-efficacy plays an important part as learners must navigate learning management systems, manage learning dependently. Research has examined various forms of self-efficacy, including computer self-efficacy and digital self-efficacy as essential to succeed in the online course (Bahçekapili & Karaman, 2020; Tseng et al., 2019). Successful experiences on assignments and in courses can help to develop learners' self-efficacy (Odabas & Kahramanoglu, 2023). Azizi et al. (2022) research added that learners' development of computer self-efficacy can alleviate their e-learning anxiety.

Emotional Factors

Learner's emotional well-being can play a significant role in shaping learners' experiences and outcomes in online environments. This includes factors such as anxiety, coping strategies, and aspects of emotional health (Breneiser et al., 2018; Idrizi et al., 2021). Online learning may intensify emotional challenges due to factors such as isolation, self-directed learning and technology-related stress. However, there can also be some emotional benefits for learning online. For instance, compared to learning in in-person environments. Idrizi et al. (2021) found an inverse relationship between performance and neuroticism, or students' degree of "instability, impulse control, and anxiety" (p. 209). However, their research also suggested that learning online may help to reduce social pressure and anxiety experienced by students with high neuroticism. This may help to explain why some, especially those in primary schools, have long turned to online learning to avoid bullying and other negative social situations (Beck et al., 2023; Beran & Li, 2007).

Social Factors

Social factors play an important role in supporting learning success in online courses by fostering a sense of belonging and connectedness. Social belonging, social skills, sense of connection with peers, need for affiliation have also been found to support learners to be successful in online courses (Seiver & Troja, 2014).

Technology Factors

Technology factors are critical contributors to online learner success and it includes both learners' access and proficiency with digital devices. Technology related factors include students' technological skills, intention to use technology (Kaur & Gopal, 2022), use of handheld devices to complete homework, digital access (Gravelle et al., 2024), and typing speed/accuracy (Geiger et al., 2014), all of which contribute to online learner success. These factors have long been recognized and Hillman et al. (1994) explained that success in online courses is highly dependent on learners' ability to effectively navigate and interact with technology.

Course and Instructor

Within the course and instructor dimension, several success factors related to course attributes of and instructional activities within the online course have been examined in prior research. These include course design and organization, course materials, instructional approaches, course activities and assessment, accessibility factors, cultural considerations, instructor characteristics, instructor presence and communication, course technology, and course support.

Course Design and Organization

Course design and organization play a central role in supporting online learner success by providing a coherent and purposeful learning structure. This factor includes thoughtful course design, alignment of learning outcomes, sequencing of learning objectives, alignment of instructional resources, and effective use of navigation tools (Aydogdu & Güyer, 2019; Mshayisa & Ivala, 2022; Ren, 2023).

Course Material

Course materials are a critical component of online learner success, as they serve as the primary means through which content is delivered in online environments. This includes teaching materials, online resources, instructional material quality, instructor created videos which have shown

to support learner engagement and understanding (McGuinness & Vlachopoulos, 2019; Sari et al., 2022).

Instructional Approaches

Instructional approaches are the various strategies that are included in the online course and plays a significant role in shaping learner engagement and achievement. Effective instructional approaches include authentic learning, project-based learning, gamification, adoption of pedagogical roles, active learning, concise lectures, just-in time teaching (Bailey et al., 2017; Zheng et al., 2020).

Course Activities and Assessment

Course activities and assessments are central to facilitating learner participation and achievement. These include varied activities and assessments, such as self-assessments, well-structured asynchronous discussions, online simulations, collaborative projects, quizzing and pre-exam quizzes, as well as formative, timely, and personalized feedback (Bognár et al., 2021; Tainsh, 2023). Jacobi's (2017) qualitative research found that well-designed collaborative activities can make students' learning feel more relevant. Zheng et al. (2020) hierarchical linear modeling and content analysis using student-, instructor-, and course-level data in online high-school English language and literacy courses found that project-based learning and high-level knowledge activities were beneficial to learning outcomes. However, these types of activities actually resulted in lower performance for online credit recovery students. As a result, when designing course activities and assessments, student preferences and abilities should be considered.

Accessibility Factors

Accessibility factors are critical to ensure that all learners including those with disabilities can fully participate and succeed in the online course. Learners with disabilities have different needs when taking online courses. Providing appropriate accommodations and accessible course design is essential for supporting their success (Betts et al., 2013; Blieck et al., 2019). However, Chatman et al. (2019) stressed that accessibility needs to be matched with best practices if it is to lead to online learner success.

Cultural Considerations

This factor involves recognizing cultural variables and providing culturally responsive and adaptive instruction to support diverse learners (Gómez-Rey et al., 2016; Newberry & DeLuca, 2014). By addressing cultural differences, instructors can create online courses that are effective for all students.

Instructor Characteristics

Instructor characteristics are key determinants of learner success in online courses and shape the quality of instruction and overall course effectiveness. This includes factors such as faculty expertise, faculty professional training, technology proficiency, instructional commitment, flexibility, rapport, and instructor longevity (Glazier, 2016). Welch et al. (2015) stressed the importance of teacher dispositions and beliefs, stressing that "dispositions manifest into actions and behaviours in the classroom affecting not only their performance as an educator, but also the performance of their students" (p. 7).

Instructor Presence and Communication

Instructor presence and communication are critical factors in supporting learner engagement. This includes instructor involvement, time spent interacting with students, social interactions, teaching presence, emotional presence, online social comfort, student-instructor interactions, sense of community, and overall communication practices (Rosser-Majors et al., 2022; Van Wart, 2020). Developing presence is a skill that can be developed by instructors. For instance, Rosser-Majors et al. (2022) designed and provided seven self-paced modules on establishing instructor presence that proved helpful in instructors' efforts to engage students and increase student success.

Course Technology

Course technology is the backbone for online teaching and learning. This encompasses technology infrastructure, interactive platforms, asynchronous discussion tools, synchronous meeting systems for live group meetings using audio and video, and secure examination systems (Alanazi et al., 2020).

Course Support

Course support encompasses a broad range of resources to help online learners successfully achieve learning outcomes. Support mechanisms include teacher support, teaching assistant support, technical support, peer support, individualized support, virtual online learning communities, aligned instructional support, instructional incentives and policies, program and legal support (Linden et al., 2022).

Program and Organization

For the Program and Organization dimension, multiple factors have been studied in research. This includes student orientation, faculty professional development, student support services, institutional sense of community and belonging and tutoring.

Student Orientation

Learning online can prove challenging in unique ways. As a result, orientation sessions and support modules can help students to develop technical skills, writing skills, and research abilities (Townley et al., 2023). Programs and institutions can also help students to develop motivational and self-regulation abilities.

Faculty Professional Development

Successful online teaching requires different competencies and skills than in-person instruction (Leasure et al., 2020). As a result, professional development provided to online instructors can positively impact their facilitation efforts, and in turn student success (Scarpena et al., 2018) and satisfaction (Kane et al., 2015). This professional development can be offered as courses or faculty-to-faculty mentoring (Zweig, 2016; Drouin et al., 2015).

Student Support Services

Institutions can provide students with both academic and non-academic support structures. Research has highlighted the importance of administrative and academic support including library services to students' academic success (Skelcher et al., 2020). Alkaabi et al. (2022) and Skelcher et al.

(2020) also highlighted the importance of aid and services such as mental health support that can have an impact on student retention.

Institutional Sense of Community and Belonging

Just as students can develop a sense of community within an individual course, students can develop a sense of community to their institution (Skelcher et al., 2020). This is often developed through students' interactions with others at the institutional level such as advisors and service providers who help to nurture a culture of caring, encouragement, and support (Stone & O'Shea, 2019).

Tutoring

Institutions and programs can provide learners with tutoring that extends the support that an online instructor can provide (Devers et al., 2022; Maré & Mutezo, 2021). This tutoring can assist students with general skills such as academic writing or more course specific content support. Tutors can also be more cost effective and available than course instructors (Linden et al., 2022).

Community

For the Community Dimension, research focused on two communities--professional and personal. While these communities do not form a part of the online course or program, they can act as both power sources that fuel persistence and success in online learning. Inversely, they can also act as loads that produce stress and demands that can be determining factors in their failure to achieve their academic goals (González & Blackford, 2022). González and Blackford (2022) warned that "excessive 'loads' or demands' that are not coupled with the required 'resources' or 'power' will hinder engagement and consequently performance (learning/grade)" (p. 3). Despite their importance, relatively few studies mentioned the importance of the environment and community outside of the course and program.

Inter-role Harmony

In addition to their responsibilities as a student, online learners fulfill multiple outside roles in their family, social groups, and employment. As online learners engage in learning activities, they may experience conflict and tension in their other roles that require them to rebalance the time and effort they put into their studies with their other responsibilities. High inter-role conflict can even cause the online learner to step away from their studies altogether. However, those students with autonomy and control in the time, place, pace, and path of their role fulfillment in and beyond the course, are more likely to achieve inter-role harmony because they can better act to balance their learning, professional, and personal responsibilities as their demands in each area ebbs and flows (González & Blackford, 2022).

Emotional Support and Encouragement

Online learners are separated by space and often time from others in their course. This can leave them feeling isolated with low affective engagement. Actors with an online learner's personal and professional communities often share the same physical space and can provide important affective support such as encouragement, understanding, and connection (Curtis & Werth, 2015; Lee et al., 2013).

Positive and Stable Home Environment

While learning online, students require an environment that is conducive to learning. For instance, more than 10% of student participants in Gravelle et al. (2024) shared that learning online proved more challenging due to a noisy home. One student shared, "I can't find a quiet place to sit and learn" (p. 455).

Assistance with Family Responsibilities

Many parents with children in their care, turn to online learning for the flexibility it provides. However, they often require assistance with family responsibilities and childcare so that they can focus on their studies. For instance, one woman in Stone and O'Shea (2019) expressed gratitude for "a husband who helps out with dinner on assignment due nights, so [she] can keep studying" (p. 104).

Online Learner Success Framework

The dimensions and factors identified from the systematic review of research is included as a visual in Figure 4.

Figure 4

Online Learner Success Framework

Learner	**Course and Instructor**
• learner readiness • personal and environmental factors • cognitive factors • self-regulation • time management • engagement level • motivational factors • learner self-efficacy • emotional factors • social factors • technology factors	• course design and organization • course material • instructional approaches • course activities and assessment • accessibility factors • cultural considerations • instructor characteristics • instructor presence and communication • course technology • course support
Program and Organization	**Community**
• student orientation • faculty professional development • student support services • institutional sense of community and belonging • tutoring	• inter-role harmony • emotional support and encouragement • positive and stable home environment • assistance with family responsibilities

Discussion

Success Dimensions and Factors

The course and instructor dimension, and the learner dimension was examined in 40.43% of studies, and the program/organization dimension was examined in 15.96% of studies when studying online learner success. Each of these three dimensions is critical to online learner success. The community dimension, in contrast, was only identified in 3.19% of the articles. This disparity does not suggest that community is less important, rather, community-related factors are often less observable and more difficult to study, especially in higher education. Despite these challenges, there is a clear need for additional research examining community influences through the lens of online learner success. Across the four dimensions, this review identified 11 learner-related factors, 10 course and instructor-related factors, five program and organization related factors, and four community related factors, resulting in a total of 30 success factors that contribute to online learner success.

Comparison with Prior Systematic Reviews

Cheawjindakarn et al. (2013) reviewed 19 studies and identified five factors, institutional management, learning environment, instructional design, services support and course evaluation. These factors largely correspond to course or program level dimensions which were also examined in this review. Arulkadacham et al.'s (2021) review identified eight categories of factors, including engagement, student characteristics, attitude towards learning and technology, academic history, psychological attributes, course structure, instructional design and social presence. These categories reflect a combination of learner and course-related dimensions. Mujjana and Luo (2019) in their review identified institutional support, the level of difficulty of the programs, promotion of a sense of belonging, facilitation of learning, course design, student behavioral characteristics, and demographic variables along with other personal variables as critical factors. These factors emphasized learner, course, and program level dimensions. While these prior reviews highlighted important aspects of online learner success, the present review identified a broader and more comprehensive set of factors within each dimension, offering a more multidimensional perspective on online learner success.

Differences Between Higher Education Learner Success and K-12 Learner Success

Skills such as academic self-efficacy, time management and organizational skills, and metacognitive and self-regulation skills (Abernathy, 2020; Basila, 2014; Lee et al., 2013; Leow et al., 2024)—skills that are often lacking in younger learners. As a result, younger learners tend to require more support than older and more developed learners. For instance, parents or other guardians also have care responsibilities over their children and adolescents and play a critical role in enrolling both in students' enrollment in online courses and their success in those courses (Krishnan, 2023; Sallaffie et al., 2021). However, often parents lack the time, ability, or needed role construction to provide students with adequate support and/or supervision (Hoover-Dempsey et al., 2005). As a result, at the K-12 level, learners taking one or two online courses are increasingly completing those courses at their local brick-and-mortar school in a classroom with an on-site facilitator or mentor who supervises and supports their learning (Borup, 2018). As learners mature into and through adulthood, they no longer require supervision and their learning becomes more independent. Parents can still play an important role at the university level, they have fewer responsibilities than parents of children or adolescents and other support actors such as friends, spouses and partners, and employers become important support sources. The outside demands on their time can also change. For instance, Stone and O'Shea (2019) examined mothers with caring responsibilities and found that they require flexible

and relevant learning responsibilities. Christenson et al. (2012) summarized that "both the individual and the context matter" (p. vi). Those designing, facilitating, and supporting online learning need to carefully consider learners' abilities and needs and provide them with the adequate support they require to be successful.

Limitations

This review has some methodological limitations. First, only five databases that the authors had access to were used in this review, which may have resulted in the omission of some studies. Second, all the articles that were reviewed were published in English, excluding potentially valuable research reported in other languages. Third, the selection of keywords used may have limited the retrieval of additional relevant studies. Adding synonyms of the keywords included could have extended the search and resulted in inclusion of additional relevant studies. Fourth, only two reviewers participated in the coding process. Although the reviewers collaborated closely and met biweekly to discuss the coding decisions, there could be coder bias involved. Finally, a separate quality assessment was not conducted in this review.

Implications and Future Directions

A key contribution of this systematic review is the development of a multidimensional framework for understanding online learner success. This framework identified critical factors across the learner, course and instructor, program and organization, and community dimensions. It benefits online students, instructors, instructional designers and administrators in making sure these factors are included in online courses that are being developed. The framework serves as a holistic lens for students to succeed beyond just academic performance. Online students need to recognize the importance of being ready to participate in online learning, managing their time, staying motivated and engaging with both the course and their peers and instructors. By developing self-regulation, leveraging technology use and seeking social and emotional support when needed, students can take ownership of their learning and be successful in online courses. Online instructors and instructional designers have to be intentional in the design and facilitation of courses, prioritize clear organization, accessible and culturally relevant materials, well-aligned activities and assessments, while maintaining consistent presence through timely communication and feedback. Administrators who lead programs and organizations should invest in effective student orientation, ongoing faculty professional development, intentionally fostering a sense of community and offering various support services including tutoring. For the community, implications include encouraging the learner and providing emotional support, offering a conducive home environment, and assisting with family responsibilities. Implications for policy include the need for institutions to establish clear guidelines for quality online courses.

The proposed framework with the dimensions and associated factors provides a foundation for future research and can be used by scholars to further examine, refine and validate determinants of online learner success in diverse contexts.

References

Articles denoted with (*) are included in the review.

*Abernathy, D. (2020). The impact of introducing self-regulated learning strategies in online graduate studies. *Journal for the Advancement of Educational Research International*, *14*(1), 7-20.

*Alanazi, A. A., Frey, B. B., Niileksela, C., Lee, S. W., Nong, A., & Alharbi, F. (2020). The role of task value and technology satisfaction in student performance in graduate-level online courses. *TechTrends*, *64*(6), 922-930. https://doi.org/10.1007/s11528-020-00501-8

*Alkaabi, N., Zaki, N., Ismail, H., & Khan, M. (2022). Detecting Emotions behind the Screen. *AI*, *3*(4), 948-960. https://doi.org/10.3390/ai3040056

Andersen, C. L., & West, R. E. (2021). "We overwhelm them with hope": How online mentors can support online learners. *Online Learning*, *25*(4), 388-415. https://doi.org/10.24059/olj.v25i4.2440

Arulkadacham, L., McKenzie, S., Aziz, Z., Chung, J., Dyer, K., Holt, C., ... & Mundy, M. (2021). General and unique predictors of student success in online courses: A systematic review and focus group. *Journal of University Teaching & Learning Practice*, *18*(8), 07. https://doi.org/10.53761/1.18.8.6

Asalla, L. K., Putri, M. R., & Pradipto, Y. D. (2017, November). The critical success factor of E-learning in higher education: A systematic literature review. In *2017 International Conference on Information Management and Technology (ICIMTech)* (pp. 192-197). IEEE. https://doi.org/10.1109/icimtech.2017.8273536

*Aydogdu, S., & Güyer, T. (2019). The Effect of Digital Concept Maps in Online Learning Environments on Students' Success and Disorientation. *Malaysian Online Journal of Educational Technology*, *7*(1), 76-93. https://doi.org/10.17220/mojet.2019.01.006

*Azizi, Z., Rezai, A., Namaziandost, E., & Ahmad Tilwani, S. (2022). The Role of Computer Self-Efficacy in High School Students' E-Learning Anxiety: A Mixed-Methods Study. *Contemporary Educational Technology, 14*(2), Article ep356. https://doi.org/10.30935/cedtech/11570

*Bahçekapili, E., & Karaman, S. (2020). A path analysis of five-factor personality traits, self-efficacy, academic locus of control and academic achievement among online students. *Knowledge Management & E-Learning*, *12*(2), 191-208. https://doi.org/10.34105/j.kmel.2020.12.010

*Baker, R., Evans, B., Li, Q., & Cung, B. (2019). Does inducing students to schedule lecture watching in online classes improve their academic performance? An experimental analysis of a time management intervention. *Research in Higher Education*, *60*, 521-552. https://doi.org/10.1007/s11162-018-9521-3

*Bailey, C. J., Salois, S. H., & Bailey, K. (2017). Just-in-Time Teaching for Online Learning: Bridging the Student/Faculty Gulf. *International Journal of Technologies in Learning*, *24*(2).

Barclay, C., Donalds, C., & Osei-Bryson, K. M. (2018). Investigating critical success factors in online learning environments in higher education systems in the Caribbean. *Information Technology for Development*, *24*(3), 582-611. https://doi.org/10.1080/02681102.2018.1476831

*Basila, C. (2014). Good time management and motivation level predict student academic success in college on-line courses. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, *4*(3), 45-52. https://doi.org/10.4018/ijcbpl.2014070104

Beck, D., Maranto, R., Tran, B., Clark, T., & Liu, F. (2023). Forced choice? Is bullying pushing non-binary students into cyber schools? *Educational Review*. Advance online publication. https://doi.org/10.1080/00131911.2023.2167068

Beran, T., & Li, Q. (2007). The relationship between cyberbullying and school bullying. *The Journal of Student Wellbeing, 1*(2), 15–33. https://doi.org/10.21913/JSW.v1i2.172

Bhuasiri, W., Xaymoungkhoun, O., Zo, H., Rho, J. J., & Ciganek, A. P. (2012). Critical success factors for e-learning in developing countries: A comparative analysis between ICT experts and faculty. *Computers & Education*, *58*(2), 843-855. https://doi.org/10.1016/j.compedu.2011.10.010

Borup, J. (2018). On-site and online facilitators: Current and future direction for research. In K. Kennedy and R. Ferdig (Eds.), *Handbook of research on K-12 online and blended learning* (2nd ed., pp. 423-442). ETC Press. https://press.etc.cmu.edu/books/handbook-research-k-12-online-and-blended-learning

Borup, J., Graham, C. R., West, R. E., Archambault, L., & Spring, K. J. (2020). Academic communities of engagement: An expansive lens for examining support structures in blended and online learning. *Educational Technology Research and Development, 68,* 807-832. https://doi.org/10.1007/s11423-020-09744-x

Borup, J. & Stimson, R. (2019). Responsibilities of online teachers and on-site facilitators in online high school courses. *American Journal of Distance Education*, *33*(1), 29-45. https://doi.org/10.1080/08923647.2019.1554984

*Breneiser, J. E., Rodefer, J. S., & Tost, J. R. (2018). Using tutorial videos to enhance the learning of statistics in an online undergraduate psychology course. *North American Journal of Psychology*, *20*(3)

*Blieck, Y., Kauwenberghs, K., Zhu, C., Struyven, K., Pynoo, B., & DePryck, K. (2019). Investigating the relationship between success factors and student participation in online and blended learning in adult education. *Journal of Computer Assisted Learning*, *35*(4), 476-490. https://doi.org/10.1111/jcal.12351

*Bognár, L., Fauszt, T., & Váraljai, M. (2021). The impact of online quizzes on student success. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(11), 225-244. https://doi.org/10.3991/ijet.v16i11.21679

Brubacher, M. R., & Silinda, F. T. (2021). First-generation students in distance education program: Family resources and academic outcomes. *The International Review of Research in Open and Distributed Learning*, *22*(1), 135–147. https://doi.org/10.19173/irrodl.v22i1.4872

Bruno, A. V. & Leidecker, J. K. (1984). Identifying and using critical success factors. *Long Range Planning, 17,* 23-32

*Chatman, T., Dick, D., Ford, P., Henry, P., Hobert, K., Keller, M., Riley, K., Tidwell, C., & Wright, R. (2019). Increasing success with online degree courses and programs in the VCCS. *Inquiry: The Journal of the Virginia Community Colleges, 22*(1), Article 9. https://commons.vccs.edu/inquiry/vol22/iss1/9

Cheawjindakarn, B., Suwannatthachote, P., & Theeraroungchaisri, A. (2013). Critical success factors for online distance learning in higher education: A review of the literature. *Creative Education*, *3*(08), 61. https://doi.org/10.4236/ce.2012.38b014

*Chen, L. (2023). Transactional distance and college students' learning engagement in online learning: The chain mediating role of social presence and autonomous motivation. *Psychology Research and Behavior Management*, 2085-2101. https://doi.org/10.2147/prbm.s409294

Chu, R. J. C. (2010). How family support and internet self-efficacy influence the effects of e-learning among higher aged adults–Analyses of gender and age differences. *Computers & Education*, *55*(1), 255-264.https://doi.org/10.1016/j.compedu.2010.01.011

Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). Preface. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement*. Springer.

*Curtis, H., & Werth, L. (2015). Fostering student success and engagement in a K-12 online school. *Journal of Online Learning Research*, *1*(2), 163-190.

*Devers, E., Devers, C., Miller, P., Alayan, A., Robertson, J., Baquet, Z., ... & Heavner, S. (2022). Online tutoring: Impact on student success. *Journal of Interactive Learning Research*, *33*(1), 57-70. https://doi.org/10.70725/818190bezwol

*Du, X., Zhang, M., Shelton, B. E., & Hung, J. L. (2022). Learning anytime, anywhere: a spatio-temporal analysis for online learning. *Interactive Learning Environments*, *30*(1), 34-48. https://doi.org/10.1080/10494820.2019.1633546

*Drouin, M., Stewart, J., & Van Gorder, K. (2015). Using methodological triangulation to examine the effectiveness of a mentoring program for online instructors. *Distance Education*, *36*(3), 400-418. https://doi.org/10.1080/01587919.2015.1081735

Ertmer, P. A., Ottenbreit-Leftwich, A., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education, 59*(2), 423–435. https://doi.org/10.1016/j.compedu.2012.02.001

Faza, A., & Lestari, I. A. (2025). Self-Regulated Learning in the Digital Age: A Systematic Review of Strategies, Technologies, Benefits, and Challenges. *The International Review of Research in Open and Distributed Learning*, *26*(2), 23–58. https://doi.org/10.19173/irrodl.v26i2.8119

*Feldman-Maggor, Y., Blonder, R., & Tuvi-Arad, I. (2022). Let them choose: Optional assignments and online learning patterns as predictors of success in online general chemistry courses. *The Internet and Higher Education*, *55*, 100867. https://doi.org/10.1016/j.iheduc.2022.100867

Gao, H., Ou, Y., Zhang, Z., Ni, M., Zhou, X., & Liao, L. (2021). The relationship between family support and e-learning engagement in college students:The mediating role of e-learning normative consciousness and behaviors and self-efficacy. *Frontiers in Psychology*, *12*, 573779. https://doi.org/10.3389/fpsyg.2021.573779

*Glazier, R. A. (2016). Building rapport to improve retention and success in online classes. *Journal of Political Science Education*, *12*(4), 437-456. https://doi.org/10.1080/15512169.2016.1155994

*Goad, T., Jones, E., Bulger, S., Daum, D., Hollett, N., & Elliott, E. (2021). Predicting student success in online physical education. *American Journal of Distance Education*, *35*(1), 17-32. https://doi.org/10.1080/08923647.2020.1829254

*Gomez del Rey, P., Barbera Gregori, E., & Fernández Navarro, F. (2016). The impact of cultural dimensions on online learning. *Educational Technology and Society*, *19*(4).

*González, C. A. H., & Blackford, B. J. (2022). Engagement as antecedent of academic achievement and the moderating impact of work-family-school inter-role conflict for online graduate students. *The International Journal of Management Education*, *20*(3), 100676. https://doi.org/10.1016/j.ijme.2022.100676

Goodrich, A. (2021). Online peer mentoring and remote learning. *Music Education Research*, *23*(2), 256–269. https://doi.org/10.1080/14613808.2021.1898575

*Gravelle, C. D., Brodsky, J. E., Lodhi, A. K., Zapparrata, N. M., Che, E. S., Ober, T. M., & Brooks, P. J. (2024). Remote online learning outcomes in introductory psychology during the COVID-19 pandemic. *Scholarship of Teaching and Learning in Psychology*, *10*(4), 442. https://doi.org/10.1037/stl0000325

*Guidry, K. (2013). Predictors of Student Success in Online Courses: Quantitative versus Qualitative Subject Matter. *Journal of Instructional Pedagogies*, *10*.

Hillman, D. C., Willis, D. J., & Gunawardena, C. (1994). Learner-interface interaction in distance education: An extension of contemporary models and strategies for practitioners. *American Journal of Distance Education, 8*(2), 30–42. https://doi.org/10.1080/08923649409526853

Hoover-Dempsey, K. V., Walker, J. M. T., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., et al. (2005). Why do parents become involved? Research findings and implications. *The Elementary School Journal, 106*(2), 105–130. https://doi.org/10.1086/499194

*Idrizi, E., Filiposka, S., & Trajkovijk, V. (2021). Analysis of success indicators in online learning. *International Review of Research in Open and Distributed Learning*, *22*(2), 205-223. https://doi.org/10.19173/irrodl.v22i2.5243

*Jacobi, L. (2017). The structure of discussions in an online communication course: What do students find most effective? *Journal of University Teaching and Learning Practice, 14*(1). https://doi.org/10.53761/1.14.1.4

*Kane, R. T., Shaw, M., Pang, S., Salley, W., & Snider, J. B. (2015). Relationships among faculty training, faculty degree, faculty longevity, and student satisfaction in online higher education.

Online Journal of Distance Learning Administration, 18(4). https://www.westga.edu/~distance/ojdla/winter184/kane_shaw_pang_salley_snider184.html

*Krishnan, S. (2023). Factors influencing educational equity in online instruction for students during the COVID-19 pandemic. *Remedial and Special Education*, *44*(6), 469-479. https://doi.org/10.1177/07419325231161322

*Latheef, Z. I., Robinson, R., & Smith, S. (2021). Realistic Job Preview as an Alternative Tool to Improve Student Readiness for Online Learning. *Online Learning*, *25*(2), 208-232. https://doi.org/10.24059/olj.v25i2.2216

Leary, H., *Dopp, C., Turley, C., Cheney, M., Simmons, Z., Graham, C. R., & Hatch, R. (2020). Professional development for online teaching: A literature review. *Online Learning, 24*(4), 254- 275. https://doi.org/10.24059/olj.v24i4.2198

*Leasure, D., Blaher, S., Davis, C., Ellsworth, E., Fortney, M., Hansen, M., ... & Willis, H. (2020). Not So Suddenly Online: Preparing UMGC's Students and Faculty for Online Success. *Journal of Literacy & Technology*, *21*(2).

*Lee, Y., Choi, J., & Kim, T. (2013). Discriminating factors between completers of and dropouts from online learning courses. *British Journal of Educational Technology*, *44*(2), 328-337. https://doi.org/10.1111/j.1467-8535.2012.01306.x

Lehan, T. J., Hussey, H. D., & Shriner, M. (2018). The influence of academic coaching on persistence in online graduate students. *Mentoring & Tutoring: Partnership in Learning*, *26*(3), 289–304. https://doi.org/10.1080/13611267.2018.1511949

Lehan, T., Shriner, M., & Shriner, B. (2020). It's complicated: The relationship between participation in academic coaching and program completion in online graduate students. *Online Learning*, *24*(3), 19-34.https://doi.org/10.24059/olj.v24i3.2142

Lenio, J. A. (2021). Employer Support as a Predictor of Online Master's Student Retention. *Journal of College Student Retention: Research, Theory & Practice*, *25*(3), 572-593. https://doi.org/10.1177/15210251211001561

*Leow, M. H. & Abdul Razak, R. (2024). Malaysian elementary learners' self-regulation, motivational beliefs and learner control motivation when experiencing online tutorials, *Turkish Online Journal of Distance Education, 25*(1), 334-351. https://doi.org/10.17718/tojde.1262408

Liu, J. C. (2019). Evaluating online learning orientation design with a readiness scale. *Online Learning*, *23*(4), 42-61. https://doi.org/10.24059/olj.v23i4.2078

*Liu, Z. (2023). Face-to-Face and Online Learning in Higher Education: Academic Achievements and Learners' Experience in the Chinese SFL Context. *SAGE Open*, *13*(4). https://doi.org/10.1177/21582440231218114

*Linden, K., Teakel, S., & Van der Ploeg, N. (2022). Improving student success with online embedded tutor support in first-year subjects.'A practice report'. *Student Success*, *13*(2), 42-50. https://doi.org/10.5204/ssj.2338

Lowes, S., & Lin, P. (2015). Learning to learn online: Using locus of control to help students become successful online learners. *Journal of Online Learning Research*, *1*(1), 17–48.

*Maré, S., & Mutezo, A. T. (2021). The effectiveness of e-tutoring in an open and distance e-learning environment: Evidence from the University of South Africa. *Open Learning: The Journal of Open, Distance and e-Learning*, *36*(2), 164-180. https://doi.org/10.1080/02680513.2020.1717941

Martin, F., Sun, T., & Westine, C. D. (2020). A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & education*, *159*, 104009.

Martin, F., & Borup, J. (2022). Online learner engagement: Conceptual definitions, research themes, and supportive practices. *Educational Psychologist*, *57*(3), 162-177.

Martin, F. (2025). Systematic Reviews in Educational Research: A Practical Guide for Research Synthesis. SAGE Publications. https://collegepublishing.sagepub.com/products/systematic-reviews-in-educational-research-1-285211

*McGuinness, N., & Vlachopoulos, D. (2019). Student Experiences of Using Online Material to Support Success in A-Level Economics. *International Journal of Emerging Technologies in Learning*, *14*(3). https://doi.org/10.3991/ijet.v14i03.9406

Moon, H. (2025). Investigating the Role of Orientation in Enhancing Student Experience within Online Learning. *The Journal of Applied Instructional Design*, *14*(2). https://doi.org/10.59668/2222.20826

*Mshayisa, V. V., & Ivala, E. N. (2022). No student left behind: Students' experiences of a self-paced online learning orientation in undergraduate studies during COVID-19 pandemic. *Education Sciences*, *12*(6), 386. https://doi.org/10.3390/educsci12060386

Muljana, P. S., & Luo, T. (2019). Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review. *Journal of Information Technology Education: Research*, *18*. https://doi.org/10.28945/4182

Nakayama, M., & Santiago, R. (2012). Learner characteristics and online learning. In *Encyclopedia of the sciences of learning* (pp. 1745-1747). Springer, Boston, MA. https://doi.org/10.1007/springerreference_302098

*Newberry, R. & DeLuka, C. (2014). Building a foundation for success through student services for online learners. *Online Learning Journal*, *17*(4). https://doi.org/10.24059/olj.v17i4.385

*Newhouse, N. K., & Cerniak, J. (2016). Student Success Factors in Graduate Psychology Professional Programs. *Online Learning*, *20*(1), 70-91. https://doi.org/10.24059/olj.v20i1.585

*Odabaş, C., & Kahramanoğlu, R. (2023). Qualitative analysis of the relationship between online homework and prior success, self-efficacy, perceived responsibility, motivation and academic success. *International Journal of Modern Education Studies, 7*(2), 619–639. https://doi.org/10.51383/ijonmes.2023.349

Oviatt, D., Graham, C. R., Borup, J. & Davies, R. S. (2018). Online student use of a proximate community of engagement at an independent study program. *Online Learning, 22*, 223-251. https://olj.onlinelearningconsortium.org/index.php/olj/article/view/1153

*Pazzaglia, A. M., Clements, M., Lavigne, H. J., & Stafford, E. T. (2016). *An analysis of student engagement patterns and online course outcomes in Wisconsin* (REL 2016–147). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. https://ies.ed.gov/ncee/rel/regions/midwest/pdf/REL_2016147.pdf

*Price, J., Luo, T., Bol, L., & Watson, G. (2021, July). An analysis of self-regulated learning strategies, academic performance, and satisfaction among online high school graduates. *I*n G. Marks (Ed.) *Proceedings of International Journal on E-Learning* (pp. 459-487). Association for the Advancement of Computing in Education (AACE) https://doi.org/10.70725/546092euekwx

*Ren, X. (2023). Investigating the experiences of online instructors while engaging and empowering non-traditional learners in eCampus. *Education and Information Technologies*, *28*(1), 237-253. https://doi.org/10.1007/s10639-022-11153-x

Roblyer, M. D., Freeman, J., Stabler, M., & Schneidmiler, J. (2007). External evaluation of the Alabama ACCESS initiative phase 3 report. Eugene, OR: International Society for Technology in Education. Retrieved from http://accessdl.state.al.us/2006Evaluation.pdf

*Rosser-Majors, M. L., Rebeor, S., McMahon, C., Wilson, A., Stubbs, S. L., Harper, Y., & Sliwinski, L. (2022). Improving Retention Factors and Student Success Online Utilizing the Community of Inquiry Framework's Instructor Presence Model. *Online Learning*, *26*(2), 6-33. https://doi.org/10.24059/olj.v26i2.2731

Samra, R., Waterhouse, P., & Lucassen, M. (2021). Combining and managing work-family-study roles and perceptions of institutional support. *Distance Education*, *42*(1), 88–105. https://doi.org/10.1080/01587919.2020.1869530

*Sari, D. P., & Rimbano, D. (2022). Online Learning Experience During Covid-19 Pandemic: Higher Education Students Satisfaction and Expectation. *Journal of Higher Education Theory and Practice*, *22*(11), 187-202. https://doi.org/10.33423/jhetp.v22i11.5423

Scagnoli, N. I. (2001). Student Orientations for Online Programs. *Journal of Research on Technology in Education*, *34*(1), 19–27. https://doi.org/10.1080/15391523.2001.10782330

*Scarpena, K., Riley, M., & Keathley, M. (2018). Creating successful professional development activities for online faculty: A reorganized framework. *Online Journal of Distance Learning Administration, 21*(1), https://www.westga.edu/~distance/ojdla/spring211/scarpen_riley_keathley211.html

*Seiver, J. G., & Troja, A. (2014). Satisfaction and success in online learning as a function of the needs for affiliation, autonomy, and mastery. *Distance education*, *35*(1), 90-105. https://doi.org/10.1080/01587919.2014.891427

Selim, H. M. (2007). Critical success factors for e-learning acceptance: Confirmatory factor models. *Computers & education*, *49*(2), 396-413. https://doi.org/10.1016/j.compedu.2005.09.004

*Skelcher, S., Yang, D., Trespalacios, J., & Snelson, C. (2020). Connecting online students to their higher learning institution. *Distance Education*, *41*(1), 128-147. https://doi.org/10.1080/01587919.2020.1724771

Springer, S. (2018). One university's experience partnering with an online program management (OPM) provider: A case study. *Online Journal of Distance Learning Administration, 21*(1). https://www.westga.edu/~distance/ojdla/spring211/springer211.html

*Stone, C. M. M., & O'Shea, S. E. (2019). My children… think it's cool that Mum is a uni student: Women with caring responsibilities studying online. *Australasian Journal of Educational Technology*, *35*(6), 97-110. https://doi.org/10.14742/ajet.5504

Sun, J. and Turner, H. A. (2022). In-House or Outsource? Chief online learning officers' decision-making factors when considering online program managers. University Professional and Continuing Education Association. https://upcea.edu/wp-content/uploads/2022/10/In-House-or-Outsource_UPCEA-and-University-of-Louisville-SKILLS-Collaborative_October-2022-1.pdf

*Tainsh, R. D. (2023). The Impact of Self-Directedness on the Academic Success and Persistence of Postsecondary Online Learners. *International Journal of Research in Education and Science*, *9*(1), 92-108. https://doi.org/10.46328/ijres.3010

*Townley, A. L., Soares, L. B., & Rahimi, R. (2023). Ready or not? A pilot study of graduate students' and instructors' perceptions of the effectiveness of online support modules for enhancing academic performance. *Education Sciences*, *13*(7), 656. https://doi.org/10.3390/educsci13070656

*Tseng, H., Yi, X., & Yeh, H. T. (2019). Learning-related soft skills among online business students in higher education: Grade level and managerial role differences in self-regulation, motivation, and social skill. *Computers in Human Behavior*, *95*, 179-186. https://doi.org/10.1016/j.chb.2018.11.035

*Van Wart, M., Ni, A., Medina, P., Canelon, J., Kordrostami, M., Zhang, J., & Liu, Y. (2020). Integrating students' perspectives about online learning: a hierarchy of factors. *International Journal of Educational Technology in Higher Education*, *17*(1), 53. https://doi.org/10.1186/s41239-020-00229-8

*Welch, A. G., Orso, D., Doolittle, J., & Areepattamannil, S. (2015). Matching student expectations with instructors' dispositions: Insight into quality of online teaching. *The Journal of Effective Teaching, 15*(2), 5–19.

Zawacki-Richter, O., & Jung, I. (2022). Shaping the field of open, distance, and digital education. In *Handbook of Open, Distance and Digital Education* (pp. 1–10). Springer Nature. https://doi.org/10.1007/978-981-19-0351-9_94-2

*Zheng, B., Lin, C. H., & Kwon, J. B. (2020). The impact of learner-, instructor-, and course-level factors on online learning. *Computers & Education*, *150*, 103851. https://doi.org/10.1016/j.compedu.2020.103851

Appendix A - Codebook

Codebook Items	Description
Article Title	Publication title
Publication Year	Year of publication
Journal name	Journal name
Learning Context	K-12 or Higher Ed
Research Methodology	Quantitative, Qualitative or Mixed-Method
Dimensions	Learner, Course and Instructor, Program and Community, Organization
Success Factors	Open-coded in first round, and then closed coded into 30 categories of success factors

Appendix B - List of articles included in review

https://docs.google.com/document/d/15P8FKnKDmunrc1X-sVllWFOWd1SS-V5B/edit?pli=1

ONLINE LEARNING
Volume 30 · Issue SI2 · 2026 · e5747
https://doi.org/10.24059/olj.v30iSI2.5747

The Integration of Generative Artificial Intelligence into Tertiary Health Education: A Scoping Review

Lok Tin Cheung
Haley Vu
Brooke Osborne
Nayana Parange
Sandhya Maranna
Adelaide University

Abstract

This scoping review aims to provide an overview of the integration of generative artificial intelligence (GenAI) in tertiary health education, including its practical applications in tertiary health education programs and how these applications are perceived by stakeholders. A literature search was conducted in five databases, including Ovid Medline, Ovid Emcare, Scopus, CINAHL, and ERIC. Articles integrating GenAI in tertiary health disciplines were included. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews checklist was followed in the reporting of results. Overall, 20 studies were included and three categories were identified: 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. Currently, GenAI is being integrated as a complementary tool to automate academic tasks and improve the critical thinking skills of tertiary health students through optimizing class designs. Although GenAI applications were well-received by students due to high accessibility, concerns over misinformation, over-reliance, and data confidentiality were consistently flagged. The review identified a lack of formal AI literacy training, highlighting the need for structured AI literacy training modules and clear ethical guidelines to support tertiary health students in using GenAI. GenAI applications are being rapidly integrated into tertiary health education, primarily for automated grading/feedback and improvement of learning outcomes. Future research is recommended to investigate the long-term impacts of GenAI and how to implement AI literacy training in current tertiary health education programs.

Keywords: Generative artificial intelligence, artificial intelligence, ChatGPT, tertiary health education, medical education, teaching and learning

Introduction

Generative artificial intelligence (GenAI) refers to a series of deep learning algorithms designed to generate new content from existing datasets through machine learning (Chan & Hu, 2023). At the core of GenAI are large language models, a type of artificial neural network trained to learn language structures and patterns from input datasets, and generate new, mostly text-based content based on prompts entered by users (Farazouli et al., 2023). The most prominent example of GenAI is the Chat Generative Pre-trained Transformer (ChatGPT), a GenAI tool that generates automatic responses to users' prompts in real time, based on patterns it learned from vast amount of training data (Gupta et al., 2023). Since its release in November 2022, GenAI technologies have undergone rapid developments, with new tools such as Stable Diffusion,

Corresponding author: Sandhya Maranna, sandy.maranna@adelaide.edu.au
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

Perplexity, and more advanced versions of ChatGPT developed and new features such as text-to-image conversion and advanced analysis (Shoja et al., 2023). With its growing capability to support interactive and iterative user engagement, GenAI has been increasingly applied across multiple sectors, including tertiary health education (Chan & Hu, 2023). One example of GenAI integration into tertiary health education includes the use of ChatGPT to generate multiple-choice questions for use in graduate medical examinations, effectively reducing the workload of university medical staff (Cheung et al., 2023). Another notable application involves employing ChatGPT to score the free-text and physical examination notes of medical students, displaying its potential as an efficient scoring and feedback tool in clinical education (Burke et al., 2024).

Tertiary Health Education and the Need for GenAI

Tertiary health education encompasses university-based education and training programs that prepare students for careers in the health professions, including medicine, nursing, midwifery, dentistry, pharmacy, allied health, and medical imaging (Allied health professions Australia, n.d; Gill & Galletly, 2025; Koontz & Podlog, 2025). Unlike many other higher education disciplines, health professions education integrates theoretical knowledge with clinical reasoning, psychomotor skill development, communication, professionalism, ethical decision-making, and supervised workplace learning. Educational approaches typically include simulation (Takhdat et al., 2026), Holderried problem-based learning, case-based discussions, laboratory training, and clinical placements, with graduates expected to demonstrate competence in providing safe, evidence-based, patient-centered care (Kotzki et al., 2026).

These distinctive characteristics make the integration of GenAI particularly important within health professions education. Healthcare systems are rapidly adopting AI-enabled technologies to support clinical decision-making, diagnostic interpretation, documentation, and patient management. Consequently, graduates require not only discipline-specific knowledge but also the ability to critically evaluate AI-generated information, understand its limitations, recognize potential biases, and use AI responsibly within professional and ethical boundaries (Daniel et al., 2025). Furthermore, GenAI offers opportunities to transform health education through personalised learning, adaptive tutoring, automated feedback, realistic clinical case generation, simulation, and support for evidence-based learning. At the same time, concerns regarding academic integrity, learner over-reliance, data privacy, equity (Atif et al., 2026), and clinical accountability necessitate careful educational integration supported by appropriate governance and curriculum design (Crompton et al., 2026).

Literature Review

Emerging research has begun to explore the role of GenAI in tertiary health education, with existing studies primarily looking at aspects such as technical performance, the perceived benefits, and potential drawbacks on student learning as seen in recent reviews by Preiksaitis & Rose, (2023) and Shoja et al., (2023). Furthermore, although current studies provide valuable insights on GenAI in tertiary health education, the evidence remains dispersed and often discipline-specific such as dentistry (Uribe et al., 2025), nursing (Hardie et al, 2026; Hinsche et al., 2026), and pharmacy (Mortlock & Lucas, 2024). This fragmentation makes it difficult for educators, curriculum designers, and policymakers to grasp the broader landscape of GenAI use, evaluate its pedagogical value, or anticipate risks emerging from its adoption. Given the rapid

expansion and diversification of GenAI technologies, a consolidated understanding of how these tools are being integrated into broad health education is needed.

Toohey et al. (2025) proposed a framework through their scoping review for a structured approach for integrating GenAI into health education, emphasizing responsible adoption across curriculum, assessment, governance, and workforce preparation. It highlights the need for AI literacy, educator capability, assessment design and evidence-informed implementation to support safe, effective, and sustainable GenAI-enhanced learning. Lin et al. (2025) in their review identified key challenges in integrating GenAI into medical education, including inaccurate or unreliable AI-generated content, academic integrity concerns, privacy and ethical risks, limited AI literacy among educators and students, over-reliance on AI, digital inequities, and insufficient evidence regarding educational effectiveness. The review highlights the need for governance, training, and responsible implementation strategies (Lin et al., 2025). A systematic review by Pham et al. (2025) investigated how students in health profession utilize GenAI using the Laurillard's six learning types and compared it to conventional and digital technologies. After evaluating 33 studies, they found that there was a shift to autonomous learning with GenAI and a reduction in collaborative learning (Pham et al., 2025).

Several scoping and systematic reviews have examined the role of GenAI in higher education; however, important gaps remain regarding its application within health professions education. Existing reviews of GenAI in higher education have primarily focused on broad educational contexts, including student engagement, academic integrity, assessment practices, and teaching efficiency, and have generated general recommendations for educators and institutions (Belkina et al., 2026; Daniel et al., 2025). While valuable, these reviews do not adequately address the distinctive characteristics of health professions education, which requires integration of theoretical knowledge with clinical reasoning, practical skill development, professional judgement, ethical decision-making, and patient-centered care.

Within health education, previous reviews have predominantly examined artificial intelligence more broadly rather than specifically focusing on GenAI and large language models (Kovalainen et al., 2025). These reviews provide important insights into the broader role of AI in healthcare education but do not capture the unique opportunities and challenges introduced by GenAI, including its capacity to generate clinical cases, provide conversational feedback, support personalised learning, assist with simulation-based education, and influence assessment practices. Furthermore, reviews focusing on specific health disciplines, such as medicine (Lin et al., 2025), provide valuable discipline-specific insights but may not represent the diversity of health professions education, including nursing, allied health, pharmacy, dentistry, and medical imaging.

Therefore, a gap exists for a comprehensive synthesis that specifically examines the integration of GenAI across tertiary health professions education. The current scoping review addresses this gap by mapping the rapidly emerging evidence base following the widespread adoption of ChatGPT and other large language models, identifying common themes across health disciplines, and highlighting discipline-specific opportunities and challenges.

The primary objective of this scoping review is to systematically explore the current landscape of how GenAI is currently being integrated into tertiary health education (i.e.,

undergraduate or postgraduate programs at university level) for health professional students (e.g., allied health, medical and dentistry cohorts). Specifically, the review aims to categorize existing GenAI applications employed, their practical roles in curriculum development within the tertiary health education setting and the perceptions of stakeholders (students and educators) towards these applications. By synthesizing themes from existing literature on the recent GenAI applications in tertiary health education, this review aimed to identify key trends and concerns in the literature that can guide future directions on how to integrate GenAI applications smoothly and effectively into the tertiary health education sector.

Methods

Study Design

This study is a scoping review conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews checklist (PRISMA ScR) (Peters et al., 2020; Tricco et al., 2018). The protocol of the scoping review was guided by the five-stage framework for scoping reviews (Arksey & O'Malley, 2005). The Population, Concept, Context (PCC) framework was utilized to help construct clear eligibility criteria and objectives of the project (Pollock et al., 2023). The protocol of this project was registered with the Open Science Framework (Cheung et al., 2025).

A scoping review was selected to address the research question since it was useful in answering broad questions, identifying gaps in literature and summarizing findings from a broad area of knowledge (Pollock et al., 2023). By mapping information from all types of study designs as well as grey literature, a scoping review was believed to effectively break down the broad and complex topic of the integration of GenAI into health education. Its comprehensive summary of the available literature could also provide breadth into the chosen research topic and help guide future research. We identified the following research question to guide the search: How is GenAI being integrated into tertiary health education for health professional students?

Eligibility Criteria

The inclusion and exclusion criteria were established using the PCC (Population, Concept, Context) framework:

- Population: Tertiary level students and educators in health and medical disciplines
- Concept: The integration of GenAI into tertiary health education
- Context: Tertiary health education sector (undergraduate level or above)

The population of interest was the tertiary level students and educators in health, and medical disciplines. These disciplines include doctors, nurses, and allied health practitioners. As defined by Allied Health Professions Australia, allied health practitioners include art therapists, audiologists, nutritionists, chiropractors, counsellors/psychotherapists, diabetes educators, dieticians, exercise physiologists, genetic counsellors, medical radiation practitioners, music therapists, occupational therapists, optometrists, orthoptists, prosthetists, osteopaths, paramedics, pedorthists, perfusionists, pharmacists, physiotherapists, podiatrists, psychologists, rehabilitation counsellors, social workers, sonographers, and speech pathologists (Allied health professions

Australia, n.d.). Studies involving any of the above disciplines were eligible for inclusion. Various inclusion and exclusion criteria were put in place, as demonstrated in Table 1.

Table 1

Table of Inclusion and Exclusion Criteria.

Inclusion Criteria	Exclusion Criteria
Peer-reviewed studies Empirical studies with all types of study designs Studies in English Grey literature	Editorials, review articles, websites, letters to editor Papers with only abstract (no full text) Studies not in English
Tertiary health education Medical education	Studies not focusing on the tertiary education sector Studies not focusing on GenAI, but only artificial intelligence

Types of Evidence Sources

This review included empirical research articles with various types of study designs, including but not limited to randomized controlled trials, experimental, qualitative and descriptive studies, if they focused specifically on GenAI and the tertiary level of health education. Conceptual, theoretical, commentary, reviews and opinion papers were excluded unless they presented original data relevant to GenAI use in tertiary health education. Editorials, websites, opinion pieces, conference abstracts, and non-English studies were excluded. Literature search was limited to publications from 2021 onwards to maintain currency of sources.

Since the research question was relatively broad, the scoping review included original research of all types of study designs and peer-reviewed sources to maximize data yield and the mapping of information. Review articles, editorials, websites and letters to editor were excluded. A preliminary attempt at searching for articles in languages other than English was made, yielding three in Japanese, one each in Chinese, French and German. However, free translation tools such as Google Translate and DeepL Translator proved unreliable in translating the true meaning of statements. Since the project did not have funding for a translator, it was decided that articles not in English were excluded. Furthermore, sources that addressed artificial intelligence instead of GenAI were excluded as they were not specific to the identified concept. Similarly, sources unrelated to the tertiary or higher education sector were excluded due to a lack of specificity towards the context.

The release of ChatGPT in November 2022 was the major catalyst for the growing number of GenAI tools developed and applied in the tertiary health education sector. However, the research team decided to begin the literature search from 2021 to ensure that relevant publications prior to the introduction of ChatGPT were also captured.

Search Strategy

A three-step search strategy was used to develop the search terms with the help of an academic librarian. Initial search on Ovid Medline helped to develop the final search terms. The final search terms, including Boolean operators, were developed (See Table 2). This comprised keywords and subject headings related to GenAI and tertiary health, and medical education such as "Artificial Intelligence" OR "Generative AI" OR "Generative Artificial Intelligence" OR "Machine learning" OR chatgpt OR chatbot OR "Large Language Model", AND "Health Education" OR "Medical Education" OR "Healthcare Student" OR "Medical Student" OR "Allied Health Student" OR "Allied Health Education", AND "Tertiary Education" OR undergraduate OR postgraduate OR "Higher education" OR “university”. Finally, five databases, Ovid Medline, Ovid Emcare, Scopus, CINAHL, and ERIC, were systematically searched for English articles published from 2021 to October 2024. The database search was conducted by the first author on 15 November 2024. The complete search string for all the databases is available as an open access dataset (Cheung et al., 2025).

Table 2

Final Search Terms

PCC	Concepts	Keywords	Subject Headings
Population	Tertiary education students and educators in health disciplines	Health educa*, Medical educa*, allied health educat*, healthcare student*, medical student* and allied health student*	Students, Health Occupations Students, Medical Health Educators Allied Health- Personnel Allied Health- Occupations
Concept	Integration of GenAI (into tertiary health education)	Generative artificial intelligence, generative AI, generative language model, artificial intelligence, machine learning, virtual, simulat*, ChatGPT, chatbot, automatic scoring, AI literacy and integrat*	Artificial Intelligence Machine Learning Simulation training
Context	Tertiary health education	tertiary health educat*, undergrad* and postgrad*	Health Education Education, Medical

Screening

The database search results were added to the Covidence systematic review software (2024) for removing duplicates and screening. A total of 598 studies underwent title and abstract screening, followed by full text screening. Two reviewers independently assessed records against the inclusion and exclusion criteria. A third author resolved the discrepancies that arose in abstract and full text screening. All conflicts were resolved through discussion and reaching a consensus among the research team. The inclusion and exclusion criteria were further refined before full-text review proceeded. The inclusion criteria stated that articles must focus on GenAI and the tertiary health education sector. Only those articles which met the inclusion and exclusion criteria with a 100% agreement rate between reviewers were included in the final tally.

Second Iteration of the Search Strategy

A second search was conducted on July 19, 2026 on SCOPUS by the corresponding author. The search string for SCOPUS was used from the search string used in the initial search after editing the dates to search for articles from November 2024 to July 2026. Additionally, a manual search of the first 50 articles on Google Scholar was also conducted by the first author. The second search was targeted rather than replicating the full database search. The search strategy, databases, date, and eligibility criteria are reported as per guidelines (Peters et al., 2020; Tricco et al., 2018) to ensure transparency and reproducibility. This consensus decision by the project team reflected the rapidly evolving AI literature. The supplementary search was intended to identify additional relevant literature rather than to claim exhaustive retrieval of all publications within this continually developing field. The search strings for Google Scholar and SCOPUS are given in Appendix A. Articles were screened for title and abstract by two reviewers as per inclusion and exclusion criteria. The articles for full-text screening were distributed to all authors such that at least two reviewers screened the articles independently. Any discrepancies were resolved by the third reviewer. Only those articles which met the inclusion and exclusion criteria with a 100% agreement rate between reviewers were included in the final tally.

Data Extraction

Data from included articles was extracted and entered in a Microsoft Excel spreadsheet in a study characteristics table to answer the research question, which included, the study, country, discipline, study design, application of GenAI, practical application and key findings. Data extraction was initially piloted for 3 articles between the first author and the corresponding author. After reaching more than 95% agreement on the data, the remaining data extraction was conducted by the first author and cross-checked by all the remaining authors. Disagreements were settled by discussion and consensus with the research team. The data extraction was consistent with PRISMA guidelines for scoping reviews; thus, the risk of bias was not assessed (Tricco et al., 2018).

Data Analysis

Subheadings included the following aspects: the characteristics of studies (publication year, country, sample size, course discipline), category of application, study design, and key findings of the studies. Categories were identified to guide the reporting themes, which were finalized after consensus by all authors.

Results

The search and selection process are reported using a PRISMA-ScR (Peters et al., 2020; Tricco et al., 2018) flow diagram (See figure 1). The initial search yielded 1112 titles, of which 514 were duplicates. After removal of duplicates, 598 studies were available for screening. Following title and abstract screening, 524 articles were removed as they did not specifically focus on GenAI, but only general AI. A total of 74 studies were included for full text screening. Full text screening resulted in removing studies which focused on the test performance of GenAI tools, instead of the integration of such tools into tertiary health education. These articles were removed after consensus among authors. A total of 13 articles were included in the final analysis. The second search, conducted in July 2026, yielded 402 articles for title and abstract screening. This resulted in 23 articles for full text screening. Finally, seven articles were added, bringing the

final tally to 20. The extracted data was categorized to answer the research question: “How is GenAI being integrated into tertiary health education?”.

Characteristics

As summarised in Table 3, the included articles were published between 2023 and 2026, reflecting the recency of the integration of GenAI into tertiary health education. Geographically, the 20 studies spanned across North America, Asia, Europe, and the Middle East. This includes the USA (n=5), China (n=5) and one each from Thailand, Taiwan, Germany, Switzerland, Bahrain, Israel, India, Turkey, Saudi Arabia and Hong Kong Special Administrative Region (HKSAR)/Singapore/Ireland/UK. 80% of studies focused on medical education (n=16), with the remainder covering nursing, anatomy, dentistry and healthcare management education. Most studies utilized ChatGPT (n=16), while two applied platform-specific adaptations, namely AnatomyGPT (Collins et al., 2024) and Glass AI (Robleto et al., 2024). Various methodologies were employed, which included randomised controlled trials (n=3), descriptive (n=5), prospective (n=3), experimental (n=1), quasi-experimental (n=6), and mixed-methods (n=2) designs. The reported methodologies of the included studies are intended to be mutually exclusive. The comprehensive data extraction table is available as an open access dataset for the initial search conducted in 2024 (Cheung et al., 2025).

Figure 1

PRISMA ScR Flowchart (Peters et al., 2020; Tricco et al., 2018)

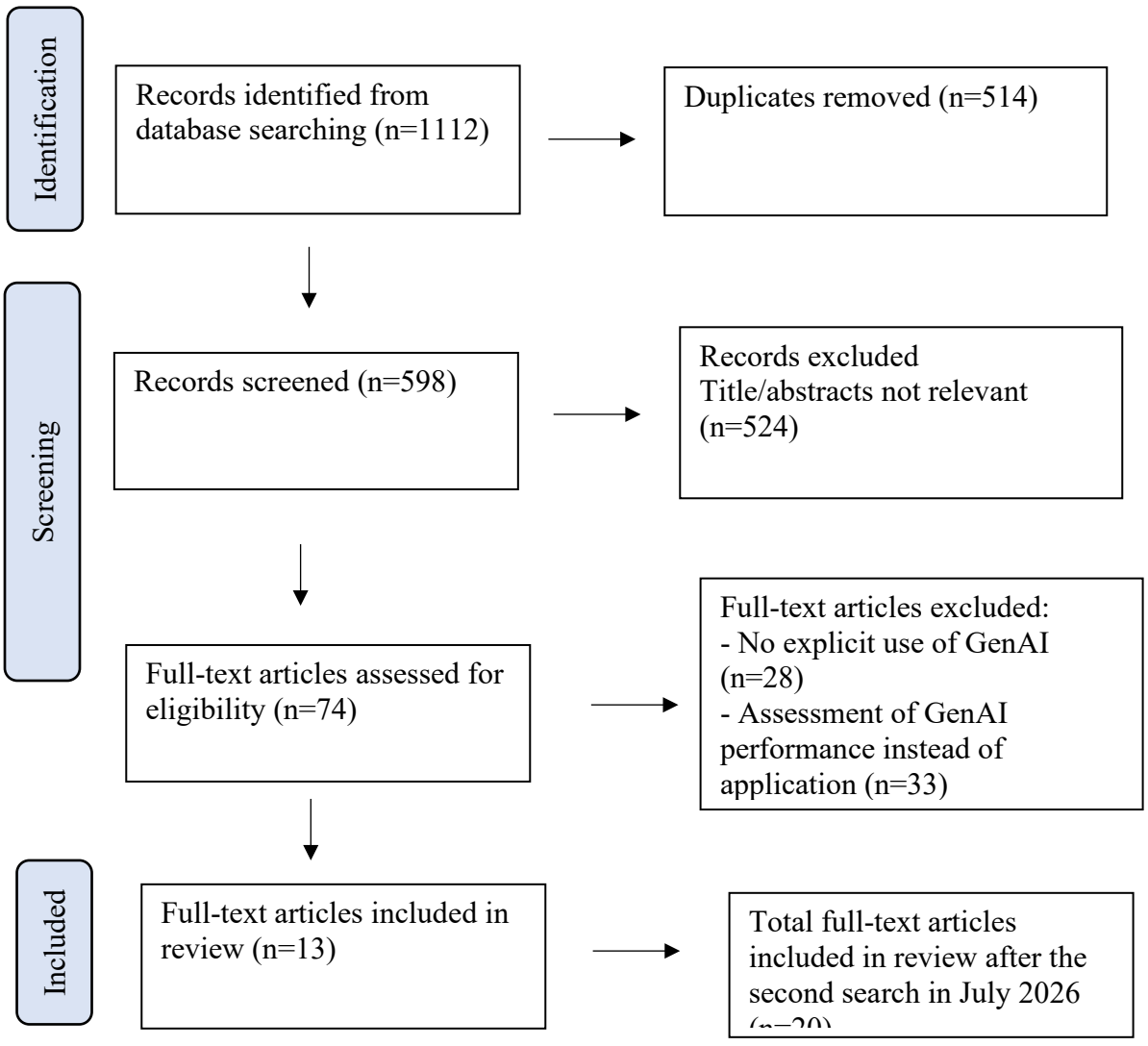


Table 3

Characteristics and key findings of all the included studies from the initial and the second search

Study	Country	Discipline	Study design	Application of GenAI	Practical application	Key findings	Concerns
1. Burke et al. (2024)	USA	Medical education	Cross-sectional descriptive	Scoring free-text clinical notes of medical students by ChatGPT	Automated grading, assessment generation, and feedback.	ChatGPT achieved a markedly lower incorrect scoring rate (1.0%) than the standardized rate (7.2%), suggesting LLMs can reliably score student notes in real time while providing immediate performance feedback.	
2. Chang et al. (2024)	Taiwan	Nursing education	Post-test quasi-experimental	Incorporation of CIDI (Clarify, Ideate, Develop, Implement) model-based ChatGPT system to promote critical thinking and problem-solving skills of students.	Improved learning outcomes	Learners using the CIDI model–based ChatGPT system (mean = 4.53; SD = 0.71) showed significantly greater gains in problem-solving skills compared with conventional methods (mean = 3.75; SD = 0.58). The interactive design promoted critical thinking by enabling conversational practice and judgment-based decision-making.	
3. Cheung et al. (2023)	HKSAR, Singapore, Ireland, and UK	Medical education	Comparative quasi-experimental	ChatGPT-generated multiple choice questions (MCQs) for medical examinations	Automated grading, assessment generation, and feedback	The time spent by ChatGPT to create 50 MCQs was 20 minutes 25 seconds, compared to 211 minutes 33 seconds by human examiners.	Reliability issues and potential bias in training data

Study	Country	Discipline	Study design	Application of GenAI	Practical application	Key findings	Concerns
4. Chiu et al.	USA	Medicine	Quasi-experimental	Breaking bad news training using predesigned scenarios in ChatGPT	Teach medical trainees the skills for breaking bad news.	Utilizing ChatGPT as a tool for communication skills training can enhance medical students' confidence levels and increase their perceived communication skills and professionalism particularly to patients	Privacy concerns
5. Cicek et al, 2024	Turkey	Medicine	Randomized controlled trial	To evaluate the effectiveness of ChatGPT-generated feedback compared to expert-written feedback in improving clinical reasoning skills among first-year medical students.	ChatGPT as supplementary tool in resource-limited settings where expert feedback is not readily available.	There was no significant difference between the mean scores of the control and experiment groups in overall performance on Key-Features Questions immediately or after 10 days. However, the control group outperformed the ChatGPT group in complicated urinary tract infection cases ($P < .001$).	While the effect of ChatGPT feedback was generally on par with the effect of expert feedback, the study identified limitations in AI-generated explanations for more nuanced diagnosis and treatment.
6. Collins et al. (2024)	USA	Anatomical Sciences education	Exploratory descriptive	Customized GenAI platform: AnatomyGPT, for anatomy education	Improvement in learning outcomes	AnatomyGPT provided more source citations than standard GPT when generating responses.	Confidentiality of materials uploaded to AnatomyGPT
7. Daniel et al (2026)	India	Medicine	Longitudinal, prospective non-randomized delayed intervention	to evaluate supervised ChatGPT-assisted active recall of previously learned material and its	As a cognitive scaffold to provide feedback & refine reasoning in real time. Improved self-awareness by prompting metacognitive	Both lower and high-achieving students benefited from the intervention, but the effect was more pronounced in the lower-achieving-learner group. Student feedback reflected high satisfaction, strong cognitive	Over-reliance, fears about reduced independent thinking or exposure to AI-generated inaccuracies.

Study	Country	Discipline	Study design	Application of GenAI	Practical application	Key findings	Concerns
				durability over time	reflection and boosting confidence.	engagement, and increased motivation.	Challenges with the clarity and flow of AI–student interactions.
8. Elsamanoudy (2026)	Saudi Arabia	Medicine	quasi-experimental mixed-methods educational evaluation (non-randomized, pragmatic intervention	to assess the effectiveness and usability of a customized AI-powered tutoring system, GPT Reproductive Physiology Companion, among third-year medical students enrolled in the Reproduction module.	As a safe and supportive supplement to traditional medical education, rather than replacing established instructional practices.	High levels of student satisfaction across various domains, particularly in conceptual understanding, formative assessment, and critical thinking support. Students consistently reported that the tool was user-friendly, aligned with curricular content, and contributed positively to their learning experience, regardless of gender, academic background, or GPA.	
9. Gao et al. (2026)	China	Medicine	Prospective, randomized controlled trial	As a conversational simulated patient using an AI-based web-based interface	Improve students' history-taking performance, particularly in completeness and organizational efficiency gathering.	The AI group achieved higher checklist completion, missed fewer key items, and obtained higher total OSCE scores than controls, with large effect sizes. They also performed better in medical interviewing and organization/efficiency, No significant differences were found in professionalism, clinical judgment, counselling skills, or overall clinical competence.	Privacy concerns

Study	Country	Discipline	Study design	Application of GenAI	Practical application	Key findings	Concerns
10. Holderried et al. (2024)	Germany	Medical education	Prospective observational	ChatGPT-simulated patient history-taking with automated feedback on history-taking skills of medical students	Automated grading, assessment generation, and feedback	The role-play and responses simulated by GPT-4 chatbot was medically plausible in more than 99% of cases, demonstrating effectiveness in providing comprehensive feedback on history-taking dialogs.	Unexpected behaviors, escape prompts, and increasing reliance on GenAI by students were raised.
11. Huang et al. (2025)	China	Dentistry	Randomized Controlled Trial	Video learning with ChatGPT-3.5 for skill acquisition	Dental skill learning	The ChatGPT group showed better performance ($p<0.01$), greater self-efficacy ($p=.04$) and motivation ($p=.02$), and lower cognitive load ($p<0.01$) than controls. Benefits were greatest among learners with low spatial ability.	
12. Laye & Wells (2024)	USA	Medical education	Practical descriptive	Formation of balanced student groups based on academic background and demographics using ChatGPT analysis	Improvement in learning outcomes	ChatGPT created balanced groups with greater synchronicity in completing class assignments, broader group participation, all under shorter group construction time. The application addressed past issues of time-consuming group formation by human staff and the imbalance of students in terms of prior experience and medical knowledge within groups.	

Study	Country	Discipline	Study design	Application of GenAI	Practical application	Key findings	Concerns
13. Li et al. (2024)	China	Medical education	Mixed methods	Application of ChatGPT to enhance the writing of English papers by non-native medical students through automatic feedback	Automated grading, assessment generation, and feedback Student perceptions	ChatGPT-assisted instruction improved students' coherence, grammar and vocabulary, increasing scores by 4.23 points and receiving 92% positive feedback for translation and real-time correction. This suggests strong potential for supporting assignment grading and reducing workload.	Authenticity and credibility of information from ChatGPT.
14. Naamati-Schneider (2024)	Israel	Healthcare management education	Explanatory descriptive	Integration of ChatGPT to problem-based learning framework to build AI and digital literacy of students	Improvement in learning outcomes Student perceptions	Students found ChatGPT highly accessible, user-friendly, supporting autonomous learning, and time efficient. Some students perceived ChatGPT as an opportunity to improve English language skills.	Inaccuracies, difficulty in phrasing clear and specific prompts, importance of critically evaluating and cross-referencing responses with other academic sources
15. Prachnakorn et al. (2024)	Thailand	Medical education	Experimental	Application of ChatGPT to transform traditional lectures to interactive workshops to enhance scientific report-writing skills of preclinical	Improvement in learning outcomes	ChatGPT successfully assisted authors in transforming a didactic lecture to a workshop, helping instructors to summarize students' inputs and increased engagement in the workshop.	Questionable reliability of GenAI, a hybrid approach combining human intellect and AI competence recommended.

Study	Country	Discipline	Study design	Application of GenAI	Practical application	Key findings	Concerns
				medical students			
16. Robleto et al. (2024)	USA	Medical education	Prospective	Application of 'Glass AI', a platform that combined GPT-4 with clinical guidelines	Improvement in learning outcomes Student perceptions	96% of participating medical students stated Glass AI increased their confidence in making the diagnosis. 97% stated Glass AI was easy to employ and effective in helping them reach the correct diagnosis.	Patient privacy, legal liability, and biases in the GenAI algorithms, including age, socioeconomic status, gender, and race.
17. Song et al. (2025)	China	Medicine	Quasi-experimental	Integration of GenAI for interactive teaching strategies in an undergraduate nursing course	Improvement in learning outcomes	The experimental group reported significantly better higher-order thinking scores (P= 0.018) and AI literacy (P<0.001) than the control group after the course	
18. Sridharan et al. (2024)	Bahrain	Medical education (pharmacology)	Descriptive	Integration of GenAI to enhance classroom instructions and student assessment	Improvement in learning outcomes	GenAI tools were useful in helping plan instructional methods in medical education programs, generating useful specific learning outcomes (SLOs) and detailed instructions.	Extracting details from unauthenticated sources. Test items created from GenAI should be reviewed by faculty members for content validity.
19. Thomae et al. (2024)	Switzerland	Medical education	Mixed methods	Integrating ChatGPT into a blended learning course, with student evaluation.	Improvement in learning outcomes and student perceptions	The integration yielded high satisfaction scores. Students found extended prompts generated information of higher quality, flagging the importance of students' ability to write specific and refined prompts.	Data protection issues

Study	Country	Discipline	Study design	Application of GenAI	Practical application	Key findings	Concerns
20. Wu et al. (2024)	China	Medical education	Controlled quasi-experimental	Integration of ChatGPT into clinical teaching of hepatobiliary surgery.	Improvement in learning outcomes	Improved students' understanding of theoretical knowledge and clinical skills. The experimental group obtained significantly higher final theoretical assessment scores than the control group	

Categories of Findings

Across the studies, GenAI was predominantly used to support formative assessment through automated feedback, evaluation of clinical reasoning, as a companion chatbot, and assessment of communication and procedural skills. As derived from the extracted data in Table 3, the practical applications of GenAI will be presented in three broad categories:

1. Applications of GenAI in Tertiary Health Education
2. Perceptions of GenAI use in health education
3. Ethical and Responsible Use of GenAI

Applications of GenAI in Tertiary Health Education

Application in Assessment

In terms of the assessment-related GenAI applications employed, GenAI was applied in seven studies for automated grading, assessment generation, and feedback (Burke et al., 2024; Cheung et al., 2023; Çiçek et al., 2024; Daniel et al., 2026; Elsamanoudy et al., 2026; Holderried et al., 2024; Li et al., 2024). Two studies (Burke et al., 2024; Holderried et al., 2024) employed ChatGPT to assess free-text clinical notes or simulate patient scenarios with immediate feedback provided; both studies reported high accuracy and reliability of the GenAI tool, with a key benefit being real-time detailed feedback generated by GenAI to students. One study utilized ChatGPT to automatically score and assist non-native medical students in writing English papers, where immediate feedback from ChatGPT led to improved student scores and writing quality (Li et al., 2024). The study by Cheung et al. (2023) utilized ChatGPT to generate MCQs for medical examinations, finding the tool to be highly time-efficient and capable of generating MCQs of comparable quality to human examiners. Two studies highlighted that automated grading reduced staff workload, freeing up time for more personalized engagement with students (Cheung et al., 2023; Li et al., 2024). Çiçek et al. (2024) demonstrated that ChatGPT-generated feedback improved students' clinical reasoning performance, although expert feedback remained superior in providing nuanced clinical judgement. GenAI also enabled immediate, iterative feedback that promoted self-reflection and independent learning, as reported by Daniel et al. (2026), when integrating ChatGPT to evaluate training in retrieval skills. Overall, these studies suggest that through automated grading and feedback, GenAI can enhance educators' productivity and student assessment experiences.

Application in Instruction and Teaching

The primary educational application across thirteen studies was GenAI functioning as an interactive tutor or learning companion. ChatGPT supported communication skills training (Chiu et al., 2025), knowledge retrieval tutorials (Daniel et al., 2026), physiology education through a customized AI tutor (Elsamanoudy et al., 2026), simulated patient interactions for gynecological history taking (Gao et al., 2026), clinical skills instruction in dental education (Huang et al., 2025), clinical reasoning (Çiçek et al., 2024), and development of higher-order thinking in nursing education (Song et al., 2025). These conversational AI systems provided personalised explanations, immediate feedback, unlimited opportunities for practice, and flexible access beyond classroom hours. Li et al. (2024) in their study reported improved students' coherence, grammar and vocabulary through ChatGPT-assisted instruction for English as second language

students. Students valued the ability to engage repeatedly with realistic clinical scenarios in a psychologically safe environment without fear of judgement.

Among seven of the included articles, there was a consistent trend of using GenAI tools to enhance engagement, collaboration, and critical thinking of students in tertiary health education courses (Chang et al., 2024; Collins et al., 2024; Laye & Wells, 2024; Prachnakorn et al., 2024; Sridharan et al., 2024; Thomae et al., 2024; Wu et al., 2024). Two studies described the transformation of traditional classroom instructions into more engaging, student-centered formats, with GenAI supporting brainstorming and planning for classes (Prachnakorn et al., 2024; Sridharan et al., 2024). The study by Laye & Wells, (2024) explored the use of GenAI to form balanced student groups, reporting shorter group construction time and broader group participation. Another study utilized ChatGPT to develop a customized GenAI platform, which provided anatomy-specific education responses to students (Collins et al., 2024). Overall, these studies suggest that integrating GenAI into course designs can lead to enhancement of learning experiences and student improvements.

Application in Course Design

Several studies demonstrated that GenAI could support curriculum development through incorporating theoretical frameworks, creation of authentic learning activities, simulated patients, structured tutorials, and adaptive learning resources. Gao et al. (2026) illustrated how GenAI can efficiently create simulated patients for history-taking exercises, while Elsamanoudy et al. (2026) embedded a customized ChatGPT physiology companion into a reproductive medicine module. Daniel et al. (2026) integrated AI-assisted knowledge retrieval sessions into existing curricula. In nursing education, Song et al. (2025) incorporated GenAI into learning activities that assessed higher-order thinking, demonstrating improvements in students' analytical and problem-solving abilities alongside AI literacy. Five studies integrated GenAI into teaching units and learning designs (Chang et al., 2024; Naamati-Schneider 2024; Robleto et al., 2024; Thomae et al., 2024; Wu et al., 2024) and reported significantly higher student learning improvement in terms of critical thinking and clinical skills. These findings indicate that GenAI can enhance curriculum scalability and reduce faculty workload when educational resources are carefully designed and reviewed by educators.

Perceptions of GenAI Use in Health Education

Perceived Benefits

A consistent finding across all studies was the positive perception of GenAI as a learning support tool. Students reported that AI enhanced engagement, motivation, confidence, and learning satisfaction. Personalised feedback and real-time responses allowed learners to progress at their own pace and revisit concepts as needed (Daniel et al., 2026; Elsamanoudy et al., 2026). GenAI also promoted active learning by encouraging questioning, reflection, and self-directed knowledge construction. Studies involving simulated patients (Gao et al., 2026) and communication skills (Chiu et al., 2025) demonstrated that AI provided safe environments for repeated practice of interpersonal skills before clinical placement. In nursing education, Song et al. (2025) further showed that AI use improved higher-order thinking skills and AI literacy, preparing students for technology-enhanced healthcare environments. Faculty similarly recognized AI’s potential to increase efficiency, expand learning opportunities, and provide scalable educational support. Four studies focused largely on student perceptions (Li et al., 2024; Naamati-Schneider, 2024; Robleto et al., 2024; Thomae et al., 2024). In all four studies, students

found ChatGPT and other GenAI tools highly accessible and user-friendly, supporting autonomous learning (Li et al., 2024; Naamati-Schneider, 2024; Robleto et al., 2024; Thomae et al., 2024). Li et al. (2024) reported 92% positive student feedback on ChatGPT as a translation and grammar correction tool; while 96% of participating medical students in the study by Robleto et al. (2024) found that GenAI increased their confidence in making diagnoses.

Perceived Risks and Barriers

Despite widespread optimism, all studies identified important limitations. The most reported concern was the potential for inaccurate or fabricated information (hallucinations), requiring students to critically evaluate AI-generated responses rather than accepting them uncritically. The study by Naamati-Schneider (2024) reported that students were sometimes surprised by the inaccuracies of ChatGPT, indirectly reminding them of the importance of critically evaluating and cross-referencing GenAI information with other academic sources. Thomae et al. (2024) added that student satisfaction depended on their ability to write specific and well-structured prompts. Overall, the studies (Li et al., 2024; Naamati-Schneider, 2024; Robleto et al., 2024; Thomae et al., 2024) suggest that students are mostly open to integrating GenAI into their learning but require foundational training in prompt literacy and appraisal skills to utilize GenAI tools effectively. Participants also expressed concerns regarding overreliance on AI, reduced independent critical thinking, and the possibility that students might substitute AI assistance for genuine learning (Daniel et al., 2026). Technical challenges, inconsistent response quality, limited contextual understanding, and occasional superficial explanations were additional barriers reported across the studies. All studies generally concluded that AI should augment rather than replace educator expertise.

Ethical and Responsible Use of AI

Despite the overall positive outlook, concerns of potential drawbacks were reported in 80% (n=16) studies, spanning across ethical, educational, and practical domains. Privacy considerations were particularly relevant for studies involving simulated clinical interactions. Gao et al. (2026) and Chiu et al. (2025) highlighted the importance of avoiding identifiable patient information when interacting with publicly available AI systems. Institutional governance, secure AI platforms, and adherence to ethical standards were considered necessary to protect confidentiality during educational activities. Three further studies raised concerns about data privacy and confidentiality, particularly when uploading sensitive academic or patient information (Collins et al., 2024; Robleto et al., 2024; Thomae et al., 2024).

Concerns about bias in training data and misinformation were flagged in six studies (Cheung et al., 2023; Holderried et al., 2024; Li et al., 2024; Prachnakorn et al., 2024; Robleto et al., 2024; Sridharan et al., 2024). Two studies highlighted potential student overreliance on GenAI for answers, reducing independent reasoning and content verification (Holderried et al., 2024; Li et al., 2024). In addition, three studies reported that students often struggled to write effective prompts, indicating the need for better instructions in GenAI literacy and structuring prompts (Li et al., 2024; Naamati-Schneider, 2024; Thomae et al., 2024).

Across all studies, accountability remained firmly with educators rather than AI systems. Faculty retained responsibility for ensuring the accuracy of educational materials, maintaining assessment integrity, monitoring student learning, and validating AI-generated feedback. Rather than replacing teachers, AI was consistently positioned as a supportive educational partner

whose outputs require human oversight, professional judgement, and continuous quality assurance.

Discussion

This scoping review demonstrates rapidly growing evidence that GenAI can enhance teaching, learning, and assessment across health professions education. Across the twenty included studies, GenAI, specifically ChatGPT, was most implemented as an intelligent tutor, simulated patient, personalised learning assistant, and feedback provider. These applications supported communication skills, clinical reasoning, knowledge acquisition, procedural training, and self-directed learning. Students generally perceived GenAI as engaging and accessible, while educators recognized its potential to improve educational efficiency, personalised feedback, and curriculum scalability. Despite these benefits, the studies consistently emphasized that GenAI should complement rather than replace educators, with human oversight remaining essential to ensure educational quality, clinical accuracy, and responsible implementation.

These findings align with previous reviews that advocate institution-wide strategies for GenAI adoption, including faculty development, AI literacy, governance frameworks, and clear academic integrity policies (Koontz, 2025; Cordero et al., 2024). Collectively, the evidence suggests that the educational value of GenAI depends less on the technology itself than on intentional instructional design. Rather than embedding AI as an isolated innovation, educators should integrate it within structured learning activities that promote reflection, critical thinking, authentic clinical reasoning, and real-world application (Crompton et al., 2026).

An important finding is that the evidence base is in its infancy. Although the search strategy included studies from 2021 onwards, all eligible publications were published after 2023, reflecting the rapid emergence of ChatGPT and related large language models. Consequently, much of the current literature focuses on feasibility, learner perceptions, and early implementation rather than rigorous evaluation of learning outcomes, clinical competence, behavioral change, or patient outcomes. There remains a need for longitudinal, experimental, and multi-institutional research to determine the sustained educational effectiveness of GenAI and identify best practices for curriculum integration (Fui-Hoon Nah et al., 2023).

Another notable finding is the dominance of medical education within the literature, with comparatively limited research conducted in allied health disciplines. This imbalance likely reflects the highly regulated nature of allied health education, where curriculum innovation must align with accreditation standards, competency frameworks, and patient safety requirements. As AI technologies become increasingly integrated into healthcare delivery, expanding research across allied health professions represents an important priority to ensure graduates develop AI capabilities appropriate to their clinical contexts (Crompton et al., 2026).

The above gaps are consistent with the review study on cross sector analysis by Eddine et al. (2025). They reported a disparity in adoption of ChatGPT across various disciplines, with high usage reported across Computer science and social sciences and relatively lower in health disciplines (Eddine et al., 2025). Of the health disciplines, medical education dominated in their results as well (Eddine et al., 2025). Atif et al. (2026), in their recent study, provide the first systematic, national-level analysis of Australian university alignment with TEQSA's guiding

principles for GenAI in assessment, offering a baseline against which future policy developments can be measured. They evaluated discipline-specific focus on information and communications technology education and identified a gap that GenAI guidance for technical assessments, is largely absent across Australian universities (Atif et al., 2026).

One of the most significant contributions of this review is the identification of AI literacy as an emerging core competency. While earlier literature has largely focused on the benefits and risks of GenAI, relatively little attention has been given to preparing students and educators to use these technologies critically and responsibly. The findings suggest that AI literacy extends beyond technical proficiency to include understanding AI limitations, recognizing bias and hallucinations, evaluating the reliability of AI-generated information, and applying ethical judgement within clinical decision-making. Frameworks such as the CSSD AI Literacy Framework and recent recommendations for assessment redesign (Al-Shawee et al., 2026; Atif et al., 2026; Cordero et al., 2024; Koontz, 2025) reinforce the need for AI literacy to become embedded throughout health professions curricula rather than taught as a standalone topic.

The review also reinforces ongoing concerns regarding academic integrity, bias, transparency, privacy, and overreliance on AI-generated content. Several studies highlighted the potential for AI systems to perpetuate biases present in training data, influencing educational resources, simulated patients, and clinical recommendations. These findings underscore the importance of institutional governance, transparent AI use, educator validation of AI-generated outputs, and explicit teaching that AI outputs require critical appraisal rather than unquestioning acceptance.

Implications

The findings have several important implications for health professions education. As AI becomes increasingly integrated into clinical practice, graduates must develop the knowledge, skills, and judgement required to work effectively alongside AI technologies. Consequently, curricula should move beyond teaching students how to use AI tools and instead focus on developing AI literacy, critical appraisal, ethical reasoning, and responsible clinical decision-making.

Successful implementation requires a whole-of-institution approach that includes curriculum redesign, faculty development, governance frameworks, and assessment reform. AI should be embedded within authentic learning experiences where students critically evaluate AI-generated outputs, compare them with evidence-based practice, and reflect on their strengths and limitations. Faculty development is equally important, enabling educators to design AI-enabled learning activities, develop AI-resilient assessments, and model responsible AI use.

For assessment, the findings support a hybrid approach in which GenAI is primarily used to enhance formative learning through personalised feedback, simulated clinical interactions, and deliberate practice, while educators retain responsibility for summative assessment, quality assurance, and professional judgement. Future assessment design should increasingly evaluate students' ability to critique, verify, and appropriately integrate AI-generated information rather than simply reproduce knowledge.

From a research perspective, future studies should prioritize robust evaluations of long-term educational outcomes, graduate preparedness, clinical competence, and patient-related outcomes. Greater attention is also needed to allied health disciplines, where evidence remains limited despite the growing integration of AI into clinical practice. Developing and evaluating accreditation-informed models for AI integration will be essential to ensure innovation occurs alongside patient safety, professional standards, and public trust.

Overall, this review suggests that the future of GenAI in health professions education lies not in replacing educators, but in augmenting learning through carefully designed, ethically governed, and pedagogically informed integration that prepares graduates for AI-enabled healthcare systems.

Limitations

As with any scoping review, this study has certain limitations that should be considered. Firstly, there is a risk of selection bias as the search strategy is limited to English literature only. It is possible that the database searches did not obtain all relevant literature as set by the eligibility criteria. Secondly, the search was initially limited to the period between 2021 and October 2024, meaning that some literature regarding the development of GenAI prior to the release of ChatGPT might have been excluded. This was mitigated to some extent by a second search on limited databases in July 2026. However, due to the limiting nature of the search, relevant current articles in other databases may have been omitted, while the most recent developments of GenAI, particularly the emergence of the DeepSeek GenAI model, were not captured. The rapidly evolving nature of GenAI technologies means that the findings may not have fully reflected the capabilities of the latest GenAI tools. Thirdly, the data extraction may have involved a degree of interpretive variability. Although the themes were developed through review and consensus among the research team, decisions regarding how findings were grouped may have been influenced by researcher judgement. As such, alternative categories may have been possible, particularly given the overlap between GenAI applications, perceived benefits, risks, and ethical considerations.

Collectively, the studies in this review suggest that GenAI is most valuable as a complementary educational technology that enhances personalised learning, formative assessment, and curriculum development. Its successful integration into health professions education requires robust governance, educator oversight, AI literacy training, and ethical safeguards to maximize educational benefits while mitigating risks associated with accuracy, bias, privacy, and inappropriate dependence on AI. Rather than drawing definitive conclusions about the effectiveness of GenAI in tertiary health education, this review illuminated where research currently stands and where it should go. Most studies were exploratory and focused on short-term implementations, leaving questions about long-term effects of GenAI on student learning and the curriculum. As this field is still in an experimental phase, this review highlighted the importance of directing future research into long-term evaluations, particularly in terms of designing AI literacy training modules and ethical guidelines.

Conclusion

This scoping review mapped the current literature on the integration of specific GenAI applications into tertiary health education. The findings demonstrated that GenAI is being rapidly integrated into tertiary health education as a supportive tool for automated grading and feedback, as well as enhanced learning outcomes and improvements in student engagement and critical thinking. Although the majority of GenAI applications employed were well-received, concerns of reliability, ethical use, and over-reliance were often cited. A key knowledge gap was identified in the lack of AI literacy training within current tertiary health education programs, which could risk underpreparing graduates. As the influence of GenAI continues to grow in the tertiary health education sector, it is imperative for educational institutions to develop AI literacy training workshops and ethical guidelines. AI literacy training modules should feature prompt literacy, informed skepticism on the quality of GenAI responses and awareness of academic integrity. Furthermore, longitudinal research on the long-term impacts of GenAI and the specific design of AI literacy training workshops is recommended.

Declarations

Ethics Statement

Due to the nature of this review, ethical approval and informed consent were not required. The project did not have any major risks as it did not involve human research and only used literature that had already been published (National Health and Medical Research Council, 2018).

Funding Statement

This research received no external funding.

AI Use Statement

Generative artificial intelligence was used for grammar and spelling using the free version of Grammarly. Generative artificial intelligence was not used to generate, analyze, or interpret research data, nor to create original scholarly content. All content was reviewed, verified, and approved by the authors, who take full responsibility for the accuracy and integrity of the work.

Conflicts of Interest Statement

The authors have no conflicts of interest related to the research, authorship, or publication of this article.

References

Alam, F., Lim, M. A., & Zulkipli, I. N. (2023). Integrating AI in medical education: Embracing ethical usage and critical understanding. *Frontiers in Medicine*, 10. https://doi.org/10.3389/fmed.2023.1279707

Allied Health Professions Australia. (n.d., first accessed May 9, 2023). *Allied health professions*. https://ahpa.com.au/allied-health-professions/

Al-Shawee, N., McElvaney, G., Strawbridge, J., & Spooner, M. (2026). When assessment theory meets generative AI: Reimagining SBA design in medical education. Perspectives on *Medical Education*, 15(1), 239–250. https://doi.org/10.5334/pme.2033

Atif, A., Jha, M., Richards, D., & Cibin, A. (2026). Institutional alignment with TEQSA's generative AI principles in Australian universities: A national document analysis with implications for ICT education. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.11001

Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in Higher education: A systematic review of Case studies. *Computers and Education: Artificial Intelligence*, 8, 100407. https://doi.org/10.1016/j.caeai.2025.100407

Benmamoun, M. (2023). Reinventing international business education: integrating the power of generative AI. *AIB Insights*. https://doi.org/10.46697/001c.90397

Boscardin, C., Gin, B., Golde, P. B., & Hauer, K. E. (2023). ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. *Academic Medicine*, 99(1). https://doi.org/10.1097/acm.0000000000005439

Burke, H. B., Hoang, A., Lopreiato, J. O., King, H., Hemmer, P., Montgomery, M., & Gagarin, V. (2025). Assessing the ability of a large language model to score free-text medical student clinical notes: quantitative study. *JMIR Medical Education*, 10, e56342. https://doi.org/10.2196/56342

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *Internation Journal of Educational Technology in Higher Education*, 20(1). https://doi.org/10.1186/s41239-023-00411-8

Chang, C.-Y., Yang, C.-L., Jen, H.-J., Ogata, H., & Hwang, G.-H. (2024). Facilitating nursing and health education by incorporating ChatGPT into learning designs. *Educational Technology and Society*, 215–230. https://doi.org/10.30191/ETS.202401_27(1).TP02

Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., Wong, R., & Co, M. T. (2023). ChatGPT versus human in generating medical graduate exam multiple choice questions-A multinational prospective study (Hong Kong S.A.R.,

Singapore, Ireland, and the United Kingdom). *PloS one*, 18(8), e0290691. https://doi.org/10.1371/journal.pone.0290691

Cheung, L. T., Maranna, S., Parange, N., Osborne, B., & Vu, H. (2025, December 22). Scoping review search strategy and data extraction: The Integration of Generative Artificial Intelligence into Tertiary Health Education. https://doi.org/10.17605/OSF.IO/GD3PV

Chiu, J., Castro, B., Ballard, I., Nelson, K., Zarutskie, P., Olaiya, O. K., Song, D., & Zhao, Y. (2025). Exploration of the role of ChatGPT in teaching communication skills for medical students: A pilot study. *Medical Science Educator*. https://doi.org/10.1007/s40670-025-02394-9

Çiçek, F. E., Ülker, M., Özer, M., & Kıyak, Y. S. (2024). ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: A randomized controlled trial. *Postgraduate Medical Journal*. https://doi.org/10.1093/postmj/qgae170

Collins, B. R., Black, E. W., & Rarey, K. E. (2024). Introducing anatomyGPT: a customized artificial intelligence application for anatomical sciences education. *Clinical Anatomy.* https://doi.org/10.1002/ca.24178

Cordero, J., Torres-Zambrano, J., & Cordero-Catillo, A. (2024). Integration of generative artificail intelligence in higher education: Best practices. *Education Sciences*, 15(1), 32-32. https://doi.org/10.3390/educsci15010032

Covidence systematic review software. (2024). Veritas Health Innovation. https://www.covidence.org. 2024

Crompton, H., Burke, D., Nickel, C., Bozkurt, A., Miao, F., Sharples, M., Greene, J. A., Parsons, D., Gill-Simmen, L., Edmett, A., Pegrum, M., de Waard, I., Bonk, C. J., Garcia, M. B., Curry, J. H., Lindsey, L., Yang, M., Marshall, S., Bali, M., … Yu, S. (2026). Governing generative AI in higher education: A global delphi study on policy and practice. *International Journal of Educational Technology in Higher Education*, 23(1). https://doi.org/10.1186/s41239-026-00602-z

Daniel, K., Msambwa, M. M., & Wen, Z. (2025). Can generative AI revolutionise academic skills development in higher education? A systematic literature review. *European Journal of Education*, 60(1). https://doi.org/10.1111/ejed.70036

Daniel, A. R., Nivethitha, V., Surya, B., Antoinette, D., & Visalakshi, R. (2026). Integrating ChatGPT into knowledge-retrieval tutorials in undergraduate medical education: A prospective evaluation of higher-order learning and feasibility. *Medical Education Online*, 31(1). https://doi.org/10.1080/10872981.2026.2639203

Eddine, R. J., Gide, E., & Al-Sabbagh, A. (2025). Generative AI in higher education: A cross-sector analysis of ChatGPT's impact on STEM, social sciences, and healthcare. *STEM Education*, 5(5), 757–801. https://doi.org/10.3934/steme.2025035

Elsamanoudy, A., Alhazimi, A., Hassanien, M., Almarabheh, A., Al-Aqidi, S., Alwalidi, R., & Alrefaie, Z. (2026). Implementation and evaluation of customized artificial intelligence personal tutor (chat GPT physiology companion) for medical students in the reproduction module. *BMC Medical Education*. https://doi.org/10.1186/s12909-026-09088-9

Farazouli, A., Pargman, T. C., Laksov, K. B., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment and Evaluation in Higher Education*, 49(3), 1–13. https://doi.org/10.1080/02602938.2023.2241676

Fui-Hoon Nah F. Zheng R, Cai J, Siau K, Chen L. (2023). Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. *Journal of information technology case and application research,* 25(3):277-304. https://doi.org/10.1080/15228053.2023.2233814

Gao, L., Zhang, L., Zhang, Y., Wu, M., Xu, J., & Lu, D. (2026). Utilizing generative artificial intelligence to create simulated patient for history taking in gynecology. *BMC Medical Education*, 26(1). https://doi.org/10.1186/s12909-026-09219-2

Gill, S., & Galletly OAM, C. (2025). 'I, robot, can help you.' applications of generative artificial intelligence in RANZCP psychiatry training. *Australasian Psychiatry*, 33(4), 689–692. https://doi.org/10.1177/10398562251344470

Glauberman, G., Ito-Fujita, A., Katz, S., & Callahan, J. (2023). Artificial intelligence in nursing education: opportunities and challenges. *Hawai'i Journal of Health and Social Welfare,* 82(12), 302–305. https://www.ncbi.nlm.nih.gov/pubmed/38093763

Gupta, B. L., Mufti, T., Sohail, S. S., & Madsen, D. Ø. (2023). ChatGPT: A brief narrative review. *Cogent Business & Management*, 10(3). https://doi.org/10.1080/23311975.2023.2275851

Hardie, P., Darley, A., Derwin, R., Eustace-Cook, J., Kearns, S., Mc Brien, B., Siddiquee, A., Zheng, D., & Mooney, M. (2026). Applications, attitudes and ethical considerations of Generative Artificial Intelligence (Gen AI) in nursing education: a scoping review. *BMC Nursing*, 25(1), 148. https://doi.org/10.1186/s12912-025-04253-9

Hinsche, L., Hasseler, M., Tischendorf, T., & Schaal, T. (2026). Rapid review on GenAI in nursing education. *Frontiers in Health Services, 5*, 1725425. https://doi.org/10.3389/frhs.2025.1725425

Holderried, F., Stegemann-Philipps, C., Herrmann-Werner, A., Festl-Wietek, T., Holderried, M., Eickhoff, C., & Mahling, M. (2024). A language model-powered simulated patient with automated feedback for history taking: Prospective study (Preprint). *JMIR Medical Education*, 10, e59213–e59213. https://doi.org/10.2196/59213

Huang, S., Wen, C., Bai, X., Li, S., Wang, S., Wang, X., & Yang, D. (2025). Exploring the application capability of ChatGPT as an instructor in Skills education for Dental medical

students: Randomized controlled trial. *Journal of Medical Internet Research*, 27, e68538–e68538. https://doi.org/10.2196/68538

Koh, E., & Doroudi, S. (2023). Learning, teaching, and assessment with generative artificial intelligence: towards a plateau of productivity. *Learning*, 9(2), 1–8. https://doi.org/10.1080/23735082.2023.2264086

Koontz, M., & Podlog, S. (2025). Integrating generative artificial intelligence in midwifery education: Balancing innovation, ethics, and academic integrity. *Journal of Midwifery & Women's Health*. https://doi.org/10.1111/jmwh.70021

Kotzki, S., Massonnet Turner, C., Gauthier, K., Minoves, M., & Vuillerme, N. (2026). Health professional students' use of generative artificial intelligence during clinical placements: Cross-Sectional online survey study. *JMIR Medical Education*, 12, e85243–e85243. https://doi.org/10.2196/85243

Kovalainen, T., Pramila-Savukoski, S., Kuivila, H.-M., Juntunen, J., Jarva, E., Rasi, M., & Mikkonen, K. (2025). Utilising artificial intelligence in developing education of health sciences higher education: An umbrella review of reviews. *Nurse Education Today*, 147, 106600. https://doi.org/10.1016/j.nedt.2025.106600

Laye, M. J., & Wells, M. B. (2024). Rapid creation of knowledge-balanced student groups using ChatGPT4. *Medical Science Educator*, 34(3), 523–525. https://doi.org/10.1007/s40670-024-02032-w

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: a study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6(100221), 100221. https://doi.org/10.1016/j.caeai.2024.100221

Li, J., Zong, H., Wu, E., Wu, R., Peng, Z., Zhao, J., Yang, L., Xie, H., & Shen, B. (2024). Exploring the potential of artificial intelligence to enhance the writing of English academic papers by non-native English-speaking medical students - the educational application of ChatGPT. *BMC Medical Education*, 24(1). https://doi.org/10.1186/s12909-024-05738-y

Lin, Y., Luo, Z., Ye, Z., Zhong, N., Zhao, L., Zhang, L., Li, X., Chen, Z., & Chen, Y. (2025). Applications, challenges and prospects of generative artificial intelligence empowering medical education: A scoping review (preprint). *JMIR Medical Education*. https://doi.org/10.2196/71125

Lye, C. Y., & Lim, L. (2024). Generative artificial intelligence in tertiary education: assessment redesign principles and considerations. *Education Sciences*, 14(6), 569. https://doi.org/10.3390/educsci14060569

Naamati-Schneider, L. (2024). Enhancing AI competence in health management: students' experiences with ChatGPT as a learning Tool. *BMC Medical Education*, 24(1). https://doi.org/10.1186/s12909-024-05595-9

Nah, F. F.-H., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

NHMRC. (2018). *National Statement on Ethical Conduct in Human Research* (2007) - Updated 2018 | NHMRC. Www.nhmrc.gov.au; National Health and Medical Research Council. https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018#block-views-block-file-attachments-content-block-1

Peters M.D.J., Godfrey, C., Mclnerney, P., Munn, Z., Tricco, A., & Khalil, H. Chapter 10: scoping reviews (2020 version). In Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. editors, *JBI Evidence Synthesis*. 2020 JBI. https://doi.org/10.46658/JBIMES-24-09

Pham, T. D., Karunaratne, N., Exintaris, B., Liu, D., Lay, T., Yuriev, E., & Lim, A. (2025). The impact of generative AI on health professional education: A systematic review in the context of student learning. *Medical Educa*tion. https://doi.org/10.1111/medu.15746

Prachnakorn, N., Preecha, K., Sri-u-thai, T., Jaroenyod, T., Sawang, K., Patwong, N., & Wattanapisit, A. (2024). Incorporating artificial intelligence into a workshop on scientific and scholarly report writing for preclinical medical students. *Medical Teach*er, 46(6), 746–748. https://doi.org/10.1080/0142159x.2024.2311267

Preiksaitis, C., & Rose, C. (2023). Generative artificial intelligence in medical education: opportunities, challenges, and future directions – a scoping review (preprint). *JMIR Medical Education*, 9(1), e48785–e48785. https://doi.org/10.2196/48785

Robleto, E., Habashi, A., Kaplan, M.-A. B., Riley, R. L., Zhang, C., Bianchi, L., & Shehadeh, L. A. (2024). Medical students' perceptions of an artificial intelligence (AI) assisted diagnosing program. *Medical Teacher*, 1–7. https://doi.org/10.1080/0142159X.2024.2305369

Shoja, M. M., Van de Ridder, J. M. M., & Rajput, V. (2023). The emerging role of generative artificial intelligence in medical education, research, and practice. *Cureus*, 15(6). https://doi.org/10.7759/cureus.40883

Song, D., Zhang, P., Zhu, Y., Qi, S., Yang, Y., Gong, L., & Zhou, L. (2025). Effects of Generative artificial intelligence on Higher-Order thinking skills and artificial intelligence literacy in Nursing undergraduates: A Quasi-Experimental study. *Nurse Education in Practice*, 88, 104549–104549. https://doi.org/10.1016/j.nepr.2025.104549

Sridharan, K., & Sequeira, R. P. (2024). Artificial intelligence and medical education: application in classroom instruction and student assessment using a pharmacology & therapeutics case study. *BMC Medical Education*, 24(1). https://doi.org/10.1186/s12909-024-05365-7

Su, J., & Yang, W. (2023). Unlocking the power of chatgpt: a framework for applying generative AI in education. *ECNU Review of Education*, 6(3), 355–366. https://doi.org/10.1177/20965311231168423

Takhdat, K., El fadely, A., Mohamed, E., Ouaamr, A., & El Adib, A. R. (2026). A systematic review of generative artificial intelligence-powered healthcare simulation for clinical reasoning skills development: Applications, outcomes and challenges. *Medical Science Educator*. https://doi.org/10.1007/s40670-026-02804-6

Thomae, A. V., Witt, C. M., & Barth, J. (2024). Integration of ChatGPT Into a course for medical students: explorative study on teaching scenarios, students' perception, and applications. *JMIR Medical Education*, 10, e50545–e50545. https://doi.org/10.2196/50545

Toohey, K., Quince, Z., Walker, F., Furness, L., Bissett, M., Daley, C., Allen, K., Munro, N., Smidt, A., Wilkie, J. C., Horstmanshof, L., Baltrotsky, K., & Naumann, F. (2025). Integrating generative AI in Health education: A scoping review and implementation framework. *Medical Science Educator*. https://doi.org/10.1007/s40670-025-02578-3

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., & Lewin, S. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467–473. https://doi.org/10.7326/M18-0850

Uribe, S. E., Maldupa, I., & Schwendicke, F. (2025). Integrating Generative AI in Dental Education: A Scoping Review of Current Practices and Recommendations. *European journal of dental education: official journal of the Association for Dental Education in Europe*, 29(2), 341–355. https://doi.org/10.1111/eje.13074

Wu, C., Chen, L., Han, M., Li, Z., Yang, N., & Yu, C. (2024). Application of ChatGPT-based blended medical teaching in clinical education of hepatobiliary surgery. *Medical Teach*er, 47(3), 445–449. https://doi.org/10.1080/0142159x.2024.2339412

Xu, X., Chen, Y., & Miao, J. (2024). Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. *Journal of Educational Evaluation of Health Professions*, 21, 6–6. https://doi.org/10.3352/jeehp.2024.21.6

Appendices and Supplemental Materials

Appendix A

Search Strings for Google Scholar and SCOPUS In the Second Search of July 2026.

Database/ website	Search string
Google Scholar	Find articles with all the words: Integration of Generative artificial intelligence in tertiary health education, generative OR tertiary OR with the exact phrase: Generative artificial intelligence with at least one of the words: health education or medical education or chatgpt
SCOPUS	(TITLE-ABS-KEY ("Artificial Intelligence" OR "Generative AI" OR "Generative Artificial Intelligence" OR "Machine learning" OR chatgpt OR chatbot OR "Large Language Model")) AND (TITLE-ABS-KEY ("Health Educat*" OR "Medical Educat*" OR "Healthcare Student*" OR "Medical Student*" OR "Allied Health Student*" OR "Allied Health Educat*")) AND (TITLE-ABS-KEY ("Tertiary Educat*" OR undergrad* OR postgrad* OR "Higher education" OR universit*)) AND PUBYEAR > 2024 AND PUBYEAR < current

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5764
https://doi.org/10.24059/olj.v30iSI2.5764

Mapping Evidence on Persistence and Retention Strategies for Non-Traditional Students in Online Higher Education: A Scoping Review

Jessica Sylvester
Melinda Kulick
University of Phoenix

Abstract

Non-traditional learners, often balancing employment, caregiving, and academic responsibilities, now constitute a growing majority of students in higher education (National Center for Education Statistics, 2024). Although their persistence and retention are central to institutional effectiveness, the evidence base informing institutional responses remains fragmented and unevenly conceptualized. Guided by the Joanna Briggs Institute (JBI) framework for scoping reviews, this study synthesized empirical and conceptual literature published between 2013 and 2025 to map how persistence and retention among non-traditional students are defined, examined, and measured in online higher education (Hadie, 2024). Findings indicate a strong reliance on age- and employment-based definitions of non-traditional status, frequent use of composite outcome measures centered on persistence and institutional retention, and limited differentiation between student-level continuation and institution-level metrics. Analysis identified five interrelated thematic domains spanning instructional, learner, and institutional levels, with an emergent emphasis on technology- and AI-mediated practices. By producing a comprehensive evidence map, this review clarifies dominant patterns, methodological tendencies, and persistent gaps in the literature, offering direction for future research and institutional practice aimed at advancing educational equity and student success in online higher education.

Keywords: Non-traditional students; online higher education; persistence and retention; scoping review; conceptualization and measurement; institutional practices; adult learners

Introduction

Online higher education has expanded rapidly over the past two decades, reshaping access, participation, and institutional accountability across postsecondary education. As online programs have scaled, they have increasingly served non-traditional learners whose enrollment patterns reflect adult age, employment and caregiving responsibilities, delayed entry, and financial independence (National Center for Education Statistics, 2024). This shift has intensified institutional attention to persistence and retention as indicators of educational quality and student success (Bawa, 2016; Cordano, 2025; Stephen, 2022).

Persistence and retention are longstanding concerns in higher education research, particularly in online learning contexts where attrition has often been reported at higher rates than in face-to-face settings. Much of the literature has emphasized learner-level explanations of persistence, such as motivation, grit, or self-regulation, while giving comparatively less sustained attention to the institutional, instructional, and structural conditions shaping students' capacity to remain enrolled (Ellis, 2019; Stephen, 2022; Walsh et al., 2024). Such framing is especially limiting for non-traditional learners, whose participation is embedded within intersecting personal, professional, and financial demands extending beyond the boundaries of the learning environment.

Corresponding author: Jessica Sylvester, jessica.sylvester@phoenix.edu
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

Although research on online student success is substantial, fewer studies examine persistence and retention specifically among non-traditional online learners. Moreover, the evidence base is dispersed across disciplines and methodologies, with inconsistent definitions of adult learner populations and considerable variation in how persistence and retention-related outcomes are conceptualized and measured. As a result, it remains difficult to synthesize findings or compare conclusions across studies. During the same period, institutions have expanded technology-mediated support strategies (e.g., learning analytics, early alert systems, and AI-supported advising), yet research examining how these strategies are embedded within instructional and institutional systems supporting adult online learners remains uneven (Seery et al., 2021; Stephen, 2022).

Prior Reviews of Persistence and Retention in Online Learning

Prior reviews have established online attrition and retention as multifactorial concerns shaped by learner motivation, self-regulation, instructional design, faculty presence, institutional support, technology access, and students' external responsibilities (Bawa, 2016; Seery et al., 2021). Reviews focused on adult learners further demonstrate online participation is shaped by work and family responsibilities, time constraints, limited institutional support, insufficient interaction, technological barriers, and isolation (Kara et al., 2019). Related reviews of online academic support also identify advising, engagement, access to services, satisfaction, motivation, self-efficacy, retention, and completion as dimensions of online student success (Walsh et al., 2024).

Although these reviews provide important foundations, they examine online learners broadly or focus on barriers, supports, or learner characteristics, as opposed to centering non-traditional students in online higher education. Existing reviews tend to emphasize either student-level attributes or specific support strategies, with less attention to how instructional practices, institutional structures, learner life contexts, and technology-mediated practices interact across levels. In addition, persistence, retention, completion, and engagement are often treated as related outcomes without consistently distinguishing student-centered continuation from institutionally defined retention metrics.

Contribution of the Present Scoping Review

This scoping review responds to these limitations by centering non-traditional online learners and mapping persistence and retention across instructional, institutional, learner-context, support, psychological, and technology-mediated domains. This review extends prior work by clarifying how persistence and retention are defined, measured, and supported in relation to non-traditional students' life contexts.

By integrating this literature, the review helps disentangle learner-level explanations of persistence from course-level pedagogical practices and institution-level structures shaping persistence and retention in online programs. Accordingly, this scoping review maps how persistence and retention among non-traditional online learners are conceptualized, investigated, and supported.

Review Questions

The purpose of this scoping review is to identify and categorize existing research on strategies, interventions, and institutional practices intended to support the persistence and retention of non-traditional students in online higher education. Consistent with scoping review methodology, the review aimed to map the breadth and characteristics of the evidence base.

Primary Review Question:

1. What strategies, interventions, and institutional practices have been examined in the research literature to support the persistence and retention of non-traditional students in online higher education?

To support systematic mapping of the evidence base, the following secondary review questions guided data extraction and synthesis:

2. How does the research literature describe institutional and instructional factors associated with student persistence in online higher education?
3. How are the characteristics and life stressors of non-traditional learners described in relation to persistence and retention in online higher education?
4. How does the literature describe features of the online learning environment associated with persistence and retention for non-traditional students, including instructional design, faculty presence, peer interaction, and technological support?

Methods

Scoping Review Design

This scoping review mapped the breadth, characteristics, and scope of existing literature examining strategies, interventions, and institutional practices supporting the persistence and retention of non-traditional students in online higher education. The review was conducted in accordance with Joanna Briggs Institute (JBI) guidance for scoping reviews and is reported in alignment with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) (Hadie, 2024).

Eligibility Criteria

Eligibility criteria were developed using the Population–Concept–Context (PCC) framework.

Population

The population included non-traditional students enrolled in online higher education. Eligible studies examined student populations with one or more non-traditional characteristics, including adult age (24 years or older), delayed enrollment or return to higher education after a significant break, full-time employment, financial independence, and/or caregiving responsibilities (National Center for Education Statistics, 2024).

Concept

The concept of interest was student persistence and retention in online higher education. Included studies examined strategies, interventions, or institutional practices intended to support continued enrollment, academic progression, or program completion. Studies examining proxy outcomes (e.g., engagement, continuity, progression) were included only when these outcomes were explicitly linked to persistence or retention.

Context

The context was limited to online higher education environments. Studies focusing exclusively on face-to-face, blended, or hybrid modalities were excluded.

Types of Sources

Peer-reviewed, empirical, and conceptual studies were included. Editorials and commentaries were excluded. Additional eligibility criteria required studies to be published between 2013 and 2025, written in English, and available in full text. Table 1 presents the inclusion and exclusion criteria used to guide study selection, organized according to the PCC framework.

Table 1

Study Inclusion and Exclusion Criteria

Criteria	Inclusion Criteria	Exclusion Criteria
Time	Published between 2013 and 2025	Published before 2013 or after 2025
Population	Non-traditional students in online higher education	Traditional students or populations not identifiable as non-traditional by the stated characteristics
Concept	Studies examining persistence and retention in online higher education, including strategies, interventions, or institutional practices.	Studies focused on outcomes not linked to persistence or retention.
Context	Online higher education environments	Face-to-face, blended, or hybrid modalities
Language	English	Articles published in languages other than English
Types of Sources	Peer-reviewed empirical and conceptual studies; full text available	Editorials, commentaries, studies not available in full text

Search Strategy

A comprehensive search strategy was developed to identify literature related to persistence and retention among non-traditional students in online higher education. Searches were conducted in ERIC, Education Source, and Google Scholar. ERIC and Education Source were selected because of their focus on education research, while Google Scholar was used to

broaden retrieval across interdisciplinary higher education literature. Search terms were organized around three concept groups:

1. non-traditional or adult learners
2. online higher education
3. persistence, retention, or institutional support mechanisms

Keywords within each concept were combined using the Boolean operator OR, and concept groups were combined using AND. The complete search strategy is provided in the online Supplemental Materials, S1 Boolean Search Strategy.

Study Selection

Study selection followed a structured, multi-stage process consistent with JBI guidance for scoping reviews (Hadie, 2024). To support systematic management of a large interdisciplinary evidence base, relevance-ranking features in ERIC and Education Source were used to prioritize records most closely aligned with the search parameters. Google Scholar results were limited to the first 10 pages for feasibility and relevance-ranking considerations.

Titles and abstracts were screened independently by two reviewers against the PCC eligibility criteria. Records were excluded if they did not focus on higher education, were not situated in online learning environments, did not include non-traditional or adult learners, or were not conceptually related to persistence, retention, or institutional support strategies. Full-text screening was conducted for all records meeting inclusion criteria at the title and abstract stage. Any disagreements at either stage were resolved through discussion and consensus. Reasons for exclusion at the full-text stage were documented to support transparency and reproducibility. Microsoft Excel was used to organize records, log reviewer screening decisions, and document full-text exclusion reasons.

Data Charting

Data charting was conducted using a structured, tiered approach aligned with the review questions. A structured charting form was developed in Microsoft Excel to support consistent extraction across source types. The form was piloted on an initial subset of included sources and refined through reviewer discussion to improve clarity, consistency, and alignment with the PCC framework. Coding decisions and refinements were documented throughout the charting process.

Tier 1 charting was applied to all included sources of evidence and captured bibliographic information, year of publication, country of origin, study type, population characteristics, definitions of non-traditional learners, level of intervention or strategy focus, and persistence or retention-related outcomes. Proxy outcomes, including engagement, progression, continuity, or completion, were charted only when study authors explicitly linked them to persistence or retention.

When differences arose in the interpretation or application of charting categories, reviewers discussed the coding decision until consensus was reached. No inter-rater reliability statistics were calculated because the purpose of data charting was descriptive mapping.

Tier 2 charting was applied to empirical studies only and captured methodological characteristics relevant to descriptive mapping, including research design, data sources, and conceptual framing. Conceptual papers and literature-based reviews were charted at the Tier 1 level only, as methodological characteristics were not applicable to these source types. Tier 1 and Tier 2 data charting tables are provided in the online Supplemental Materials, S2 Tier 1 Data, and S3 Tier 2 Data.

Data Synthesis & Reporting Standards

Extracted data were synthesized using descriptive and thematic analytic approaches. Descriptive synthesis was used to summarize publication characteristics, study designs, population definitions, and categories of institutional strategies and interventions.

A descriptive thematic analysis was conducted to identify recurring patterns in how non-traditional learners were conceptualized, how studies framed responsibility for persistence and retention (e.g., as primarily individual, instructional, or institutional), and how persistence and retention outcomes were defined and operationalized. Thematic analysis emphasized the organization and mapping of conceptual and methodological patterns across the evidence base rather than the evaluation of effectiveness of intervention. Consistent with JBI guidance for scoping reviews, no formal critical appraisal of study quality was conducted. Across study selection, data charting, and synthesis, no digital automation tools were used to determine eligibility, extract or chart data, code themes, or analyze findings.

Conceptual Framework

To interpret the mapped evidence, this review drew on two established persistence frameworks as analytic lenses: the Model of Non-traditional Student Attrition (Bean & Metzner, 1985) and the Composite Persistence Model for Online Learning (Rovai, 2003). These frameworks did not determine study inclusion but guided coding and interpretation by foregrounding (a) external life-context and institutional fit, and (b) conditions specific to online learning environments. Applying these frameworks analytically supported examination of how responsibility for persistence was framed (individual, instructional, or institutional) and how interventions and outcomes were positioned across levels.

Model of Non-traditional Student Attrition

Bean and Metzner's (1985) Model of Non-Traditional Student Attrition (MNSA) foregrounds the role of external life context in shaping persistence, including employment, financial constraints, family responsibilities, and other environmental stressors. The model conceptualizes attrition as strongly influenced by misalignment between institutional structures and learners' lived realities. In this review, the model informed analysis of whether studies accounted for external constraints when defining non-traditional learners and interpreting persistence or retention outcomes, as well as how responsibility for persistence was framed across institutional and individual levels (Ellis, 2019; Motte, 2020).

Composite Persistence Model for Online Learning

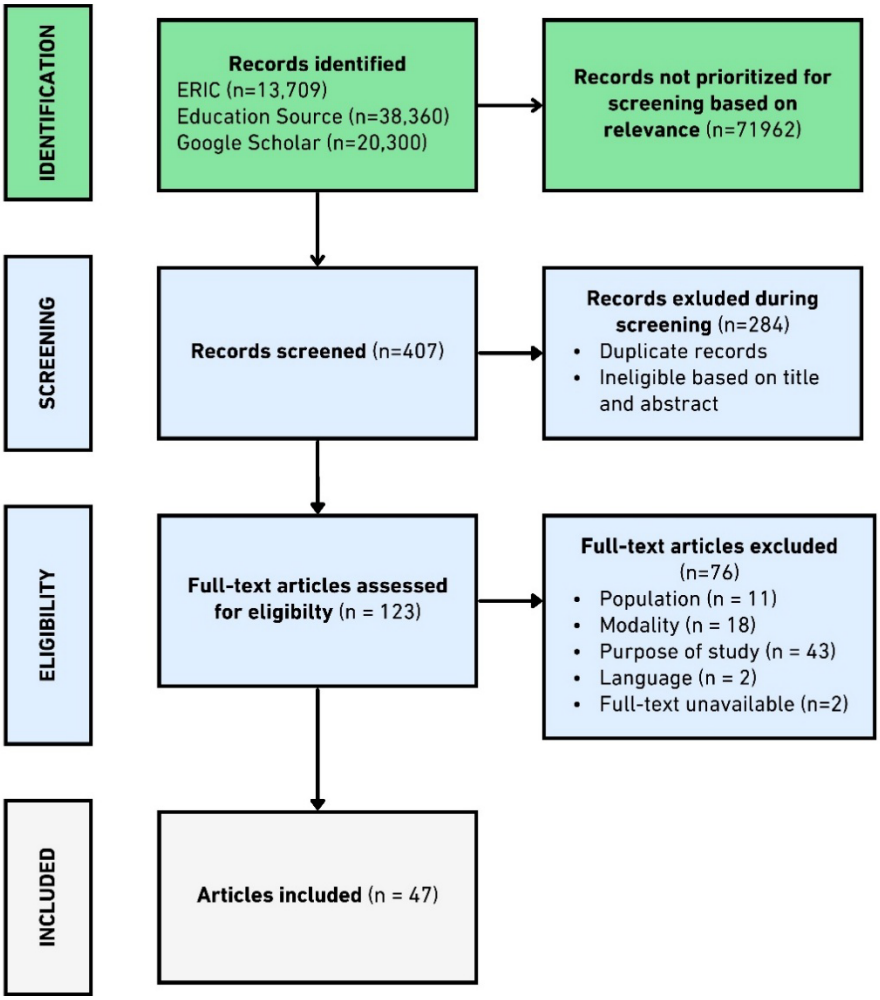
The Composite Persistence Model for Online Learning (CPMOL), developed by Rovai (2003), incorporates conditions specific to online learning environments, including academic integration, social presence, technological infrastructure, and institutional support. The model informed analysis of how instructional and institutional strategies were described across studies, particularly with respect to course design, faculty presence, peer interaction, and access to academic and student support services. The model also supported examination of how persistence and retention outcomes were defined and measured (Belt et al., 2025; Choi & Park, 2018). Integrated, these frameworks provided a complementary analytic structure for interpreting the literature by situating external life context alongside online learning design and institutional support, enabling examination of persistence as a multi-level and system-mediated outcome.

Results

The database search yielded 72,369 records (ERIC: 13,709; Education Source: 38,360; Google Scholar: 20,300). To manage retrieval volume, database relevance-ranking features were used to prioritize records for screening, resulting in 407 records retained for title and abstract screening. Of these, 284 records were excluded due to duplication or failure to meet PCC eligibility criteria. Full-text reports were retrieved for 123 articles assessed for eligibility. Following full-text review, 76 articles were excluded due to population mismatch, modality misalignment, or scope misalignment (e.g., studies focused on technical tool evaluations or conceptual frameworks without explicit application to persistence or retention outcomes). In total, 47 studies met the inclusion criteria and were included in the review. The study selection process is summarized in the PRISMA-ScR flow diagram, Figure 1.

Figure 1

PRISMA-ScR: Flow of Records Through the Scoping Review Process



Characteristics of Included Studies

The 47 included studies were published between 2013 and 2025, with volume increasing over time. Early publications were limited, followed by more consistent representation from 2018 onward, reflecting sustained scholarly attention to persistence and retention among non-traditional students in online higher education.

Most studies were conducted in the United States (U.S.) (n = 40). International contexts were less frequently represented, including Australia (n = 2), Belgium (n = 1), South Korea (n = 1), Sri Lanka (n = 1), and multinational or international samples (n = 2).

With respect to level of analysis, studies most often examined persistence and retention at the institutional level (n = 18), followed by the program level (n = 14) and course level (n = 11). Fewer studies addressed multiple levels (n = 3) or did not clearly specify a level of analysis (n = 1).

The studies included quantitative (n = 16), qualitative (n = 16), and mixed-methods (n = 7) designs; conceptual and literature-based studies accounted for a smaller subset of the evidence base (n = 8). Quantitative studies primarily relied on institutional administrative data and survey measures, while qualitative studies commonly used interviews, focus groups, and program artifacts. Mixed-methods studies integrated multiple data sources using sequential or convergent designs. Table 2 summarizes key characteristics of the included studies, contextualizing the scope and composition of the literature. The Appendix provides visual summaries of the descriptive characteristics of the included studies, including database distribution, publication year distribution, non-traditional student descriptors, and reported outcome domain.

Table 2

Characteristics of Included Studies

	Characteristic	Frequency (*n*)
Publication Period	2013-2016	12
	2017-2020	14
	2021-2025	21
Country	United States	40
	Australia	2
	Belgium	1
	South Korea	1
	Sri Lanka	1
	Multinational	2
Level of Analysis	Institutional level	18
	Program level	14
	Course level	11
	Multiple levels	3
	Not specified	1
Methodological Approach	Quantitative	16
	Qualitative	16
	Mixed methods	7
	Conceptual / literature-based	8

Characteristics and Definitions of Non-Traditional Students

Definitions of non-traditional students varied in explicitness but followed a largely consistent pattern across studies, most often operationalized through composite criteria rather than a single defining characteristic. While all included studies focused on non-traditional learners, the degree of definitional clarity varied, with some studies articulating formal inclusion criteria and others relying on descriptive inference. Explicit definitions were provided in a substantial subset of studies (*n* = 20), most commonly combining adult learner status (typically age 24 or older) with full- or part-time employment and frequently supplemented by caregiving responsibilities.

Across the evidence base, adult age and employment status emerged as the dominant defining characteristics of non-traditional students. The majority of studies incorporated employment status (n = 37) and adult learner status (n = 30), often treating these criteria as sufficient indicators of non-traditional status. Family or caregiving responsibilities were also commonly included (n = 19), though typically as an additional layer rather than a primary defining feature. In contrast, enrollment trajectory characteristics, such as delayed enrollment, or extended time to degree completion were referenced less frequently (n = 8). Capacity- or access-related constraints (e.g., academic readiness, digital access, language proficiency, or technology skills) appeared in only a small subset of studies (n = 6). Definitions of non-traditional students were heterogeneous and multidimensional, with substantial variation in how characteristics were combined and operationalized across studies. Table 3 summarizes the frequency and operationalization of descriptors used across studies to define non-traditional student populations.

Table 3

Frequency of Non-Traditional Student Descriptors

Descriptor	Frequency (*n*)	Operationalization
Full-time employment	37	Concurrent employment while enrolled
Adult learners (Age 24+)	30	Mature or adult learner designation
Family/caregiving responsibilities	19	Childcare or dependent care obligations
Enrollment patterns	8	Delayed enrollment, extended time to completion
Capacity, readiness, access	6	Academic readiness, digital access

Note. Frequencies reflect non-mutually exclusive categories; studies frequently used multiple descriptors in combination.

Conceptualization and Measurement of Persistence and Retention

Persistence and retention were central outcomes across the included studies and were most often examined in combination with one another. Persistence was the most frequently reported outcome domain (n = 34) and was typically defined as continued enrollment, term-to-term continuation, or sustained participation. Persistence was commonly operationalized using enrollment continuity indicators or self-reported persistence measures.

Retention was reported as a primary or co-primary outcome (n = 24) and was generally framed as an institutionally defined metric. Retention measures most often relied on administrative indicators such as enrollment status, re-enrollment rates, or year-to-year continuation. Persistence and retention were used interchangeably or reported together (n = 14) without clear distinction between student-level continuation behaviors and institution-level retention metrics.

Across studies, the most frequently reported outcome domains were persistence and continuation, followed by completion and engagement-related outcomes. Completion outcomes (e.g., course completion, graduation, credential attainment) were reported (n = 13). Engagement, integration, or belonging outcomes appeared (n = 11), while satisfaction or perceived learning outcomes were reported less frequently (n = 6). Academic performance indicators like grades, or grade point average (GPA) were least commonly examined (n = 5). Outcome measurement was predominantly multi-dimensional, with variation in outcome combinations, temporal scope, and data sources ranging from short-term course-level indicators to institution-level metrics and incorporating both self-reported and administrative data. Table 4 illustrates the frequency of outcome domains examined across the included studies, highlighting how persistence, retention, and related constructs were operationalized and reported.

Table 4

Frequency of Reported Outcome Domains

Outcome Description	Frequency (*n*)
Persistence, continuation	34
Institutional retention, re-enrollment	24
Completion, graduation, attainment	13
Engagement, integration, belonging	11
Satisfaction, perceived learning	6
Academic performance	5

Note. Frequencies reflect non-mutually exclusive categories; studies report multiple outcome domains.

Identified Thematic Domains

Analysis of the charted data identified recurring patterns in how persistence and retention among non-traditional students were examined across the online higher education literature. Studies frequently contributed to multiple domains, reflecting the interconnected nature of persistence and retention across learner, instructional, and institutional levels.

As presented in Table 5, five interrelated thematic domains were identified: (1) learning environment and pedagogical design, (2) learner characteristics and life contexts, (3) institutional structures and policy, (4) support strategies and interventions, and (5) psychological capital and self-regulation. An emergent sixth theme, technology and AI-mediated practices, captured enabling infrastructures supporting instructional delivery, monitoring, and coordinated outreach in online learning environments.

Table 5

Thematic Domains with Framework Alignment

Theme	Description	Frequency (*n*)	RQ Alignment	Framework Alignment
Learning Environment & Pedagogical Design	Course-level instructional and relational practices through which institutional support is enacted, including instructor presence, communication, and pedagogical design.	38	4	CPMOL: Academic integration; Social presence
Learner Characteristics & Life Contexts	Demographics as contextual factors, role conflict, and lived realities.	47	3	MNSA: External life context
Institutional Structures & Policy	Institutional policies, infrastructures, and delivery models shaping access, continuity, and learner progression.	35	2	MNSA: Institutional fit and responsiveness
Support Strategies & Interventions	Structured institutional practices connecting learners to academic systems, expectations, and resources.	32	1	CPMOL: Institutional support; Social integration
Psychological Capital & Self-Regulation	Grit, self-regulation, motivation, and human agency.	29	3	MNSA, CPMOL: Learner agency
Emergent: Technology & AI-Mediated Practices	Enabling infrastructures supporting instructional delivery, monitoring, and coordinated outreach.	11	4	CPMOL: Technological infrastructure

Note. Categories are non-mutually exclusive.

Discussion

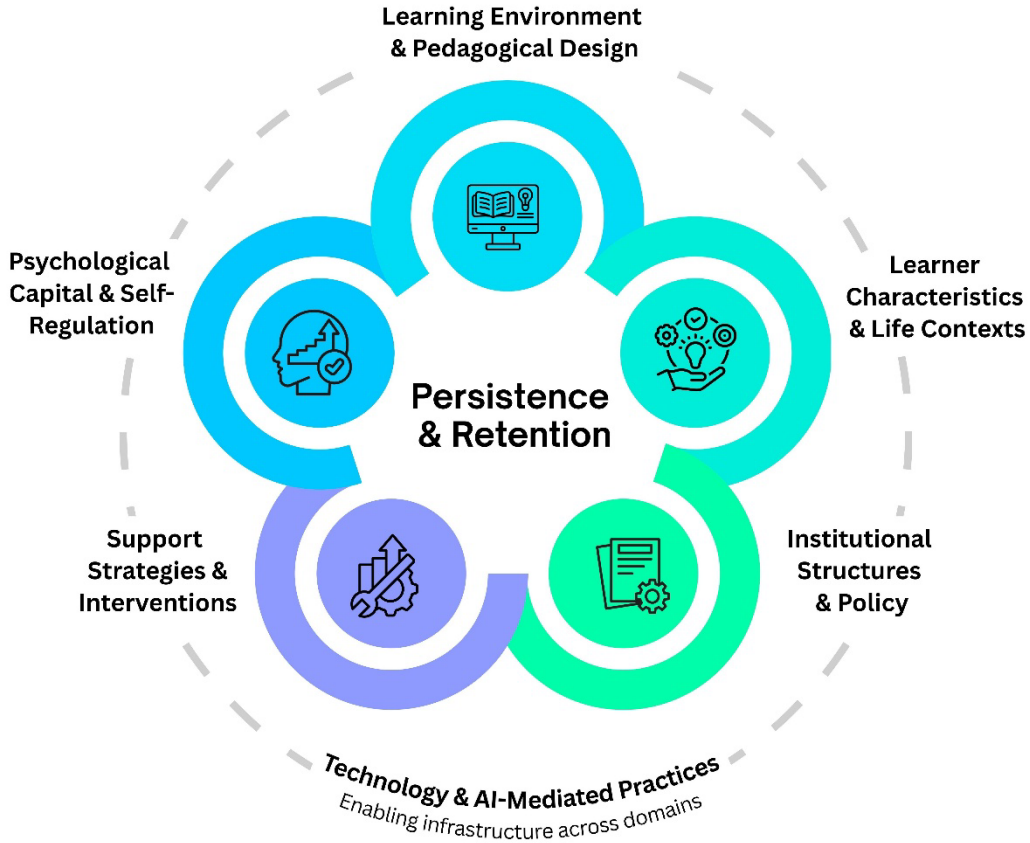
Persistence and retention among non-traditional online students are best understood as outcomes produced through the interaction of institutional structures, instructional environments, coordinated supports, and learner agency (Lehan et al., 2018; Stephen, 2022). Across the mapped evidence, continuation depended on how effectively course-level practices, institutional policies, and support systems aligned with learners' external responsibilities and life circumstances; when alignment weakened, structural friction disrupted sustained progression (Chaffin & Shatila, 2023; Motte, 2020; Sutton, 2014). Interpretation of these findings is complicated by the frequent conflation of persistence, retention, attrition, and completion across the literature, as well as by the strong concentration of studies within the U.S. context, which shapes how these constructs

are defined, measured, and reported. When these constructs are used interchangeably or without explicit definition, it becomes difficult to distinguish learner decision-making from instructional design effects, policy constraints, or institutional accountability, particularly within the U.S.-based reporting frameworks such as the Integrated Postsecondary Education Data System and the National Center for Education Statistics, which rely on standardized time-to-degree metrics and may underrepresent non-linear enrollment patterns common among non-traditional learners (De Paepe et al., 2018; Haydarov et al., 2013). These patterns align closely with established persistence frameworks foregrounding external life context, institutional fit, (Bean & Metzner, 1985) and conditions specific to online learning environments (Bean & Metzner, 1985; Rovai, 2003), while maintaining the descriptive and mapping purpose of a scoping review.

Synthesis reflects a shift away from explanations of attrition grounded in learner motivation or preparedness and toward a systemic framing emphasizing environmental fit, institutional responsiveness, and intentional design (Chaffin & Shatila, 2023; Stephen, 2022). The literature converges around five interrelated domains organized by how persistence and retention were conceptualized: learning environment and pedagogical design; learner characteristics and life contexts; institutional structures and policy; support strategies and interventions; and psychological capital and self-regulation. The following domains functioned together: courses translated institutional intent into lived experience, support mechanisms mediated complexity across systems, and psychological resources were activated or constrained by the alignment among course design, institutional policy, learner life context, and support availability. Technology and AI-mediated practices emerged as an enabling infrastructure, strengthening or weakening alignment depending on how tools are embedded in instructional and institutional routines. Figure 2 presents an applied visual synthesis of the thematic domains identified in the scoping review, illustrating persistence and retention as system-level outcomes produced through the interaction of instructional environments, institutional structures, learner contexts, coordinated supports, and psychological capital, with technology functioning as enabling infrastructure.

Figure 2

Applied Synthesis of Persistence and Retention Domains



Learning Environment and Pedagogical Design

The learning environment operates as the primary site where institutional commitments to persistence were enacted and experienced by non-traditional online learners. Across the literature (n = 38), course-level instructional and relational practices, particularly instructor presence, communication, and pedagogical design, were consistently framed as mechanisms through which institutional support becomes tangible in everyday learning experiences (Bedford, 2023; Belt et al., 2025; Rosser-Majors et al., 2022). These practices positioned persistence as co-produced within instructional relationships, distributing responsibility across faculty, learners, and institutional design.

Instructor presence emerged as a prominent pedagogical condition shaping persistence aligning with Community of Inquiry research. Recent studies further extend this work, suggesting emotional presence and care-oriented facilitation are associated with student satisfaction and short-term persistence (Jiang & Koo, 2020; Stone & Springer, 2019). Teaching presence (intentional design and facilitation), social presence (humanizing the virtual space), and cognitive presence (supporting meaning-making through sustained discourse) structured learners'

experience of clarity, connection, and academic direction in online learning environments (Chaffin & Shatila, 2023; Oregon et al., 2018). Emerging attention to emotional presence further highlighted the role of care, empathy, and responsiveness in shaping learner confidence and satisfaction (Bedford, 2023). Presence was conveyed through both direct communication (e.g., individualized outreach, timely feedback) as well as cues embedded in course organization, instructional media, and instructor persona.

Relational connection was reinforced through predictable communication patterns, transparent expectations, and feedback functioning as a primary relational anchor. Peer interaction also contributed to social integration, including forms of "passive collaboration" through which non-traditional online learners developed a sense of belonging by observing and engaging with others' work (Belt et al., 2025; McNeil et al., 2025; Trespalacios & Uribe-Florez, 2020). Relational practices reduced isolation and lowered affective barriers, positioning faculty as key conduits through which institutional structures were experienced interpersonally.

Pedagogical effectiveness was closely tied to the stability and coherence of course design. Although flexibility was widely emphasized, the literature distinguished between flexibility intentionally embedded in course structures and discretionary accommodations enacted on an ad hoc basis (Chaffin & Shatila, 2023; Rosser-Majors et al., 2022). Courses that built flexibility into their structure through modular organization, predictable assessments, and clear extension policies were more consistently associated with student persistence. Studies of program redesign and course structure similarly emphasized predictable rhythms and modular designs reduced administrative burden and improved continuation among non-traditional learners (Aversa & MacCall, 2013; Rosser-Majors et al., 2022; Sutton, 2014). Predictability emerged as a consequential, though often understated, feature of effective design: stable interaction rhythms, clear timelines, and consistent procedures reduced cognitive and administrative load for learners balancing work, caregiving, and financial responsibilities (McNeil et al., 2025; Rosser-Majors et al., 2022; Sutton, 2014). In this domain, the reliability of the learning environment appeared to function as a stabilizing force for persistence.

Learner Characteristics and Life Contexts

Learner characteristics and life contexts constituted a central analytic domain across the literature ($n = 47$), with studies consistently treating demographics as situated and dynamic conditions shaping educational participation rather than as fixed predictors of persistence (Ellis, 2019; Kara et al., 2019; Stephen, 2022). This perspective marked a move away from deficit-oriented interpretations of attrition, framing persistence as a contingent outcome unfolding within the limits and possibilities of learners' lived circumstances. Non-traditional learners were most often portrayed as managing competing roles, commonly as students, employees, and caregivers, within constrained temporal, financial, and emotional environments (Bourdeaux & Schoenack, 2016; Burke, 2020; Shatila, 2023).

Although non-traditional status was frequently operationalized using broad demographic and role-based descriptors, these characteristics functioned primarily as contextual labels rather than as analytically differentiated conditions. The literature emphasized substantial heterogeneity beneath the non-traditional umbrella, shaped by employment intensity, caregiving responsibilities, prior educational disruption, and gendered distributions of domestic labor

(Banerjee & Zhang, 2024; Dumais et al., 2013). Role conflict emerged as a pervasive barrier, particularly for women, whose academic participation was often negotiated alongside primary responsibility for childcare, eldercare, and household management (Burke, 2020). International studies suggest additional context-specific factors, including socio-political conditions or culturally grounded conceptions of belonging, may also shape persistence in ways less frequently captured in U.S.-focused literature. Employment demands, frequently involving full-time or variable schedules, further limited learners' capacity to meet rigid academic timelines.

Within this context, withdrawal and stop-out were commonly framed as deliberate and rational life decisions rather than indicators of academic failure. Disengagement tended to occur when the cumulative demands of work, family, health, or financial instability exceeded available capacity, with many learners exhibiting non-linear patterns across time and institutions (Ellis, 2019; Layne et al., 2013; Sylvester, 2025). Spatial and environmental factors further shaped persistence, including unstable broadband access, limited technology, lack of private study space, and the relational isolation associated with online learning environments (Cordano, 2025; Swartzwelder et al., 2023).

Despite widespread acknowledgment of external life demands, these realities were infrequently translated into instructional or institutional design decisions (Brown, 2017; Walsh et al., 2024). When misalignment was framed primarily as a learner-level challenge, responsibility shifted away from institutional systems and practices. Persistence was therefore conceptualized as a situated process emerging from the interaction between learners' lived realities and instructional, institutional, and environmental conditions highlighting the limits of context-light approaches and calling for intentional alignment with the lives of non-traditional learners (Cato Parnell, 2022; Walsh et al., 2024).

Institutional Structures and Policies

Institutional structures and policies shaped access, continuity, and learner progression across the literature (n = 35). Policies, infrastructures, and delivery models were consistently framed as institutional infrastructure; the invisible scaffolding mediating the fit between learners' complex lives and academic requirements. Within this domain, persistence and retention were understood as consequences of organizational design, with persistence treated as a student-centered behavior and retention as an institutional metric (Motte, 2020; Stephen, 2022).

Academic flexibility functioned as a central structural enabler, particularly through accelerated formats and asynchronous delivery. These features addressed the temporal and physical constraints faced by working adults and caregivers, enabling coursework to be integrated into complex and shifting schedules (Burke, 2020; Cordano, 2025; Tolouian et al., 2022). By contrast, rigid academic calendars, densely packed semesters, and inflexible pacing structures repeatedly emerged as sources of institutional friction. Policies governing course load expectations, grading timelines, withdrawal procedures, and satisfactory academic progress often operated as gatekeeping mechanisms when misaligned with learners' lived realities (Motte, 2020; Stephen, 2022).

Financial scaffolding also played a significant role in enrollment continuity, particularly within the U.S.-based studies where federal aid systems such as Pell Grants are frequently

examined as primary drivers of persistence, reflecting a founding model not directly comparable to centralized or publicly funded systems in other national contexts. Financial aid, including Pell Grants and layered funding supports, was described as supporting credit accumulation and reinforced academic momentum, whereas financial uncertainty constrained learners' ability to remain enrolled over time (Dello Stritto et al., 2025; Sutton, 2014). Administrative clarity and institutional integrity further shaped persistence. Transparent onboarding processes, recognition of prior learning, and alignment between institutional mission and enacted practice facilitated navigation of academic systems (Motte, 2020; Walsh et al., 2024). Digital infrastructure including intuitive learning management systems and accessible technical support extended institutional commitment into the virtual environment, whereas poorly designed systems contributed to frustration and disengagement (Dello Stritto et al., 2025; Preuss et al., 2023).

Although institutional policies were frequently referenced, they were more often described than analytically examined. Institutional friction was typically inferred through accounts of stop-out or withdrawal attributed to "external factors," obscuring how policy design itself shaped persistence trajectories. Where institutional responsiveness was addressed, it appeared uneven and was supported by limited empirical evaluation (Cato Parnell, 2022; Stephen, 2022; Walsh et al., 2024). Overall, the literature reflects a gradual shift toward recognizing persistence as a shared institutional responsibility, contingent on the interaction of governance practices, administrative systems, and delivery models with non-traditional learners' life circumstances.

Support Strategies and Interventions

Support strategies and interventions formed a key domain across the literature (n = 32), centering on structured practices connecting non-traditional learners to academic systems, expectations, and resources. These practices were consistently conceptualized as *bridging mechanisms*: intentional, personalized interventions mediating the relationship between learners' complex lives and institutional demands. Support operated as a relational layer translating institutional structures into guidance and reduced isolation across the student lifecycle (Amundsen, 2023; Motte, 2020; Seery et al., 2021; Walsh et al., 2024).

Proactive and strength-based approaches were most prominent, particularly intrusive advising models using risk indicators to initiate outreach before disengagement occurred. Advising was frequently described through a concierge model, in which a single, consistent point of contact supported learners in navigating academic, administrative, and personal challenges. In this role, advisors and coaches served as interpretive interfaces, helping learners make sense of institutional rules, timelines, and expectations while reducing uncertainty around progression (Gaytan, 2013; Kish, 2025; Preuss et al., 2023).

Academic coaching was distinguished from advising by its emphasis on developing learner competence and self-directed learning skills, including time management, study strategies, and academic planning. Coaching addressed barriers often invisible within formal academic systems yet consequential for persistence (Motte, 2020; Preuss et al., 2023). Scaffolded onboarding and transition supports extended assistance beyond initial entry facilitating academic integration and later-stage mentoring. Credit for prior learning functioned as both a structural and symbolic bridge, reducing time-to-degree while validating learners' professional experience and identity (Motte, 2020; Sylvester, 2025).

Despite their prevalence, support strategies were unevenly integrated across the student lifecycle. Many were described as episodic interventions activated at specific moments. Limited alignment across faculty, advising, financial aid, and student services constrained the capacity of supports to address cumulative challenges, particularly for learners managing fluctuating work, caregiving, or financial conditions. As a result, the effectiveness of support strategies depended less on their presence than on their coordination and coherence within the broader institutional system (Motte, 2020; Walsh et al., 2024).

The literature positions support strategies as connective tissue between institutional structures and course-level experiences. When integrated and aligned, these practices support persistence as an institutionally mediated process rather than an individual burden.

Psychological Capital and Self-Regulation

Psychological capital (n = 29) emerged as a distinct analytic domain across the literature, encompassing non-cognitive, motivational, and self-regulatory resources through which learners enact agency in online learning environments. This domain emphasized the psychological stamina required to sustain commitment amid the demands and uncertainty characterizing non-traditional students' educational trajectories. Across studies, psychological resources were framed as operating in relation to instructional and institutional conditions rather than as independent determinants of persistence (Gullixson, 2015; Shatila, 2023).

Grit and long-term stamina, defined as sustained passion and perseverance toward superordinate goals, were associated with higher GPA and program completion, particularly when learners understood education as a prolonged endeavor (Cross, 2013; Louie, 2024; McClendon et al., 2017). A "marathon mindset" was especially salient in graduate and doctoral contexts, where progress unfolds over extended timeframes and interruptions are common (Ellis, 2019; Stephen, 2022). Grit was frequently distinguished from momentary self-discipline, functioning instead as persistence oriented toward long-horizon goals.

Persistence was also linked to mechanisms of human agency, including self-regulation, self-direction, and online learning self-efficacy. Learners' capacity to set goals, manage time, monitor progress, and navigate technological environments independently supported continued enrollment, particularly in highly autonomous online contexts (McNeil et al., 2025; Stephen, 2022). Locus of control further shaped persistence trajectories, with learners who attributed success to their own actions more likely to persist through challenge than those who perceived outcomes as externally determined (Su & Waugh, 2018; Yoo & Huang, 2013).

Motivational orientation reflected a balance between intrinsic motives, such as personal fulfillment, identity development, and role modeling, and problem-centered extrinsic goals related to workplace application and career advancement (Beattie, 2022; Sylvester, 2025). Persistence was strongest when learners perceived clear relevance and utility in their academic work. At the same time, psychological constructs were often treated as stable individual traits, detached from the environments in which they were activated. Experience-based measures such as engagement and satisfaction were frequently used as proxies for persistence, limiting insight into sustained enrollment (Bawa, 2016; Louie, 2024).

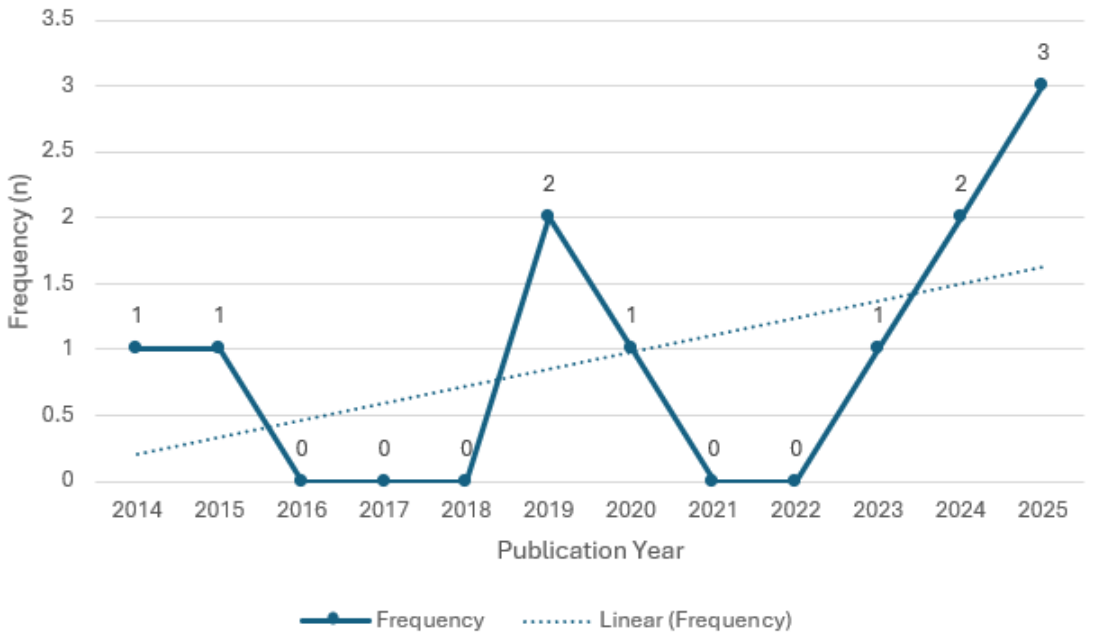
The evidence suggests psychological capital functions most meaningfully within supportive educational contexts. When institutional and pedagogical systems reduce uncertainty and friction, learners are better positioned to mobilize psychological resources in service of persistence (Lehan et al., 2018; Motte, 2020; Walsh et al., 2024).

Emergent Theme: Technology and AI-Mediated Practices

Technology extended the system-level pattern evident across the preceding domains by coordinating instructional, institutional, and support practices in online learning environments. Technology and AI-mediated practices appeared as an emergent domain in the literature (n = 11), most often described as background conditions supporting instructional delivery, engagement monitoring, advising coordination, and administrative response. Learning management systems, analytics tools, and early alert mechanisms augmented the work of faculty and advisors by facilitating timely communication and targeted outreach (Choi & Park, 2018; Kish, 2025; Rosser-Majors et al., 2022). Across the mapped evidence, these tools were not framed as independent solutions to attrition; instead, they linked course activity, advising workflows, institutional policy, and learner outreach. Figure 3 illustrates that this domain became more visible in recent publications.

Figure 3

Technology and AI–Mediated Practices by Publication Period



The cross-cutting nature of this domain was evident in its connection to each of the preceding themes. Learning management systems structure the learning environment through content access, communication, pacing, assessment, and feedback. Analytics and early alert systems link learner activity to advising and support, positioning technology as a bridge between instructional design and institutional response. However, for non-traditional learners, digital

traces often fail to capture the full context of participation. Irregular logins, delayed submissions, or episodic engagement may reflect employment demands, caregiving responsibilities, financial strain, limited study space, or asynchronous participation patterns and do not necessarily indicate disengagement (Burke, 2020; Dumais et al., 2013; Kara et al., 2019; Stephen, 2022). These indicators therefore require contextual interpretation before being classified as risk, particularly when used to trigger advising, outreach, or administrative intervention (Kish, 2025; Walsh et al., 2024).

Empirical evaluation of technology's impact on persistence remained limited. Where discussed, analytics and AI-enabled systems were positioned as extensions of existing advising and support structures, with human interpretation and contextual judgment emphasized as necessary complements (Kish, 2025; Lehan et al., 2018; Walsh et al., 2024). Although references to technology- and AI-mediated practices increased over time, the literature provides limited insight into how these tools reshape responsibility, trust, or learner agency. This gap is consequential because technology-mediated systems may shift how responsibility for persistence is distributed between learners and institutions. When embedded within coordinated systems, analytics may enable earlier identification of barriers and more coherent institutional response. When used primarily for monitoring, they may shift responsibility onto learners by flagging risk without addressing underlying conditions (Stephen, 2022; Walsh et al., 2024). The central issue is not simply whether technology is used, but how it is governed, interpreted, and integrated into institutional practice. Technology and AI-mediated practices raise underexamined questions related to transparency, bias, privacy, surveillance, and student agency. Few studies examined whether learners understood how engagement data were collected or used, whether risk indicators reflected adult learner participation patterns, or whether students had opportunities to contextualize automated alerts. Without such attention, technology-mediated systems risk reproducing deficit interpretations of constrained participation by treating limited or uneven engagement as evidence of insufficient motivation, readiness, or commitment.

Technology appeared most effective when embedded within coherent institutional ecosystems aligning policy, pedagogy, and support practices. When thoughtfully designed and governed, these tools extended the reach of existing advising, instructional communication, and support systems (Ranasinghe et al., 2025; Stephen, 2022). Poor integration may instead amplify fragmentation by detecting risk without altering underlying conditions. Early alerts are actionable only when linked to advising, flexible policy, financial guidance, and relational outreach (Kish, 2025; Lehan et al., 2018; Walsh et al., 2024). Visibility alone does not reduce structural friction.

Accordingly, technology and AI-mediated practices are best understood as institutionally situated practices. Their contribution to persistence depends on whether they enhance institutional responsiveness, strengthen human support, and reduce navigational burden for non-traditional learners. Technology supports persistence only when it is transparent, context-sensitive, human-mediated, and connected to the institutional conditions shaping learners' capacity to continue.

The literature consistently frames persistence and retention among online non-traditional students as outcomes produced through the alignment of instructional environments, institutional structures, coordinated supports, and learner agency. Persistence depends on system coherence,

as course-level relational practices, structural flexibility, and mediating support mechanisms interact with learners' life contexts to shape friction across the student lifecycle. Within this system, psychological resources are activated or constrained by institutional conditions, and technology strengthens persistence only when embedded within human-mediated practices. Yet the evidence remains uneven in explaining how these domains interact over time, particularly as technology and AI-mediated systems become more embedded in institutional practice.

Gaps in Evidence

This scoping review identified several gaps in the evidence base examining persistence and retention among non-traditional students in online higher education. The evidence base is also geographically concentrated, with limited examination of how persistence and retention are conceptualized and supported across differing national policy, funding, and cultural contexts. A central limitation was not simply the absence of evidence, but the limited examination of how instructional, institutional, and support systems interact over time to produce persistence or attrition. Although persistence and retention remain core outcomes, they are conceptualized inconsistently, obscuring how responsibility for continuation is framed within educational systems.

Persistence is most often operationalized as a student-level behavior, such as continued enrollment or course completion, while retention is typically treated as an institution-level metric based on re-enrollment or graduation. The frequent interchangeable use of these constructs, or their use without explicit definition, limits the field's ability to distinguish learner decision-making from instructional design effects, policy constraints, or institutional accountability. This conceptual ambiguity constrained cumulative knowledge building and weakened the interpretability and comparability of findings across studies.

Gaps were also evident in how persistence is temporally and structurally examined. Short-term indicators such as term-to-term enrollment and course completion dominate the literature, while longer-term and non-linear trajectories, including stop-out, re-entry, and delayed completion remain underexamined. This emphasis reflected an implicit assumption of linear enrollment pathways not in alignment with the lived realities of adult online learners, for whom persistence is often episodic and contingent on changing life conditions.

Methodological limitations further restrict understanding of persistence as a system-level phenomenon. Cross-sectional and short-term survey designs predominated, limiting causal inference and obscuring interaction effects across instructional, policy, and support domains. Reliance on single-institution or program-specific studies constrains insight into how different institutional configurations distribute structural friction or support. Longitudinal designs following learners across multiple enrollment decisions, including stop-out and re-entry, remain rare, and intervention studies often lacked sufficient detail on implementation context or resource requirements to support broader application.

Representation gaps were also evident. While studies frequently referenced "non-traditional" learners as a category, few disaggregated findings by subgroup or examined how institutional practices differentially affect learners navigating distinct combinations of work, caregiving, financial precarity, or access constraints. Limited attention to caregivers, first-

generation students, low-income learners, adult literacy learners, and individuals facing language-related access barriers restricts understanding of how persistence mechanisms function across intersecting life circumstances. Peer connection and social presence were infrequently examined as focal persistence mechanisms, despite their relevance in online learning environments.

Although technology and AI-mediated practices appear with increasing frequency, the literature provides limited insight into their systemic effects. Most studies emphasized detection, monitoring, or efficiency gains, with minimal examination of how these tools redistribute responsibility for persistence between learners and institutions or reshape trust, agency, and relational dynamics. Empirical evaluation of technology-enabled interventions remains sparse, particularly with respect to learner experience and governance. These gaps suggest the evidence base is better suited to identifying isolated correlates of persistence than to explaining how persistence is produced through the alignment, or misalignment, of institutional systems across the learner lifecycle.

Implications

Implications for Institutional Practice and Policy

The findings of this scoping review indicate supporting persistence and retention among non-traditional students in online higher education requires sustained attention to institutional coherence. Persistence emerged most reliably when institutional policies, instructional practices, and support structures were intentionally aligned with the lived realities of learners managing work, caregiving, financial precarity, and other external demands.

At the institutional level, leaders should examine how policies governing enrollment, academic progress, course sequencing, and financial aid interact with instructional expectations and support availability. This policy review should also include withdrawal, stop-out, re-entry, transfer credit, prior learning assessment, and satisfactory academic progress policies, particularly where institutional requirements assume linear enrollment or uninterrupted progression. Policies that were administratively coherent but insufficiently responsive to fluctuating life circumstances often intensified structural friction and elevated attrition risk, whereas applying a persistence-oriented lens to policy review can help identify misalignments between institutional requirements and learners' capacity to comply, particularly in online programs serving working adult learners.

At the instructional level, evidence points to the importance of course design practices prioritizing clarity, predictability, and structured flexibility. Consistent communication, transparent assessment criteria, stable pacing, and coherent course organization support sustained participation more reliably than ad hoc accommodations or pedagogical novelty. Embedding these expectations into program-level standards and instructional design guidelines reduces variability in student experience and limits reliance on individual faculty discretion as the primary source of flexibility.

Support strategies are most effective when integrated across the student lifecycle and coordinated across functional units. Shifting from episodic, reactive models toward proactive and

navigational approaches enabled learners to anticipate academic and administrative demands over time rather than respond to crises as they arose. Strengthening coordination among faculty, advising, financial aid, and student services reduces the interpretive burden placed on learners and improves continuity of guidance across enrollment phases. Structured onboarding, early-term check-ins, and coordinated handoffs between instructional and support staff are particularly important during transition points in online learning environments when attrition risk is elevated.

Technology and AI-enabled tools offer potential efficiencies for monitoring engagement and facilitating outreach, but the evidence cautions against treating these systems as substitutes for relational support or structural alignment. Institutional leaders should attend to how analytics and automated tools were embedded within broader advising, instructional, and policy frameworks supporting adult online learners, ensuring transparency, human oversight, and responsiveness to learner context. Policy should also clarify data privacy, bias monitoring, student agency, interpretation of risk indicators, and responsibility for outreach. When aligned with coherent institutional design, technology can extend the reach of support systems; when poorly integrated, it risks reinforcing fragmentation.

Finally, institutional policy should support disaggregated assessment of persistence and retention outcomes by non-traditional learner characteristics, including employment intensity, caregiving status, age, first-generation status, financial independence, and re-entry status. Aggregate retention metrics may otherwise obscure differential barriers among adult online learners.

Persistence for non-traditional online learners was best supported when institutions designed policies, courses, supports, and technologies as interdependent components of a coordinated system. Institutional capacity to support continuation is strengthened not by adding discrete services, but by reducing friction and increasing alignment across the structures shaping learners' daily experience of online learning environments.

Implications for Future Research

Future research on persistence and retention among online non-traditional students should move beyond identifying isolated predictors and instead examine how institutional, instructional, and support systems interact over time to enable or constrain continuation. Advancing the field will require greater conceptual clarity, methodological rigor, and sustained attention to how persistence is produced through system alignment.

Conceptually, researchers should explicitly distinguish between persistence as a student-centered behavior and retention as an institutional metric, specifying levels of responsibility and temporal scope. Clear operationalization of these constructs, including defined time horizons, is essential for cumulative knowledge building and meaningful comparison across studies. Greater consistency in terminology will also help clarify whether observed relationships reflect learner decision-making, instructional design, institutional policy, or measurement artifacts.

Methodologically, longitudinal and lifecycle-oriented designs are needed to capture non-linear enrollment trajectories common among non-traditional learners, including stop-out, re-entry, and delayed completion. Studies following learners across multiple enrollment decisions

can illuminate how alignment, or misalignment, across course, program, and institution-level systems shapes persistence over time. Multi-institutional and comparative designs would further strengthen transferability and support examination of how different institutional configurations distribute structural friction or support.

Equity-centered research remains a priority. Future studies should disaggregate findings by learner subgroup, including caregiving status, employment intensity, socioeconomic position, first-generation status, and language-related access barriers, to examine how institutional designs differentially affect learners navigating intersecting life circumstances. Moving beyond aggregate "non-traditional" categories will allow more precise identification of which strategies supported persistence for whom, and under what conditions.

Expanding use of technology and AI-mediated practices warrants deeper empirical investigation. Future research should examine not only predictive accuracy or efficiency gains, but how these tools are designed, governed, and experienced by learners. Attention to issues of agency, trust, transparency, and surveillance is critical to understanding how technology reshapes responsibility for persistence and interacts with instructional and institutional contexts. These directions point toward a research agenda treating persistence as a system-level outcome—one that can be intentionally designed, rigorously evaluated, and iteratively improved through coordinated, multi-level inquiry spanning institutional structures, instructional and support practices, and learner experiences.

Limitations

This scoping review has several limitations to consider when interpreting the findings. The review was bounded by predefined temporal (2013–2025), language, and contextual criteria, including inclusion of only peer-reviewed English-language studies situated in online higher education. As a result, relevant research published in other languages, prior to 2013, or conducted in adjacent contexts, such as community-based adult education or workplace learning, may not be represented.

Despite a comprehensive search strategy, relevant studies may have been missed due to variability in terminology used to define non-traditional learners, persistence, and retention, as well as inconsistent indexing across databases. Although broad and inclusive terms were used, studies employing alternative terminology or relying on implicit definitions without clear keywords may not have been captured.

Consistent with JBI guidance for scoping reviews, this review did not include formal appraisal of methodological quality or risk of bias. Consequently, the findings map the breadth and characteristics of the evidence base rather than the effectiveness or comparative strength of specific strategies or interventions.

Conclusion

Persistence and retention among non-traditional students in online higher education are not individual problems to be fixed one learner at a time; they are system outcomes emerging when course design, institutional policy, coordinated supports, and learner agency align. This

scoping review synthesizes 47 studies into a usable taxonomy of five core domains and identifies technology and AI-mediated practices as an emergent area.

For institutions serving adult, working, and caregiving students, the map offers practical leverage points. These include prioritizing predictable and transparent course design and sustained instructor presence, auditing policies and financial supports for sources of friction, and integrating advising, coaching, and technology into coordinated, human-centered workflows rather than treating them as add-ons.

Scholarly contributions include reframing persistence as a relational and institutional design challenge, documenting the rise and limited empirical evaluation of technology and AI-mediated supports, and providing a conceptual foundation to guide future research toward system-level explanations and tests of alignment-focused strategies. This review offers an organizing scaffold for institutional improvement efforts and future studies examining which coordinated changes support longer-term continuation and completion among non-traditional students in online higher education.

Declarations

Comprehensive data and evidence mapping supporting this submission can be found in the online Supplemental Materials.

This research received no external funding.

The authors declare no conflicts of interest related to the research, authorship, or publication of this article.

Generative artificial intelligence (GenAI) was used in the preparation of this manuscript for grammar and figure refinement, using ChatGPT 5 and Canva. GenAI was not used to generate, analyze, or interpret research data, nor to create original scholarly content. All content was reviewed, verified, and approved by the authors, who take full responsibility for the accuracy and integrity of the work.

References

Amundsen, S. (2023). Focus on their strengths: Recruiting & retaining online adult learners in a hyper-competitive environment. *International Society for Technology, Education, and Science*.

Aversa, E., & MacCall, S. (2013). Profiles in retention part 1: Design characteristics of a graduate synchronous online program. *Journal of Education for Library and Information Science, 54*(2), 147–161. http://www.jstor.org/stable/43686942

Banerjee, M., & Zhang, H. (2024). Determining the influence of learner characteristics on online course success for underserved students in higher education. *Journal of Educators Online, 21*(3). https://doi.org/10.9743/jeo.2024.21.3.1

Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—A literature review. *Sage Open, 6*(1). https://doi.org/10.1177/2158244015621777

Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional Undergraduate Student Attrition. *Review of Educational Research, 55*(4), 485-540. https://doi.org/10.3102/00346543055004485

Beattie, E. (2022). The power of the positive: Enhancing online student engagement for adult literacy learners. *Adult Literacy Education: The International Journal of Literacy, Language, and Numeracy, 4*(1), 20-35. https://doi.org/10.35847/ebeattie.4.1.20

Bedford, L. (2023). Supporting online graduate students through the perspective of the affective filter hypothesis: A multiple case study of faculty interactions with students. *Journal of Online Graduate Education, 6*(2). https://doi.org/10.65201/001c.141601

Belt, E. S., Friesen, N., Lowenthal, P. R., & Snelson, C. (2025). Teacher–student relationships in an online graduate program: A phenomenological analysis. *Online Learning, 29*(3). https://doi.org/10.24059/olj.v29i3.4604

Bourdeaux, R., & Schoenack, L. (2016). Adult student expectations and experiences in an online learning environment. *The Journal of Continuing Higher Education, 64*(3), 152–161. https://doi.org/10.1080/07377363.2016.1229072

Brown, C. G. (2017). The persistence and attrition of online learners. *School Leadership Review, 12*(1), 47–58. https://scholarworks.sfasu.edu/slr/

Burke, C. M. (2020). Retention of women in online education. *Quarterly Review of Distance Education, 20*(4), 27–44. https://doi.org/10.1108/qrde-07-2020-0002

Cato Parnell, M. A. (2022). *Perceptions of higher education online students and faculty regarding attrition in online courses* (Order No. 29213827) [Walden University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/perceptions-higher-education-online-students/docview/2679598848/se-2

Chaffin, K. M., & Shatila, S. (2023). The case for online education: Success factors in an MSW program through a critical feminist lens. *Social Work Education, 43*(7), 1812–1830. https://doi.org/10.1080/02615479.2023.2221992

Choi, H. J., & Park, J.-H. (2018). Testing a path-analytic model of adult dropout in online degree programs. *Computers & Education, 116*, 130–138. https://doi.org/10.1016/j.compedu.2017.09.005

Cordano, V. (2025). Navigating online horizons: Fully online experiences of adult returning students at the University of Hawai'i at Mānoa. *International Education Journal Comparative Perspective*. https://doi.org/10.70830/iejcp.2402.19865

Cross, T. M. (2013). *Staying the course: Grit, academic success, and non-traditional doctoral students* (Order No. 3587687) [Pepperdine University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/staying-course-grit-academic-success-non/docview/1427344572/se-2

Dello Stritto, M. E., Skinner, D., Aguiar, N. R., & Underhill, G. R. (2025). Financial aid matters: The relationship among financial aid, retention, and graduation rates in online students. *Journal of Student Financial Aid, 54*(2). https://doi.org/10.55504/0884-9153.1866

De Paepe, L., Zhu, C., & DePryck, K. (2018). Drop-out, retention, satisfaction and attainment of online learners of Dutch in adult education. *International Journal on E-Learning, 17*(3), 303–323. https://www.learntechlib.org/primary/p/174173/

Dumais, S. A., Rizzuto, T. E., Cleary, J., & Dowden, L. (2013). Stressors and supports for adult online learners: Comparing first- and continuing-generation college students. *American Journal of Distance Education, 27*(2), 100–110. https://doi.org/10.1080/08923647.2013.783265

Ellis, H. (2019). Pursuing the conundrum of non-traditional student attrition and persistence: A follow-up study. *College Student Journal, 53*(4), 439–449. https://link.gale.com/apps/doc/A612581075/AONE?u=anon~e08a31a4&sid=googleScholar&xid=6b1669af

Gaytan, J. (2013). Factors affecting student retention in online courses: Overcoming this critical problem. *Career and Technical Education Research, 38*(2), 145–155. https://doi.org/10.5328/cter38.2.147

Gullixson, A. (2015). Rethinking retention: Using engagement to increase adult distance learner persistence. University Digital Conservancy. https://hdl.handle.net/11299/175446

Hadie, S. N. H. (2024). ABC of a scoping review: A simplified JBI scoping review guideline. *Education in Medicine Journal*, *16*(2), 185-197. https://doi.org/10.21315/eimj2024.16.2.14

Haydarov, R., Moxley, V., & Anderson, D. (2013). Counting chickens before they are hatched: An examination of student retention, graduation, attrition, and dropout measurement validity in an online master's environment. *Journal of College Student Retention: Research, Theory & Practice, 14*(4), 429–449. https://doi.org/10.2190/CS.14.4.a

Jiang, M., & Koo, K. (2020). Emotional presence in building an online learning community among non-traditional graduate students. *Online Learning, 24*(4). https://doi.org/10.24059/olj.v24i4.2307

Kara, M., Erdoğdu, F., Kokoç, M., & Cagiltay, K. (2019). Challenges faced by adult learners in online distance education: A Literature Review. *Open Praxis*, *11*(1), 5. https://doi.org/10.5944/openpraxis.11.1.929

Kish, D. (2025). Recommendations and best practices to increase online student retention. *Journal of Educators Online, 22*(4). https://doi.org/10.9743/JEO.2025.22.4.9

Layne, M., Boston, W. E., & Ice, P. (2013). A longitudinal study of online learners: Shoppers, swirlers, stoppers, and succeeders as a function of demographic characteristics. *Online Journal of Distance Learning Administration, 16*(2), 15–25.

Lehan, T. J., Hussey, H. D., & Shriner, M. (2018). The influence of academic coaching on persistence in online graduate students. *Mentoring & Tutoring: Partnership in Learning, 26*(3), 289–304. https://doi.org/10.1080/13611267.2018.1511949

Louie, Y. (2024). *Successful online adult learners' motivation for persistence to graduation* (Order No. 31633836) [Walden University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/successful-online-adult-learners-motivation/docview/3121559190/se-2

McClendon, C., Neugebauer, R. M., & King, A. (2017). Growth mindset, and deliberate practice in online learning. *Journal of Instructional Research, 6*(1). https://doi.org/10.9743/jir.2017.2

McNeil, S. G., Ahlf, M., & Ford, C. D. (2025). Exploring nontraditional graduate students' online engagement: an instrumental case study of learners' perceptions of purposefully designed activities. *International Journal of Educational Technology in Higher Education, 22*(1). https://doi.org/10.1186/s41239-025-00549-7

Motte, K. (2020). The intersection of stakeholder, educator, and community support for distance education doctoral education students. *Journal of Online Graduate Education, 3*(3). https://doi.org/10.65201/001c.141630

National Center for Education Statistics. (2024). *Non-traditional undergraduates/ definitions and data*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/pubs/web/97578e.asp

Oregon, E., McCoy, L., & Carmon-Johnson, L. (2018). Case analysis: Exploring the application of using rich media technologies and social presence to decrease attrition in an online graduate program. *Journal of Educators Online, 15*(2). https://doi.org/10.9743/jeo.2018.15.2.7

Preuss, M. D., Bennett, P. J., Renner, B. J., & Wanstreet, C. E. (2023). The factors influencing retention of online adult learners: A case study of a private institution. *International Journal of Research in Education and Science, 9*(1), 1–16. https://doi.org/10.46328/ijres.3005

Ranasinghe, K., Fernando, T. L. D., Vineeshiya, N., & Fernando, G. W. A. R. (2025). Improving student persistence in open and distance learning; An investigation of demotivational challenges. *Open Praxis, 17*(4), 648–663. https://doi.org/10.55982/openpraxis.17.4.850

Rosser-Majors, M. L., Rebeor, S., McMahon, C., Wilson, A., Stubbs, S. L., Harper, Y., & Sliwinski, L. (2022). Improving retention factors and student success online utilizing the

Community of Inquiry framework's instructor presence model. *Online Learning, 26*(2). https://doi.org/10.24059/olj.v26i2.2731

Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *The Internet and Higher Education*, *6*(1), 1-16. https://doi.org/10.1016/s1096-7516(02)00158-6

Seery, K., Barreda, A., Hein, S., & Hiller, J. (2021). Retention strategies for online students: A systematic literature review. *Journal of Global Education and Research, 5*(1), 72–84. https://doi.org/10.5038/2577-509x.5.1.1105

Shatila, S. L. (2023). Not alone when I'm feeling stressed: Online adult learner connection and retention. *Adult Education Quarterly, 74*(1), 43–61. https://doi.org/10.1177/07417136231184570

Stephen, J. S. (2022). Persistence of non-traditional undergraduate online students: Towards a contemporary conceptual framework. *Journal of Adult and Continuing Education, 29*(2), 438–463. https://doi.org/10.1177/14779714221142908

Stone, C., & Springer, M. (2019). Interactivity, connectedness and 'teacher-presence': Engaging and retaining students online. *Australian Journal of Adult Learning, 59*, 146–169.

Su, J., & Waugh, M. L. (2018). Online student persistence or attrition: Observations related to expectations, preferences, and outcomes. *Journal of Interactive Online Learning, 16*(1), 63–79.

Sutton, R. (2014). Unlearning the past: New foundations for online student retention. *The Journal of Educators Online, 11*(3). https://doi.org/10.9743/jeo.2014.3.5

Swartzwelder, K., Murphy, J., & Myers, J. (2023). Factors influencing student persistence in online courses during the pandemic: A mixed-methods study. *Journal of Educators Online, 20*(2). https://doi.org/10.9743/jeo.2023.20.2.12

Sylvester, J. (2025). *Navigating the journey of non-traditional women in online higher education: A narrative inquiry* (Publication No. 32114090) [Doctoral dissertation, University of Phoenix]. ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/navigating-journey-nontraditional-women-online/docview/3223079889/se-2

Tolouian, A., Wholeben, M., & Rankin, D. (2022). Student retention and success in an RN-BSN program: Determining variables and methods. *European Journal of Educational Sciences, 9*(2), 107–120. https://doi.org/10.19044/ejes.v9no2a107

Trespalacios, J., & Uribe-Florez, L. J. (2020). Developing Online Sense of Community: Graduate Students' Experiences and Perceptions. *Turkish Online Journal of Distance Education, 21*(1), 57-72. https://doi.org/10.17718/tojde.690340

Walsh, C., Bragg, L., Heyres, M., Yap, A., & Ratcliff, M. (2024). A systematic literature review of online academic student support in higher education. *Online Learning, 28*(2). https://doi.org/10.24059/olj.v28i2.3954

Yoo, S. J., & Huang, W. D. (2013). Engaging online adult learners in higher education: Motivational factors impacted by gender, age, and prior experiences. *The Journal of*

Continuing Higher Education, 61(3), 151–164.
https://doi.org/10.1080/07377363.2013.836823

Appendix

Descriptive Characteristics

Figure A1

Database Distribution of Included Studies

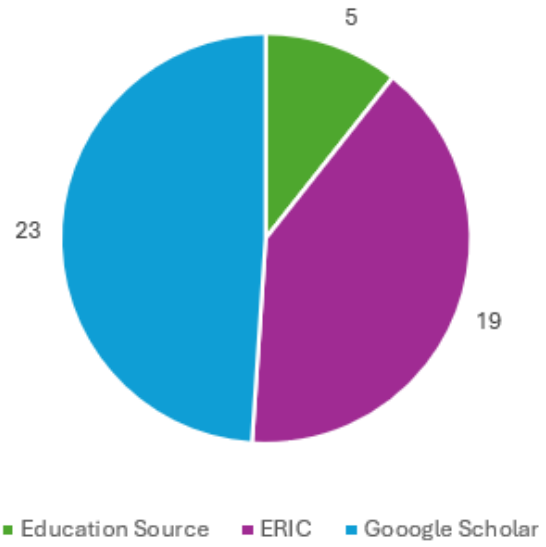


Figure A2

Publication Distribution of Included Studies by Year

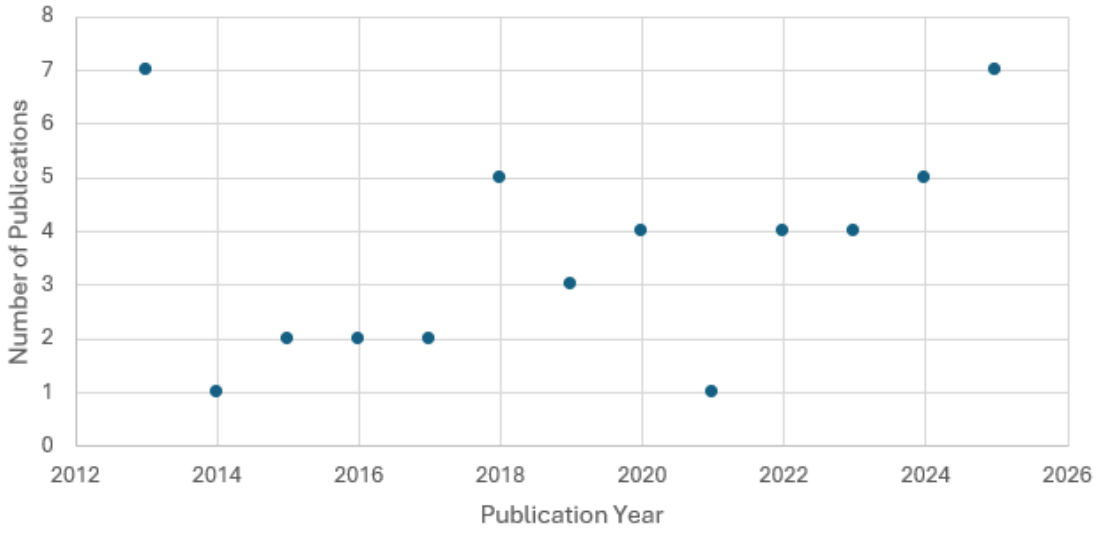


Figure A3

Frequency of Non-Traditional Student Descriptors Across Included Studies

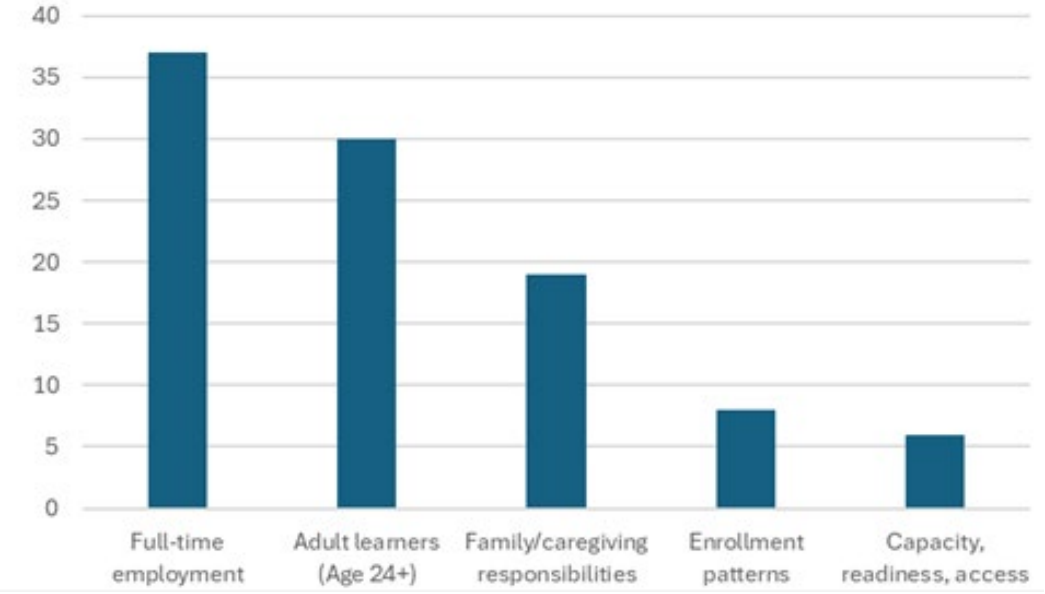
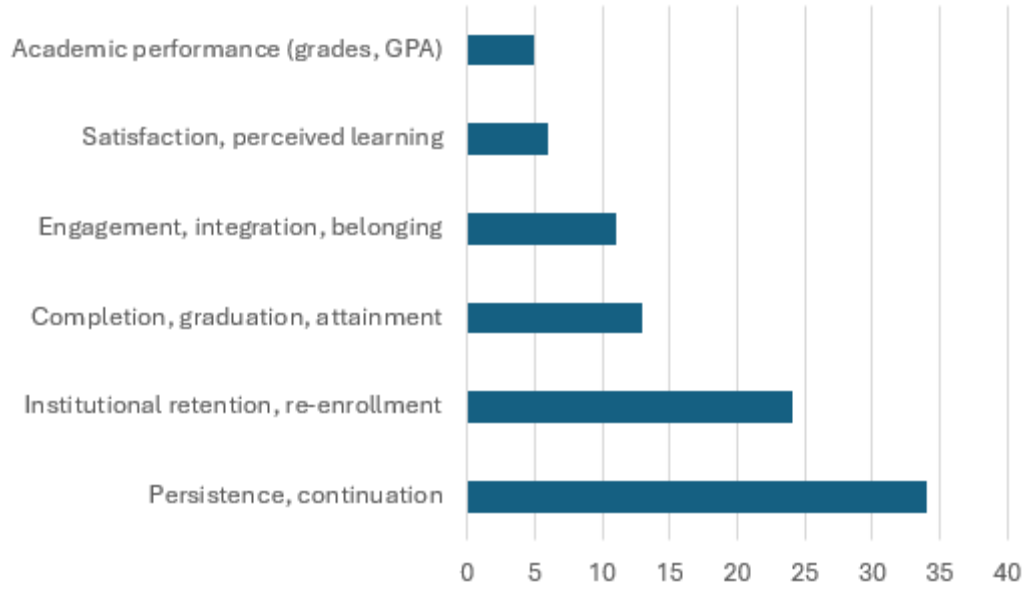


Figure A4

Frequency of Reported Persistence and Retention Outcome Domains



ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5781
https://doi.org/10.24059/olj.5781

Fostering Critical Consciousness Online: A Systematic Review of Strategies, Opportunities, and Gaps in the Literature

Jennifer Darling-Aduana
Qurrat ul Ain Rasheed
Ijeoma Mbaezue
Georgia State University
Kristin Hemingway
American Institutes for Research (AIR)

Abstract

This systematic review examines how learning in online environments and using online tools foster critical consciousness. A key tenet of critical and liberatory pedagogy - critical consciousness includes developing the skills to think critically about and take action to address structural inequities. Student participation in activities that foster critical consciousness are associated with higher academic, socio-emotional, and citizenship outcomes. However, there is limited documentation of critical consciousness raising online. Further, there is a risk that technology exacerbates rather than mitigates the long-standing inequities for individuals belonging to marginalized groups through embedded normative assumptions, algorithmic biases, and surveillance practices. Grounded in critical theory, we fill that gap and examine the strategies and components of critical consciousness fostered in online and blended learning environments across 96 studies published between 2010 and 2024. The most frequently observed critical consciousness raising strategy was students' engagement with power, and the least documented one was enactment of change inside and outside the classroom. The online environments and tools provided several unique affordances and advantages to fostering critical consciousness online, including digital content creation and sharing, critical media inquiry, online advocacy and community participation, digital cultural repositories simulations, and cross-cultural collaborations. Findings from this systematic review study highlight best practices within online teacher training, professional development, instructional design, and policy to support the enactment of liberatory pedagogy online that both inspires the mind and transforms communities. (231)

Keywords: critical consciousness, liberatory pedagogy, critical theory, educational equity, online learning, instructional design, teacher education

Corresponding author: Jennifer Darling-Aduana, jdarlingaduana@gsu.edu
This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 The Author(s).

Introduction

Critical consciousness raising is one of the primary goals and means of enacting liberatory education (Freire, 2000). Critical consciousness is also the third (and ultimate) tenet of one of the most prominent equity-based pedagogical frameworks, Culturally Responsive Pedagogy (CRP; Ladson-Billings, 1995, 2014). With a focus on equity, empowerment, and interactivity, critical consciousness includes obtaining the skills to think critically about and take action to address structural inequities (Freire, 1994, 2000; Jemal, 2017; Watts et al., 2011). Students experience individual as well as collective benefits through the development of critical consciousness including improved academic achievement (Gutstein, 2016; O'Conner, 1997), engagement and test scores (Luter et al., 2017), adolescent mental health (Zimmerman et al. 1999), and career outcomes (Diemer, 2009). Important even in the most progressive of times, critical consciousness raising takes on greater importance in eras such as this one of increased political censorship and the rolling back of human rights protections.

At the same time, online learning has become ubiquitous in K-16 education. While prevalent before the COVID-19 pandemic, the rapid shift to virtual schooling resulted in increased exposure to online learning tools and environments. Technological infrastructure and professional development investments reduced barriers to online and blended instruction (Fütterer et al., 2023; Sidi et al., 2023). It is now standard for children as young as Kindergarteners to receive some direct instruction through online videos and other resources, engage with self-directed digital programs as a regular component of the curriculum, have their behavior monitored and rewarded via online tracking software, and complete asynchronous learning activities during school closures. The dual pandemics of COVID-19 and systemic racism drew increased attention to the impact of long-standing inequities for individuals belonging to marginalized groups, with technology more often exacerbating than mitigating existing inequities (Darling-Aduana et al., 2024). For instance, students tend to be exposed to differential technology use by socioeconomic background with students from lower income and minoritized groups less likely be asked to use technology to create content and more likely to be told to use technology for drill and practice tasks (Delgado et al., 2015; Rafalow, 2020). Further the coding and content within online tools and programs exacerbate inequity through embedded normative assumptions, algorithmic biases, and surveillance practices (Benjamin, 2019; Darling-Aduana et al., 2022, 2024).

Despite the limitations of modal enactment, there are opportunities and bright spots in the use of online learning to support equity-based pedagogy. A recent systematic review identified promising strategies to enact CRP online across 42 studies (Viano & Darling-Aduana, 2025). However, most of these strategies focused on fostering academic excellence and cultural competence, the first and second tenets of CRP. Only 17 percent of studies included a strategy aimed at fostering critical consciousness, the third and final tenet. The strategies for fostering critical consciousness identified in the systematic review included acknowledging the role of power and systemic bias in society and encouraging critical action. None of the studies involved students in political movements or demonstrated a shared power structure in the classroom. Ladson-Billings (2014) has called out researchers and educators for this persistent overlooking of critical consciousness by CRP adopters. Accordingly, there is still a need to document strategies to foster critical consciousness in online spaces as well as to conceptualize the extent to which technology-based tools and learning modes may or may not align with the goals of liberatory education. We accomplish this, in part, by expanding the theoretical lens to include all critical and liberatory pedagogical frameworks.

In conducting this review, we aim to more clearly identify instructional strategies that foster critical consciousness online and expand the theoretical approaches from which we identify promising strategies. To accomplish this, we identified, reviewed, and placed in conversation studies across several fields, including instructional design, teacher education, and critical studies. Operating within critical theory (Apple, 2004; Freire, 2000; Giroux, 2010; Ladson-Billings, 1995, 2014), we comprehensively mapped existing literature to identify best practices and gaps in knowledge for further research through an examination of the following research questions: (1) What were the types and prevalence of strategies used to foster critical consciousness among studies examining online learning and learning using online tools? And (2) compared to face-to-face modes of learning, what are the opportunities for and constraints of fostering critical consciousness online? We examined this second research question using two modes of inquiry. First, we identified intervention characteristics that were unique or enhanced when critical consciousness raising online. Second, we examined variation in the prevalence of studies fostering various components of critical consciousness by time period, grade span, and online learning mode to identify opportunities and constraints specific to each of these elements. We further contribute to the literature by outlining recommendations that can be used to inform online learning policy, teacher training programs, and curriculum development for online learning with the aim of providing equitable, inspiring, and empowering learning experiences for K-16 students online.

Background

Conceptual and Theoretical Framework

Critical pedagogy is a collection of strategies regarding curriculum, instruction, and assessment with common assumptions and goals (Apple, 2004; Freire, 2000; Giroux, 2010; Ladson-Billings, 1995, 2014). Common tenets include the belief that there is no such thing as a neutral education (Apple, 2004; Freire, 2000). As a social institution, schools replicate dominant norms, expectations, and power dynamics unless purposefully identified, challenged, and transformed (Ladson-Billings, 1995, 2014). Schooling as a dominant cultural institution dehumanizes students and other actors (Freire, 2000), creating a self-fulfilling prophecy by which differential outcomes created by unjust social systems and inequities are internalized as personal failures (Jemal, 2017). When not explicitly countered, this internalization serves dominant groups by preventing individuals belonging to marginalized backgrounds from questioning or acting against the social structures and power dynamics that harm them.

Critical pedagogy aims to humanize education by helping students "gain greater control over their identities and lives, protect human rights, and reduce social injustice" (Jemal, 2017, p. 605). One of the primary means through which this is accomplished is by fostering critical consciousness, defined as providing students with the tools to recognize, critique, and work to rectify social inequities (Freire, 1970/2018; Ladson-Billings, 1995, 2014). Dimensions of critical consciousness include (a) critical reflection, (b) political efficacy, and (c) critical action (Jemal, 2017; Watts et al., 2011).

Critical reflection requires an understanding of the extent to which individuals are constrained and can exert agency upon systemic forces and structural oppression (Diemer et al., 2016; Giroux 1983). To build critical reflection skills, students must learn to recognize the underlying norms, beliefs, and assumptions driving individual, institutional, and systemic actions as well as be able

to place those norms, beliefs, and assumptions within appropriate sociohistorical and political context. Carlson, Engbretson, and Chamberlain (2006) identified four stages of critical reflection, including: (1) passive adaptation, (2) emotional engagement, (3) cognitive awakening, and (4) intention to act.

Political efficacy describes the extent to which students come to see themselves as capable of affecting social and political change, either individually or as part of a collective group (Watts et al., 2011). For instance, the Sociopolitical Control Scale (SPCS) assesses youth political efficacy by assessing agreement to statements regarding desire to be a leader and perceived leadership competence, whether they believe individual political action and voice have the ability to shape outcomes, and the value they place on being heard as part of the political process (Zimmerman & Zahniser, 1991). Political efficacy is an important part of fostering critical consciousness because individuals who believe they have the ability to enact change are more likely to feel empowered to advocate for themselves and others.

Lastly, critical action includes participating in individual or collective acts with the goal of changing the systems and processes creating systemic inequities (Diemer et al., 2014; Watts et al., 2011). For instance, the political activities subscale of the Youth Inventory of Involvement (YII) includes measures for participating in a protest as well as making presentations to organizations or politicians (Pancer et al., 2007). Another important element of critical action – desire to engage in political action specifically to reduce inequities - is measured by the Activism Orientation Scale (AOS) (Corning & Myers, 2002).

Critical Consciousness in the Classroom

Students achieve better academic, social-emotional, and citizenship outcomes when exposed to critical pedagogy, with notable effects for students belonging to minority groups (Aronson & Laughter, 2016). Strategies to support the development of critical consciousness tend to be process-intensive (i.e., problem-based learning) and include open conversations about inequity and critical reflection. These activities help develop the ability to identify the "common sense" beliefs and assumptions that dehumanize and disenfranchise individuals belonging to non-dominant groups (Jemal, 2017; Watts et al., 2011). Critical consciousness raising activities create an environment where students are empowered to actively construct knowledge by "integrat[ing] activities that will enable learners to experience, reflect, think, explore, feel, and ask questions. (Junco et al., 2023, pp. 95). These processes further support deep learning by encouraging students to engage in applied complex tasks that require critical thinking with more than one possible outcome (Newmann et al., 1996; Reeves et al., 2000).

Critical consciousness raising is most often conceptualized as occurring through co-learning and other social processes, as modeling, discussion, and social support appear key processes (Jemal, 2017; Watts et al., 2011). As such, fostering critical consciousness inherently places students in a place of power where they can construct their own knowledge by critiquing injustice and imagining new structures for social justice (Flores, 2024; Morrison et al., 2008). Students also tend to engage in learner-centric activities such as content creation, social justice work, storytelling, and service-learning (de los Rios, 2020; Morrison et al., 2008; Thanapornsangsuth & Holbert, 2020).

Lastly, action is a prerequisite to critical consciousness raising as application fosters new skills and understanding both of social processes and individual identity as active citizens and change makers (Freire, 2000; Jemal, 2017). Critical consciousness raising within the classroom empowers students to take transformative action in their own lives and society by engaging in social justice work and providing opportunities for students to engage in community service and citizenship (Morrison et al., 2008). Critical consciousness practices in the classroom also make room for students to hold power in the class by making decisions on curriculum and assessments, as well as class policies (Flores, 2024; Gutstein, 2016). While the aforementioned strategies have been documented in some face-to-face classrooms, they are hardly standard practice, and there remains a gap in understanding how critical consciousness translates to online settings. Similarly, reviews of critical consciousness (i.e., Jemal, 2017; Watts et al., 2011) provide helpful theoretical refinement and support for how the pedagogical components described above foster critical consciousness. However, none includes a discussion of how these theoretical concepts or pedagogical strategies translate to online learning.

Envisioning Equity-based Online Learning

Pre-pandemic, K-16 online learning was most often used for enrichment and credit recovery models, with negligible to negative associations with learning depending on the program model and student population served (Fitzpatrick et al., 2020; Heinrich et al., 2020; Molnar, 2021). Post-pandemic, the rise in demand for online and blended education has sparked interest in further research on creating optimal learning environments online. Despite well documented reduced learning opportunities during the COVID-19 precipitated shift to virtual crisis-schooling (Goldhaber et al., 2022; Heinrich et al., 2022; Kuhfeld et al., 2022; Molnar et al, 2021), post-pandemic enrollment in full-time virtual schools has doubled compared to 2019 (Molnar et al, 2023). At the same time, the proportion of students being homeschooled has doubled, which experts believe is due at least in part to increased access to online curricula and instructional resources (Cheng & Hamlin, 2023; Musaddiq et al., 2022). As online learning becomes the default learning mode for an increasing percentage of students in the United States and is increasingly delivering content and facilitating instruction even within brick-and-mortar schools, deepening understanding of how to foster critical consciousness in this unique, often dehumanizing context is essential to enacting liberatory education.

Despite typical practices, there is potential to effectively enact one of the most prominent equity-based frameworks – Culturally Responsive Pedagogy (CRP) - when learning online. In a recent systematic review of 42 studies that enacted at least one component of CRP online, many interventions demonstrated high, culturally responsive expectations of students (the first CRP tenet) by (a) modeling and scaffolding challenging curriculum, (b) building on students' strengths and interests, (c) encouraging open communication, (d) and providing regular, supportive and formative feedback (Viano & Darling-Aduana, 2025). Many of the studies accomplished this first tenet of CRP by incorporating student-directed, problem-based, or inquiry-based learning. Within a liberatory framework, student-directed, as well as problem- and inquiry-based, learning expands choice and autonomy, empowers students as the primary actors in their own learning, and provides opportunities to apply their funds of knowledge authentically (Darling-Aduana, 2021; Gutstein, 2016). While both problem and inquiry-based learning are student-directed, inquiry-based learning engages students in developing and investigating self-generated questions (Pedaste et al., 2015), while problem-based learning encourages collaborative problem-solving to real-world challenges (Hmelo-Silver, 2004).

Additionally, 41 of the 42 identified studies documented a strategy to foster the second tenet of CRP, cultural competence (Viano & Darling-Aduana, 2025). Strategies included reshaping the curriculum (i.e., integrating multicultural content, highlighting POC as experts, emphasizing sociopolitical context), having students serve as experts, being responsive to students' backgrounds and needs, forming family partnerships, and revising narratives within the culture of power. There was also an emphasis on virtual simulations and digital storytelling to allow students of diverse backgrounds to critically reflect & use their voices to shape society for more equity, while also providing student-centered learning that promotes high academic outcomes.

However, the same study found that only seven studies within the sample fostered critical consciousness (Viano & Darling-Aduana, 2025). Five of those studies examined the role of power, while two encouraged critical action. The study called for more research that examined critical consciousness raising online, including expanding the definitions and theoretical orientations examined to provide a more holistic picture of practices beyond CRP. In this study, we aim to respond to that call and in doing so expand understanding of how our increasingly online world and educational system can better serve the needs and develop the competencies of all students using equity-based online pedagogy. By taking a wider theoretical lens than solely CRP, we identify a wider range of strategies for critical consciousness raising and bring into conversation interdisciplinary literature and communities with similar pedagogical orientations and goals.

Methodology

We conducted a systematic review guided by PRISMA 2020 (Page et al., 2021) to allow for a comprehensive examination of best practices and research gaps on fostering critical consciousness online. We identified themes across these studies and systematically summarized their findings. Below, we detail each component in this process including search strategy, study screening, data extraction and analysis, and limitations.

Search Strategy

We searched the ProQuest database (including ERIC), EBSCO's Academic Search Complete, and Google Scholar between August 15–18, 2024. Searches in ProQuest and Academic Search Complete were conducted across abstracts, keywords, and titles, while Google Scholar search was titles only. As such, Google Scholar was considered a supplemental search. The search term incorporated three conceptual components designed to identify studies at the intersection of formal education, critical pedagogies, and online learning: (education* OR pedagog* OR school* OR instruct* OR teaching*) AND ("critical consciousness" OR "critical literac*" OR "critical digital literac*" OR "power dynamics" OR "sharing power" OR "critiqu* soc*" OR "sociopolitical awareness" OR "sociopolitical consciousness" OR "social justice" OR "community action" OR "communitybased learning" OR "service learning" OR "activism" OR "liberatory" OR "resistance" OR "participatory" OR "student voice" OR "critical lens" OR "community partnership" OR "digital dialogue" OR "co-creat" OR "culturally relevant pedagogy") AND (synchronous OR asynchronous OR online OR virtual).

The first parenthetical set of terms captured research situated in formal educational settings, including both classroom-based and structured out-of-school contexts. The second set of terms reflected scholarship emphasizing critical consciousness and related frameworks—such as critical

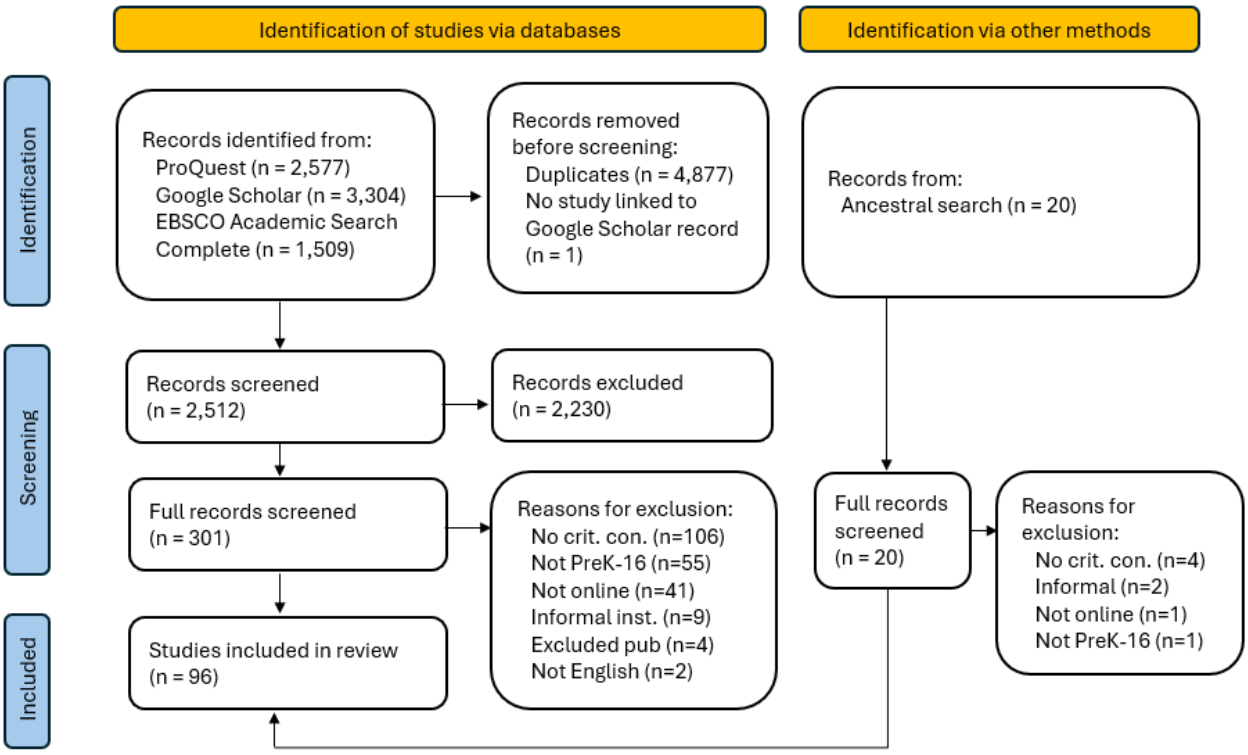
literacy, sociopolitical awareness, and community partnership. These keywords also included overlapping concepts commonly associated with equity- and justice-oriented digital education, such as "co-creation," "digital dialogue," and "community partnership." Finally, the third parenthetical restricted the results to studies addressing online or technology-mediated instruction, encompassing asynchronous, synchronous, and virtual modalities. To be classified as technology-mediated instruction, the technological tool or program had to deliver or otherwise be directly involved in instructional activities. For instance, using a Learning Management System (LMS) to access files did not count, but instruction delivered through a math video game did.

We limited all searches to English-language works published between January 2010 and August 2024, including peer-reviewed journal articles, dissertations, conference proceedings, and educational or governmental reports. The 2010 start date was selected to align with the transformation of online learning environments following the introduction of widely accessible, classroom-ready technologies such as the Apple iPad (Apple, 2010) and the first Chromebooks, which entered U.S. schools through pilot programs that same year (Google, 2021). This period marked a fundamental shift in the infrastructure and accessibility of online learning, supported by the nationwide expansion of broadband internet (Horrigan, 2010) and the Federal Communications Commission's E-rate Modernization Order (2014) to ensure broadband access in all U.S. schools. Limiting the review to studies published after 2010 ensured that the included research reflected the pedagogical, technological, and infrastructural landscape of contemporary online education.

The combined searches produced 7,390 total records (2,577 from ProQuest, 1,509 from Academic Search Complete, and 3,304 from Google Scholar). After removing duplicates and correcting for invalid citations, 2,512 unique studies remained for screening. See Figure 1 for the PRISMA flow diagram (Page et al., 2021). Authors double-coded the titles and abstracts of these studies to assess whether they met the three inclusion criteria using an Excel document: (1) focused on formal instruction in K-16 or equivalent educational settings, (2) examined online, virtual, or technology-mediated teaching and learning, and (3) included evidence of or potential for fostering critical consciousness or related constructs such as critical literacy, sociopolitical awareness, or community engagement. Raters assigned either a 1, indicating that the study met all three criteria, or 0, indicating that the study met fewer than three criteria. We then compared responses, tagging any studies where raters disagreed for review and reconciliation with a third rater and through discussion to clarify original ratings.

Figure 1

PRISMA Flow Diagram



After title and abstract screening, we retained 301 studies for full-text review. During this stage, we reapplied inclusion and exclusion criteria related to publication type, methodological rigor, and substantive focus. We excluded works outside PreK-16 education, those emphasizing informal or non-instructional contexts (e.g., social media participation, online forums unrelated to schooling), and non-academic sources such as magazines. (See Appendix A for the Full Text Coding Manual.)

Full-text inclusion decisions were double-coded using a Google Form to capture responses to the Full Text Coding Manual. Disagreements were resolved through discussion. Following full-text screening, 84 studies met all inclusion criteria. We added to these studies articles cited within the manuscripts in the full text review that appeared to meet inclusion criteria. We captured potentially relevant articles through a question on the Google Form. After which the article was assigned to two raters to double-code using the same process described above. This additional search yielded 12 new studies that satisfied inclusion criteria, resulting in a final sample of 96 studies for analysis. The final dataset thus represents over a decade of scholarship on PreK-16 online and blended learning, encompassing diverse geographic regions and methodological approaches.

Data Extraction and Analysis

We developed a structured data extraction protocol to facilitate consistent collection of information across the full-text corpus. Data extraction was implemented using a Google Form that aligned with our research questions. After piloting the form with sample studies, all the authors reviewed their coding for consistency and refined item wording for clarity and comprehensiveness. Once finalized, all remaining articles were coded using the refined Google Form. Each article was double-coded, with at least two researchers independently completing the form for each study. We reviewed and merged responses from both coders into a single composite record per study, adding a third coder to reconcile any discrepancies.

Using the form, we first coded descriptive characteristics of the study including publication type, research design, grade-level, instruction context, online learning type, and COVID-19 timeframe. Next, we coded each article for the presence of critical consciousness raising strategies that emerged through our conceptual review of prior literature on critical consciousness (see Table 1).[1] Importantly, while we coded each study for whether digital content creation was present based on its prominence and importance in our conceptual review, we decided to summarize findings from this strategy under the unique affordances of critical consciousness raising online section, since it is not inherently a strategy to foster critical consciousness. We report the presence of each strategy within our study sample in the Finding section. Study-by-study coding is available in the supplementary materials.

Table 1

Strategies to Foster Critical Consciousness Coded for during Full Text Inclusion

Strategy	Definition
Power	Are students asked to engage with power?
Critical Literacy	Is there evidence of critical literacy "inviting students to question, examine, and even dispute the power relations between writers and readers" (Morrison et al., 2008 pp. 441)?
Critical Digital Literacy	Is there evidence of critical DIGITAL literacy (i.e., encouraging students to deconstruct the meanings, uses, and messages of digital artifacts to make judgments about online tools and content and/or apply ethical perspectives and interrogate power dynamics present in the creation and use of digital artifacts)?

[1] We considered studies to meet the criteria for fostering critical consciousness if at least one strategy was present. However, as we considered engaging with power to be a core tenet of fostering critical consciousness, there was a prompt to re-evaluate whether critical consciousness was present if the rater indicated that students were not asked to engage with power. Ultimately, most studies coded as demonstrating elements of one strategy were coded as meeting others as well given commonalities in the underlying skill set and assignments integrated.

Strategy	Definition
Digital Content Creation[2]	Is there evidence of digital content creation?
Social Justice Skills	Is there evidence of developing the skills and socio-political awareness to engage in social justice work (i.e., by contextualizing academic content in social justice topics)?
Classroom Change	Is there evidence of students enacting change in their classrooms (i.e., co-designing curriculum, reshaping power dynamics in the classroom)?
Community Change	Is there evidence of students enacting change outside their classrooms (i.e., engaging in social justice work or providing services)?

Qualitative Analysis

In addition to calculating descriptive statistics of quantitative coding responses, we exported all narrative descriptions of instructional strategies from Google Forms into Dedoose (Version 10.0.35), a secure web-based platform for mixed methods data analysis. Within Dedoose, we uploaded and organized the extracted text segments to facilitate inductive coding, theme development, and analytical memo writing. Two authors completed an initial cycle of open coding to identify emergent patterns in how studies operationalized elements of critical consciousness online. These initial codes were then refined through team discussions and memo exchanges, which allowed for collaborative comparison and synthesis of ideas.

Through iterative coding cycles, we generated preliminary themes describing the ways in which critical reflection, political efficacy, and critical action were fostered in digital environments using the strategies identified. We then used analytic memos to synthesize these themes across studies, noting both convergences and contextual differences. This process yielded clearly defined, recurring patterns in how online learning settings supported critical consciousness raising. We used the resulting narrative descriptions to provide rich detail on promising instructional strategies. In addition, we used qualitative data to provide nuance, understand patterns in the quantitative data, and consider alternative hypotheses and understandings.

Findings

We first present descriptive statistics of the studies included in the systematic review to define our sample and highlight trends in prevalence by category. Of the 96 studies included in the final synthesis (see Table 2), most were peer-reviewed journal articles (78.1%), followed by doctoral dissertations (11.5%), and a smaller proportion of book chapters (5.2%), conference proceedings (4.2%), and books (1%). Nearly half of the studies employed intervention designs (49%), and another 42.7% used observational approaches, with only a few conceptual or review pieces. In terms of educational level, undergraduate (47.9%) and high school (30.2%) contexts were most

[2] During our analysis, we determined this was most appropriately discussed under the section on unique affordances of critical consciousness raising online, as digital content creation without a critical framework was insufficient to foster critical consciousness on its own

common, while elementary and middle-level settings represented smaller portions of the corpus. We did not identify any studies examining critical consciousness enactment in pre-Kindergarten through second grade. The majority of studies (78.1%) were conducted in formal classroom environments. Approximately one-fifth (21.9%) took place in structured out-of-school instructional contexts. Regarding online modalities, synchronous learning was the most prevalent (46.9%), followed by in-person instruction that integrated online tools (32.3%), asynchronous courses (16.7%), and a small number of face-to-face interventions that taught digital literacy without online tools (4.2%). Nearly half of the studies were conducted before the COVID-19 pandemic (46.9%), with the rest of the studies conducted during (40.6%) and after (12.5%) COVID-19.

Table 2

Included Study Characteristics

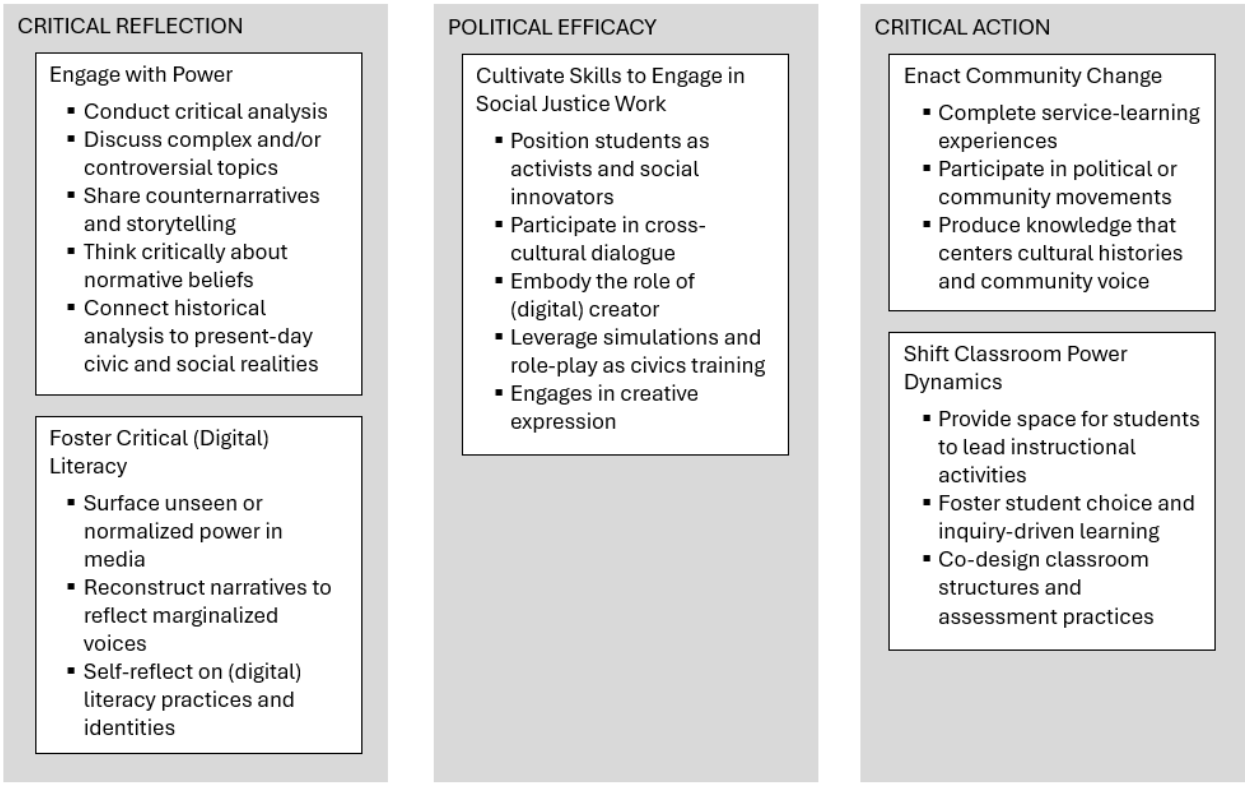
Domain	Category	Number	Percentage
Study Type	Peer-reviewed journal	75	78.1
	Book chapter	5	5.2
	Book	1	1.0
	Doctoral dissertation	11	11.5
	Peer-reviewed conference proceeding	4	4.2
Research Design	Review	1	1.0
	Observational	41	42.7
	Intervention	47	49.0
	Conceptual / Theoretical	7	7.3
Grade Level	Elementary (3–5)	9	9.4
	Middle (6–8)	12	12.5
	High school (9–12)	29	30.2
	Undergraduate	46	47.9
Instructional Context	Formal instruction, in a classroom environment	75	78.1
	Formal instruction, in an out-of-school environment	21	21.9
Online Learning Type	In-person instruction using a fully online tool or resource	31	32.3
	Fully asynchronous online	16	16.7
	Synchronous online	45	46.9
	Face-to-face on the topic of digital literacy	4	4.2
COVID-19 Timeframe	Before COVID-19	45	46.9
	During COVID-19	39	40.6
	After COVID-19	12	12.5
Critical Consciousness Strategies	Engage with power	93	96.9
	Social justice skills	75	78.1

Domain	Category	Number	Percentage
	Critical literacy	44	45.8
	Critical digital literacy	39	40.6
	Community change	32	33.3
	Classroom change	27	28.1
	Digital content creation	62	64.6

Next, we report descriptive analysis of the 96 studies included in the review by critical consciousness strategy. Overall, 97 percent of studies documented instructional strategies that asked students to engage with power. The majority also cultivated the skills to engage in social justice work (78 percent). We observed elements of critical literacy in 46 percent of studies and critical digital literacy in 41 percent of lessons. The least frequently observed elements included enacting community change (i.e., change outside of the classroom) (33 percent) and classroom change (i.e., shifting classroom power dynamics and processes to give student voice) (28 percent). We identified key strategies and insights used to foster each of these elements of critical consciousness in Figure 2. We discuss and provide examples for each strategy in greater detail below. Additionally, 65 percent of studies incorporated digital content creation (other than Word documents or essays), which we discuss in in the unique affordances and enhancements of critical consciousness raising section.

Figure 2

Strategies to Foster Critical Consciousness Online



Strategies that Engage with Power

As a core element of critical pedagogy, 97 percent of studies asked students to engage with power. The three studies without this element all detailed critical literacy projects that incorporated critical action to address environmental and globalism related issues, but without identifying systemic inequities as a contributing factor. In this way, the studies that did NOT ask students to engage with power did support students in developing political efficacy and taking action without requiring critical reflection (Jemal, 2017; Watts et al., 2011). Strategies to support students in engaging with power and inequality included (a) conducting critical analysis, (b) thinking critically about normative beliefs, (c) sharing counternarratives and storytelling, (d) discussing complex and/or controversial topics, and (e) connecting historical analysis to present-day civic and social realities. We discuss each in greater detail below.

Around half of the studies documented students **conducting critical analysis** as a means to make visible and critique systems of power and inequality. Critical analysis often included introducing students to an applicable theory, engaging with the theory through real-world or virtual activities, and forming informed, individual opinions and reflections (i.e., Dancis, 2024; Kumar, 2024; Muriel & Acosta, 2015; Porto, 2018). For instance, one undergraduate course grounded critique in political philosophy (Kaye & Spurgin, 2011). Students examined John Rawls' Theory of Justice alongside real-world inequalities supplemented by virtual simulations in Second Life to

model civic engagement and resource distribution (Kaye & Spurgin, 2011). Also with undergraduates, Wittebols (2016) describes a series of iterative assignments designed to build analytical power that transfer to civic judgement. Assignments included bias self-audits and cross-source research. In another study, middle school students examined U.S. foreign policy and military interventions through the Just War Project (Lawrence & O'Brien, 2016). The project provides expert video interviews, articles, and training to support ethical decision-making in modern warfare, situating global systemic critique in an accessible classroom format (Lawrence & O'Brien, 2016). We observed the use of similar strategies with elementary school students although at a more basic level. For instance, one fifth grade teacher used a series of blogging exercise to develop students' critical literacy skills (Chamberlain, 2017). Particularly relevant to critical analysis, students were prompted as part of their blog writing and commenting to recall previous knowledge (engaging funds of knowledge), make connections, decide what was important to remember from the text, and articulate an opinion (Chamberlain, 2017). All these experiences positioned students as critical investigators of social systems rather than passive learners, providing a process for students to apply their own experiences and form their own opinions on class content.

Relatedly, a little over half of all studies prompted learners to **think critically about normative beliefs**. By naming and questioning "common sense" assumptions—about knowledge, identity, institutions, media/tech, and civic life—interventions supported students in developing more complex, justice-oriented understandings of the world (Hicks, 2022; Philpot et al., 2024). Assignments emphasized fact-checking, rewriting, or deconstructing dominant narratives (Hilton, 2013; Huang, 2014). In one study, students re-wrote one-way policy documents as two-sided policy discourses (Garcia & Mirra, 2021). Media literacy projects taught students to spot bias, resist stereotypes, and critique power in texts (Nash, 2021; Wittebols, 2016). One assignment took an additional approach to encourage critical reflection on normative beliefs by asking students to take and reflect on their responses to the Harvard Implicit Bias test (Rippy & Muñoz, 2019). Classroom activities placed results in their larger sociopolitical context, discussing relevance to future careers and social issues. Similarly, a simulation exercise encouraged deeper reflection on systemic barriers; the teacher randomly provided some groups with truncated instructions for a group project (Schnurr, 2013). Students rated and discussed the merits of each project before being told about the experiment and asked to reconsider in light of the different instructions provided.

Importantly, around one-fourth of interventions involved **sharing counternarrative and storytelling** – in digital, oral, written, and performative forms - to help students surface, interrogate, and contest power dynamics. These interventions highlighted how personal narratives made structural inequities tangible. Students' artistic and multimodal projects — such as artwork expressing marginalization (Murray et al., 2024), spoken word and performance of lived histories (Kaler-Jones, 2021), or podcasts, comics, and digital storytelling about disability, religion, identity, and belonging (Matias & Grosland, 2017; Murray et al., 2024; Price-Dennis et al., 2015) — exemplify how creative practices enabled participants to communicate private, painful, and often silenced experiences. In Catena (2020), a high school journalism project, students developed first-person narratives that functioned as civic texts—using personal stories to address community issues and public concerns. These projects both humanized abstract issues like discrimination and created emotional entry points for peers and broader audiences to grapple with inequities. In these instances, storytelling was used not only as a pedagogical tool but also as a form of civic intervention, validating lived experience as knowledge alongside academic or statistical data.

Another strategy for encouraging reflections on power – observed in a little under one-fourth of studies - included **discussing complex and/or controversial topics** such as systemic inequality, race, gender, and economic injustice. Rather than avoiding contentious issues, these interventions framed dialogue, debate, and critical reflection as central to the learning process. Across settings and levels, students were prompted to confront privilege and oppression. Undergraduate courses tasked students with "Difficult Dialogues" assignments that developed skills in challenging conversations (Hicks et al., 2023) or structured discussions on systemic inequality and sociopolitical awareness (Weiner et al., 2020). Similarly, when working with younger students, a virtual Freedom School used Children's Defense Fund-approved texts that centered the experiences of individuals from marginalized groups and on social justice topics, with guided discussion prompts and teacher modeling to encourage discourse on these complex topics (Harris et al., 2025).

Lastly, eight particularly impactful studies detailed interventions that **connected historical analysis to present-day civic and social realities**. Across diverse grade levels, contexts, and online tools, students were guided to interrogate how historical narratives—and their exclusions—continue to shape inequities in contemporary life (Mills et al., 2011). These interventions spanned middle school through undergraduate levels. Many of the interventions were synchronous (e.g., live discussions, performances), reflecting the importance of dialogic, collective processing when working with contentious histories (Şenbayrak & Hart, 2024). Yet asynchronous spaces (e.g., online archives, wiki projects) also emerged as important venues for preserving and sharing historical narratives (Boetto & Bell, 2015).

At the undergraduate level, history was explicitly mobilized to expose systemic inequities in professional and disciplinary contexts. For instance, technical writing courses interrogated the systemic inequities and gender-based violence (Jones, 2021), while health-focused courses examined how historical exclusion of marginalized groups underpins mistrust in the healthcare system today (Donovan & Barcus, 2021). Both cases illustrate how embedding historical perspectives in disciplinary content can challenge students' assumptions about neutrality in professional fields. For middle and high school students, history served as a mirror for civic engagement. Case studies, conflict analyses, and social movements were analyzed not only to understand the past but also to build civic literacy and consider strategies for contemporary conflict resolution (Lawrence & O'Brien, 2016). In science classrooms, students were asked to interrogate why the "canon" of famous physicists is dominated by white men, sparking discussions about exclusion from STEM and the importance of diverse representation (Ross & Grant, 2022). These examples highlight how historical narratives — who is remembered, who is omitted — shape students' understanding of justice in present domains. These approaches did more than transmit historical knowledge — they emphasized history as living and actionable, a resource for shaping present civic identities.

Strategies that Cultivate Skills to Engage in Social Justice Work

While asking students to engage with power supports critical reflection (the first tenet of critical consciousness), developing political efficacy (the second tenet of critical consciousness) requires fostering the skills to engage in social justice work. In all, 78 percent of studies integrated strategies that fostered students' capacity to move from reflection to action, developing the confidence and agency to enact social and political change. Some of the strategies used to cultivate the skills to engage in social justice work within our study sample included (a) positioning students

as activist and social innovators, (b) participating in cross-cultural dialogue, (c) embodying the role of (digital) creator, (d) leveraging simulations and role-play as civics training, and (e) engaging in creative expression.

By embedding social justice within subject content, educators moved beyond identifying and critiquing power dynamics to emphasizing the need for action. Several interventions used identity-based pedagogies to **position students as activists and social innovators.** Jones (2021) integrated victim advocacy and critique of the justice system into undergraduate learning. Thompson (2012) guided teacher candidates to integrate disability justice into broader social justice agendas. Another study documented student reflections on the activist role of social media in mobilizing around #BLM (Beck et al., 2021). High school students connected inequities in healthcare to advocacy organizations and campaigns (Mills et al., 2011) and proposed social entrepreneurship projects to address social issues in health care that they had the opportunity to share in a science fair format (Huang, 2014). High school students in Baltimore learned from museum activists and visited museums to critique and suggest ways to expand access, diversity, and the relevance of exhibits (Zeleny, 2023). Environmental projects, such as the Watershed Council initiative, developed middle schoolers' awareness of the political dimensions of ecological issues and helped them identify actors who should be involved in solutions (Marques et al., 2020).

Several interventions support the development of political efficacy through student **participation in cross-cultural dialogue**, either through global exchanges or by examining a global perspective. For instance, elementary students engaged in cross-cultural dialogues to build empathy and global awareness (Kerkhoff et al., 2024). In another study, Şenbayrak and Ortaçtepe Hart (2024) gave undergraduates a platform for global civic dialogue through language learning. While Dawson and Robinson (2021) fostered embodied learning tied to activism, blending online and offline spaces with social justice issues.

Additionally, a strong cluster of studies focused on computing programs designed for marginalized groups. These interventions encouraged students to **embody the role of (digital) creator**. Several studies described culturally responsive computing initiatives for high school students, particularly adolescent girls of color in urban districts (Lee et al., 2021; Scott et al., 2010; Scott & White, 2013). For instance, the COMPUGIRLS program reconceptualized technology as raced, gendered, and classed, framing digital education as a site of identity affirmation while positioning students as creators of technology for social justice (Scott & Garcia, 2016). Similarly, da Silva and colleagues (2024) aligned course design with culturally responsive practices by encouraging students to share art that represented their daily lives, reflect on the historical precedent for normative preferences in landscape design, and re-envision outdoor spaces together. In doing so, the activity affirmed injustices while offering inspiration for action.

Other interventions **leveraged simulations and role-play as civics training**. These interventions used immersive experiences to model political or social change (Basile, 2021; Brown et al., 2021; Levett-Jones et al., 2015; Kaler-Jones, 2021; Rippy & Muñoz, 2019; Schnurr et al., 2013). For instance, Shliakhovchuk (2023) employed a refugee simulation game, connecting gameplay to concrete activism (donating, sharing, advocacy). Bell et al. (2015) had undergraduates use economic justice simulations to explore structural inequity and fairness, while Huang (2014) assigned students to play roles (e.g., mayor, police chief) in community decision-making, prompting reflection on power and action. Games, simulations, and issue-driven cases provide

opportunities for students to delve deeply into the contextual factors and systemic pressures of a specific case. Simulations and games show structural patterns beneath personal experience.

Interestingly, **engaging in creative expression** often functioned as a bridge between reflection and advocacy. Kaler-Jones (2021) engaged high school students in arts-based performance as a call to action. The Black Girls Movement (BGM) program also encouraged middle school participants to use performance-based expression (spoken word, public storytelling, artistic presentations) as a form of advocacy and healing in their communities (Smith-Purviance et al., 2022). Price-Dennis (2016) worked with elementary students to produce stop-motion films on gender inequality. Brown and colleagues (2021) required undergraduates to create photo essays on health inequities, while da Silva and colleagues (2024) encouraged online students to use visual culture projects to address inequities and propose solutions. In Ireland, 11- to 16-year-olds participated in a socially engaged digital art summer program that used storyboarding, sketching, and script writing to engage with complex issues; the program ultimately led to student-produced newspaper articles and memes (Murray et al., 2024).

Strategies that Foster Critical (Digital) Literacy

The next most common strategies for critical consciousness raising included critical literacy (46 percent) and/or critical digital literacy (41 percent). In total, 59 percent of studies engaged with one or both frameworks. Critical literacy and critical digital literacy are overlapping, related concepts from different subfields with the primary difference that critical digital literacy specifies the media format examined. Therefore, although we separated these elements when identifying studies to ensure we captured studies employing either approach, we summarize strategies across both elements here due to substantial overlap, since our sample was inherently limited in its focus on online learning. Strategies used to foster critical (digital) literacy online included (a) surfacing unseen or normalized power in media, (b) reconstructing narratives to reflect marginalized voices, and (c) self-reflecting on (digital) literacy practices and identities. Digital content creation was also a crucial element of critical digital literacy enactment, which aims to support students in developing an identity as digital creators. However, since content creation was enacted within various frameworks and to support several tenets of critical consciousness, we summarize this element separately in a later section.

Almost all studies that fostered critical (digital) literacy did so by helping learners **surface unseen or normalized power in media**, such as political and algorithmic bias, gatekeeping, datafication, curricular canons, and platform economics. An understanding of edtech power and datafication was a fast-growing theme within the data. Learners surface algorithmic and institutional power that usually feels "neutral." Students reject platform neutrality, linking datafication to how school systems "reduced their participation… to a static metric" (Pangrazio et al., 2022). Several studies framed digital platforms (e.g., YouTube, social media, online discussion forums) as both tools for learning and objects of critique. In the Critical Media Creators Club, middle school students investigated YouTube algorithms, questioning why certain videos appeared and whether targeted ads reflected or distorted their identities (Adams-Grigorief, 2023). Students scrutinized authorship, frames, and omissions to reveal who is centered and who is erased. Across interventions, students used structured, repeatable prompts to fact-check and evaluate media and other digital platforms (Huang, 2014; Nash, 2021). In one classroom, students identified racist/sexist ads "not… noticeable at first glance" (Rippy & Muñoz, 2019). In other assignments, undergraduates are prompted to ask, "who pays for it?" and "who held power within

the text" to help identify media–political economy ties (Wittebols, 2016, pp. 7). These studies highlight that students learned not only how to develop and use (digital) literacy skills across a range of relevant contexts, but also why and for whom media and online platforms function as they do.

Within this category, several programs explicitly centered identity, intersectionality, and epistemic power to reveal how knowledge gets authorized. COMPUGIRLS made visible the "power asymmetry" in tech design (Scott & Garcia, 2016). Integrating Decolonial/Indigenous lenses exposed canon power and epistemic dominance (Ripero-Muñiz et al., 2021; Dawson & Robinson, 2021). While international and global citizenship interventions used cross-cultural comparison to name asymmetries and privilege (Waghid, 2024; Sernbo et al., 2024; Fuchs et al., 2010).

Beyond identifying media bias, several interventions asked students to **reconstruct narratives to reflect marginalized voices**. For instance, high school students "analyzed journalistic conventions" before creating new media that better represent the narratives, perspectives, and experiences of individuals belonging to marginalized communities (Catena, 2020). Others engaged high school students with policy reports to reveal how official documents framed their communities, then produced fuller portraits of their schools as acts of counter-storytelling (Garcia & Mirra, 2021). These activities foregrounded students' ability to challenge deficit-based narratives and recenter community perspectives.

In addition to engaging with and critiquing (digital) literacy, several interventions encouraged students to **self-reflect on (digital) literacy practices and identities**. Nash (2021) guided students to ask culturally grounded questions such as "how does my own culture or background influence the links I click on?", encouraging reflexivity and relevance in digital literacy practices. The COMPUGIRLS program positioned technology as a "vehicle by which students reflect and demonstrate understanding of their intersectional identities" (Scott & Garcia, 2016). Programs like Black Girls S.O.A.R. (Kaler-Jones, 2021) and other feminist digital pedagogy projects (Eudey, 2012; Donovan & Barcus, 2021) used digital storytelling and activism to connect students' lived experiences with broader struggles against systemic inequality. Such work foregrounded race, gender, and cultural identity as necessary entry points for digital literacy, reframing learning as both personal and political. Taken together, this body of work demonstrates how digital literacy education can become transformative when it explicitly embraces the tensions of power, identity, and inequality. By situating students in spaces where reflection, discomfort, and critique are expected, these programs helped learners not only deepen media literacy but also understand its roots in systemic power relations.

Through strategies aimed at fostering critical digital literacy specifically, students learned that algorithms shape identity, representation, and visibility, raising questions about fairness, bias, and digital selfhood. These studies discuss how digital literacy was positioned not just as a technical skill, but as a civic responsibility, equipping students to interrogate how online tools mediate democracy and justice. In addition, digital projects – discussed in greater detail in a following section - emphasized production (videos, stories, collaborative digital artifacts) as a pathway to empowerment and to foster equity, representation, and democratic representation.

Strategies that Enact Community Change

Beyond critical reflection and political efficacy, a key component of critical consciousness is critical action. In total, 33 percent of identified studies integrated activities that extended students' learning into civic and community spaces. These projects positioned students as civic actors, educators, and advocates. Some of the strategies used to involve students in critical action in their communities included (a) completing service-learning experiences, (b) participating in political or community movements, and (c) producing knowledge that centers cultural histories and community voice.

The most prevalent strategy for critical action in our sample, particularly in undergraduate classrooms, involved **completing service-learning experiences**. Assignments went beyond volunteering to explicitly address structural inequities through structured community service experiences connected to and supported by classroom learning on topics such as health disparities (Brown et al., 2021), food insecurity (Muriel & Acosta, 2015), and racial justice (Rapp, 2023). University students partnered with local or global communities (Lyons & Howarth, 2022). Several service-learning experiences, particularly during the COVID-19 pandemic, were conducted virtually, whether through advocacy websites tied to the Black Lives Matter movement (Ordem, 2023) or multimodal campaigns shared with the public (DeWaard & Roberts, 2021). By situating students as both learners and change agents, programs disrupted traditional hierarchies of service, reframing it as reciprocal, justice-oriented, and transformative.

While service-learning courses tended to provide semester-long community service experiences aligned with classroom instruction, other interventions provided the opportunity for (often shorter) direct **participation in political or community movements.** Projects tied to the Black Panther Party's political framework (Kaler-Jones, 2021) or to transnational collaborations (Lyons & Howarth, 2022) underscored how online tools and assignments can connect learners to broader struggles for justice. Undergraduate and high school students engaged in structured community service experiences—addressing racial justice through oral history projects (McDaniel, 2024), supporting food pantries (Muriel & Acosta, 2015), or creating digital campaigns tied to Black Lives Matter (Ordem, 2023). High school students also engaged in local civic projects such as online petitions, protests, and campaigns for social issues (Dias-Fonseca & Potter, 2016). Similarly, Giunco (2020) described middle school students synthesizing their learning into public speeches with calls to action, inviting community members to join their advocacy efforts, while elementary school students participated in a National Day of Social Action on voting rights (Harris et al., 2025). Students also organized protests, engaged in public storytelling, and shared digital products with family, schools, and broader communities (Price-Dennis et al., 2017).

In many cases, assignments explicitly required students to engage in digital activism — from community media campaigns (Murray et al., 2024; Amgott, 2018) and advocacy websites (Ordem, 2023) to digital storytelling projects addressing racism, food insecurity, and environmental justice (Brown et al., 2021; Muriel & Acosta, 2015; Lyons & Howarth, 2022). Amgott (2018) described undergraduates designing online media campaigns about food insecurity that included educational resources and actionable steps for community response, which they shared publicly online. Other instructors treated online petitions, campaigns, and multimodal presentations as authentic political action (DeWaard & Roberts, 2021; Dias-Fonseca & Potter, 2016; Lee et al., 2021). In doing so, these activities connected students to larger social movements, providing an avenue for broader community engagement.

Many critical action interventions also encouraged students to educate others by **producing knowledge that centers multicultural histories and community voices.** Students often completed these activities in collaboration with museums, community centers, and other organizations with the means to disseminate this knowledge. In other interventions, students contributed to online archives such as Civic Voices (Mills et al., 2011, Marques et al., 2020). For instance, students created digital storytelling projects that documented the experiences of underrepresented groups, connecting historical struggles to current social issues (Marques et al., 2020). Multimedia outputs (infographics, digital campaigns, and online debriefs) allowed students to circulate historical insights beyond the classroom, engaging broader publics in debates about justice. These examples highlight how community-centered knowledge production foregrounds cultural histories as sources of civic insight and resistance (Ordem, 2023; Lee et al., 2021).

Strategies that Shift Classroom Power Dynamics

The final way interventions engaged students in critical action was by empowering students through shifting classroom power dynamics and structures. This was the least observed strategy to foster critical consciousness with only 28 percent of studies in our sample enacting classroom change. At the same time, programs that disrupted traditional classroom hierarchies stood out as especially impactful. In these studies, many of the following normative educational practices were reimagined: positioning the teacher as "giver of knowledge", prioritizing respect for authority, and expecting all student to engage in the same teacher-directed and designed activities. Some of the strategies used to shift these classroom power dynamics included (a) providing space for students to lead instructional activities, (b) fostering student choice and inquiry-driven learning and (c) co-designing classroom structures and assessment practices.

One of the most prevalent strategies used to shift classroom power dynamics included **providing space for students to lead instructional activities**. Taking leadership in instructional delivery could look like co-leading lessons, facilitating discussions, or organizing guiding projects. In one high student class, students led class dialogues and shaped instructional flow using online blogging (Muriel & Acosta, 2015). Middle schoolers led international citizenship projects, producing deliverables and exchanging leadership roles across borders (Porto, 2018). In another example, Spanish-speaking students' linguistic and cultural skills became recognized sources of instructional authority by serving and translators and team leads in collaborations with Columbia partners, while non-Spanish-speaking students were asked to contribute more through reflection on how narratives, values, and language may need to be presented differently to different cultural groups (Lyons & Howarth, 2022).

Across contexts, teaching strategies also shifted classroom power dynamics by **emphasizing student choice and inquiry-driven learning** as ways to empower students to direct their own learning. Elementary school students completed independent oral history projects on voting rights and inequities exposed by COVID-19 (Harris et al., 2025). Similarly, youth research that integrated their lived experiences into topic selection, asset-building, and public-facing communication (Scott et al., 2010) reinforced that student choice and co-curation can challenge deficit narratives.

Lastly, some studies documented teachers and students **co-designing classroom structures and assessment practices**. Shifting who decides what is taught and how learning happens allows students to experience power-sharing firsthand. Practices such as self-grading, co-designing course goals, and peer lectures enabled students to directly shape learning outcomes (Philpot et al., 2024;

Dancis, 2024). Students also developed case studies, debate topics, and mini syllabi based on their own sociopolitical interests (Sernbo et al., 2024; Lawrence & O'Brien, 2016; Ordem, 2023). Dancis (2024) used several democratic power-sharing and transformative accountability practices; students co-designed syllabi, self-graded, and planned collective action projects. Marques et al. (2020) demonstrated how middle school students collaboratively designed environmental solutions, merging curricular and civic agency.

One commonly observed structure for co-design was Youth Participatory Action Research (YPAR), which empowered students to investigate and challenge local power structures (Amgott, 2018). Full realization of YPAR tenets included opportunities to act on research findings—such as students analyzing pandemic-related disparities and proposing school policy solutions (Blue & Redick, 2022) or building advocacy units and community collaborations (Giunco, 2020). Participatory justice research in elementary curricula likewise framed young learners as co-researchers investigating fairness, equity, and community issues in their own school contexts (Little et al., 2024). In Little et al. (2024), children and adults collaboratively explored whose voices were heard or silenced in classroom decision-making, demonstrating that participatory justice research can surface power relations even in early schooling. In other studies, students acted as co-researchers, conducting investigations of digital platforms and sharing findings (Guo, 2019). Across grade levels and modalities, co-creation strategies disrupted traditional hierarchies, validated students' lived experiences, and fostered inclusive and democratic learning environments.

Unique Affordances of Online Tools & Settings in Fostering Critical Consciousness

In addition to documenting the form and prevalence of strategies to foster critical consciousness online, we examined whether there were any unique affordances to fostering critical consciousness online. Across the studies in our sample, only five used online technologies in a manner that could not be modified to achieve similar goals in face-to-face instruction. Although small in number, these studies highlight particular affordances within online learning that enabled forms of critical engagement that extend beyond what can be achieved in traditional face-to-face instruction. These affordances included expanded access to public sharing and digital dialogue, multimodal authorship, and digital cultural repositories.

One critical digital literacy assignment that could not have achieved a similar goal without the use of online tools asked K-12 students to use Blendspace, Flipgrid, and Spotify to curate and circulate counter-narratives on race and justice. Blendspace interactive digital boards allowed students to access, engage with, and juxtapose multimodal artifacts—news articles, images, and videos—to critique dominant media representations (Pangrazio et al., 2022). Not only did these platforms enhance the ease, impact, and sharing of comparisons, but the media's virtual nature also necessitated the use of online tools.

Three studies leveraged the increased ease of sharing digital content to expand impact. For instance, one activity conducted with undergraduate students used virtual museums and interactive archives synchronously in a classroom environment to democratize access to cultural institutions (Zeleny, 2023). Students navigated and annotated digital exhibits, reframing museum literacy as civic activism by curating their own counter-collections—something constrained by logistical and spatial limits of in-person museum visits (Zeleny, 2023). In another study, high school students used digital petition platforms and moderated youth forums to expand their audience beyond the classroom, transforming critique into consequential public authorship (Dias-Fonseca & Potter,

2016). These online tools connected local inquiry projects to authentic civic spaces, enabling visible participation in public discourse that could not be simulated through in-class debates or school displays.

Another intervention also leveraged the unique affordance of online platforms, but this time to provide safe, identity-affirming spaces where Black, middle school girls blended self-expression with social critique. The digital format allowed participants to control pacing, edit their contributions, and engage in reflective exchanges without the surveillance and social pressures often present in in-person discussions (Price-Dennis & Muhammad, 2021). These digital dialogues were used as an online supplement to traditional, face-to-face instruction to minimize some of the barriers and judgement of participation for a minoritized student group in a way that offered support without stigma.

Lastly, in one intervention, high school students in an out-of-school environment engaged in coding and game design using fully online tools. The program positioned girls of color as technologists and social theorists. The iterative digital creation process—debugging, testing, and remixing—allowed learners to continuously revise representations of identity and power within their games, an affordance not easily transferable to face-to-face analogs (Scott et al., 2010). Additionally, the focus on inherently digital skills, such as coding, requires interfacing with online tools (if not fully online).

Online Tools & Settings Enhancements to Fostering Critical Consciousness

In addition to examining interventions that could not be enacted without the use of online tools or in online settings, we explored interventions that could be enacted in a traditional, face-to-face setting, but only with substantial modification. While not unique to online contexts, the resulting elements highlight some of the ways online learning can enhance current strategies to foster critical consciousness in the classroom. Collectively, these studies illustrate that affordances such as simulations, cross-cultural collaboration, critical media inquiry, and digital content creation uniquely enhance critical consciousness raising in online environments by extending dialogue across distance, democratizing participation, and embedding reflection into multimodal practice.

As one example of effective simulation use, undergraduate nursing students engaged with Wiimali, a simulated community that used reflective exercises and narrative immersion to help students examined how bias and structural injustice shape patient experiences (Levett-Jones et al., 2015). In-person, similar learning could be achieved through community-based simulations, role-plays, or case-based storytelling circles that preserve the reflective and dialogic engagement central to Wiimali. However, the online platform reduces the burden on educators to design and facilitate the simulation.

Similarly, in a university environmental humanities course, students engaged in cross-cultural collaboration with Colombian community partners to create bilingual digital stories addressing environmental justice (Zamparutti, 2022). The online, multilingual format enabled cross-border dialogue and public authorship through a digital storytelling platform that extended learning beyond the university. Adapted for in-person settings, comparable critical exchanges could be structured through joint workshops with local community partners, gallery exhibits of student-created bilingual narratives, or cross-campus exchanges that retain the principles of co-authorship and global solidarity.

In another study, students in a critical media literacy course developed news analysis projects to uncover bias and political economy in contemporary media (Wittebols, 2016). Online assignments required students to track their own confirmation biases, analyze news sources, and critique how corporate and political power shape public discourse. Similar outcomes could be fostered in-person through structured media labs, peer analysis sessions, and collaborative news mapping activities that preserved the course's reflective and dialogic dimensions. However, enacting similar activities in a face-to-face environment would likely have required more time and teacher (versus student) facilitation.

Ultimately, however, we identified digital content creation as the most consistent enhancement to fostering critical consciousness online. Digital content creation is a core component of fostering critical digital literacy. However, it can also be integrated in classroom activities outside of a critical digital framework to engage with power dynamics and/or critical action. Creating digital content is an inherently technology-based process that situates students as empowered creators, critics, and change makers. Emphasizing its centrality, 61 percent of identified studies integrated some element of digital content creation. Studies used media making to help sharpen students' perceptions of what is usually invisible, construct and exert empowered identities, and extend student work beyond the classroom to authentic audiences.

Several studies used student-created film, photography, and digital art to challenge social issues (DeWaard & Roberts, 2021; Thompson et al., 2022). In particular, digital storytelling projects allowed students to use online tools to explore representation, identity, and power. These projects included video creation, Jamboard brainstorms, and community-based digital stories around health, race, and equity (Chappell & Hetherington, 2024 Murray et al., 2024; Rapp, 2023). In another study, students used digital platforms like Pinterest, Prezi, and hip-hop media to generate critique & counter-narratives (Price-Dennis et al., 2017). These activities enabled students to connect personal experiences with broader systemic issues through emphasizing voice and perspective-taking. Engaging in these activities shifted students from consumers of media to critical creators by emphasizing how storytelling provided students with agency to create digital counter-narratives.

Digital content creation was also used to construct and exert empowered identities. The COMPUGIRLS program, for instance, invited high school girls of color to challenge deficit narratives by constructing empowering online identities (Scott & White, 2013). Similarly, projects in undergraduate classrooms asked students to reflect critically on memoirs of sexual assault (Lawrence & O'Brien, 2016) or to co-create antiracist classroom cultures by acknowledging BIPOC students' experiences of exclusion (Dancis, 2024). In these cases, students were not simply documenting stories but also disrupting silencing mechanisms and producing new, justice-oriented ways of knowing. A notable thread across the studies is the use of digital storytelling as both methodology and praxis. Students employed blogs, documentaries, podcasts, and social media campaigns (Doerr-Stevens, 2017; Catena, 2020; Levett-Jones et al., 2015) to amplify their voices and link their local experiences to broader social movements.

Many of the most transformative uses for digital content creation extended student work beyond the classroom to authentic audiences and community connections. In one case, students developed murals and digital counter-stories displayed in community spaces (Garcia & Mirra, 2021). The COMPUGIRLS program had adolescent girls produce digital documentaries and interactive projects on systemic injustice, culminating in public showcases (Scott & White, 2013;

Lee et al., 2021). Other initiatives involved digital campaigns and research products for nonprofits and social justice movements (Eudey, 2012; Hicks, 2022; Lyons & Howarth, 2022; Early & Lasker, 2018). These programs illustrate how authentic audiences elevate creation from a classroom exercise into civic participation.

Building critical digital literacy through digital content creation transforms students' relationships with technology, media, and society. By shifting students into the role of digital creators, educators enable them to interrogate how power operates in digital environments, while simultaneously affirming their voices as contributors to public knowledge. The strongest models integrated multimodal production with explicit attention to platform politics and systemic inequities, supported by authentic dissemination opportunities. In doing so, they advanced students' civic identities and positioned digital content creation not merely as a technical skill but as a practice of critical participation in democratic life.

Critical Consciousness Strategies by Context

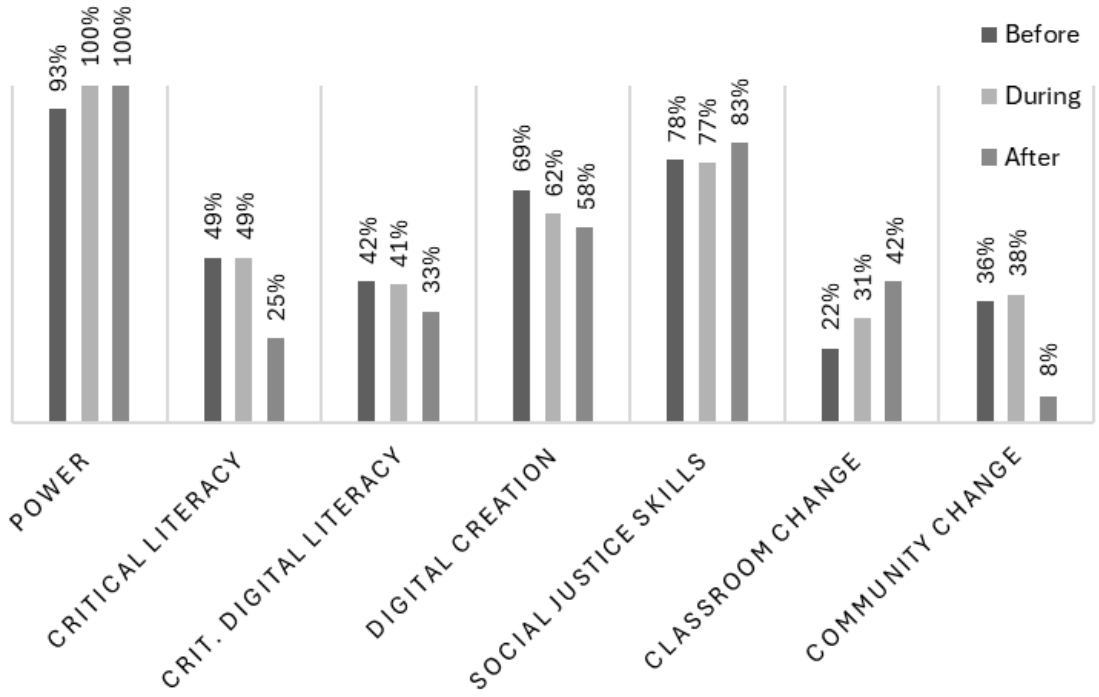
Lastly, in recognition of different opportunities and constraints across time, grade, and online instructional type, we disaggregated strategies along each dimension. In doing so, we examined unique affordances and constraints within each context. We highlight key differences and their implications below.

COVID-19 Timeframe

In comparing critical consciousness raising strategies before, during and after COVID-19, we identified a shift from outward-facing, community-based action toward inward, relational, and reflective practices within the classroom (see Figure 3). These patterns underscored educators' adaptive responses to post-pandemic constraints and the evolving role of technology, pedagogy, and teacher positionality in sustaining justice-oriented learning. Post-COVID interventions more frequently focused on redistributing classroom power, amplifying student voice, and building identity-affirming, care-centered environments that supported belonging and critical dialogue.

Figure 3

Strategies to Foster Critical Consciousness Overall and by COVID-19 Timeframe



Qualitative insights offered several possible explanations for this trend. Educators described how post-pandemic constraints—including increased surveillance of online instruction and the need to attend to socioemotional recovery—often redirected energy away from public-facing activism toward internal classroom equity and relationship-building. In some cases, teachers framed small shifts in classroom voice and participation as meaningful critical action within newly fragile learning communities. Student artifacts showed strong engagement with power analysis (e.g., unpacking algorithmic bias or media stereotyping) but often stopped short of organizing or advocacy beyond the class, reflecting institutional caution and competing priorities during recovery periods. By contrast, some undergraduate instructors used asynchronous forums, podcasts, or digital exhibits to sustain limited forms of civic engagement, suggesting that while the scale of social justice action narrowed post-COVID, the intent to connect critical reflection to action persisted in new modalities. Rather than a retreat from critical engagement, this shift reflects an adaptation—where the classroom became the primary space for practicing equity, reflection, and collective empowerment in a period marked by social disruption and heightened political sensitivity and censorship.

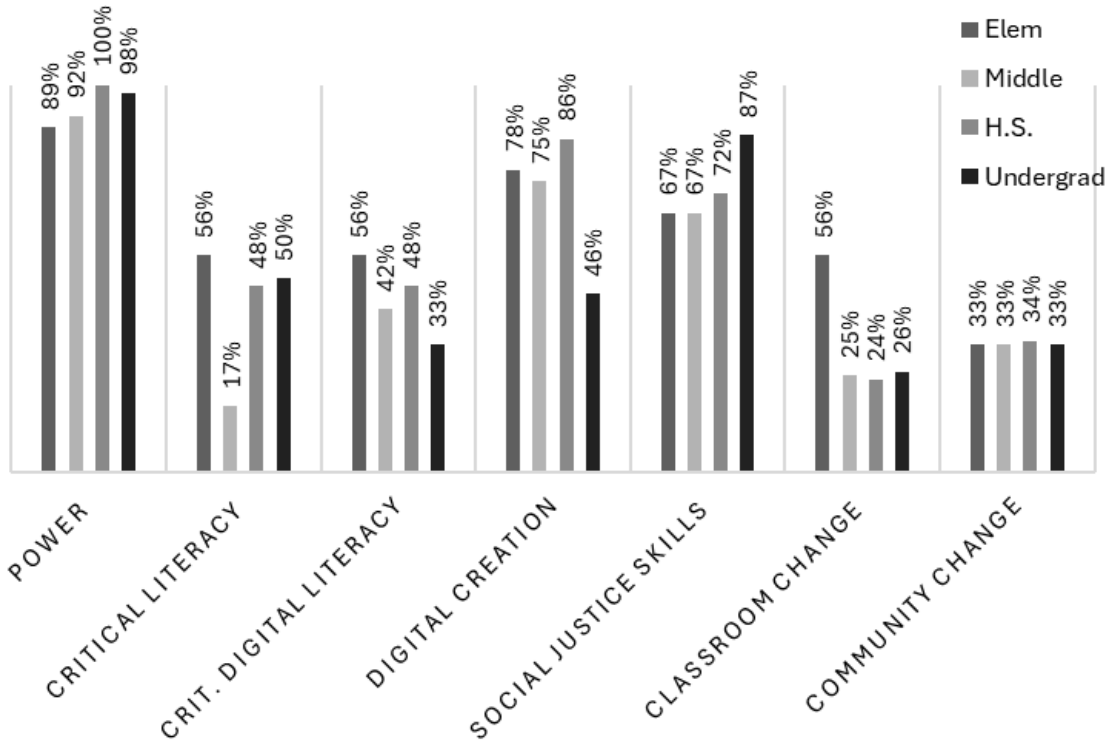
Grade-span

When examining critical consciousness raising strategies by grade-span (see Figure 4), we noted several divergent trends. First, a high proportion of studies at the elementary level (56 percent) involved classroom change compared to studies of older students (24 to 26 percent).

Power sharing tends to take different forms across grade levels, reflecting developmental, curricular, and institutional differences in how agency and responsibility are enacted. In elementary classrooms, power sharing often appears through collaborative rule-setting, choice-based learning, and collective inquiry, where teachers intentionally redistribute voice by inviting students to co-design discussion norms, select project topics, or evaluate peers' ideas. These practices are deeply relational and build from children's social and emotional development, emphasizing voice, belonging, and mutual respect as foundations for early critical consciousness. Teachers at this level frequently use storytelling, dialogue circles, and art-based reflection to help younger students recognize fairness, inclusion, and difference, thereby modeling participatory democracy within the classroom itself.

Figure 4

Strategies to Foster Critical Consciousness by Grade-span



In contrast, power sharing with older students (middle school through postsecondary) tended to focus less on classroom norms and more on shared intellectual authority and civic agency, for example, involving students in curriculum design, issue selection, or facilitation of critical discussions. These classrooms often expected students to take ownership of inquiry and engage with real-world systems of power (e.g., through media critique, research, or activism). However, institutional constraints, grading systems, and curricular pacing in older (tested) can make such power redistribution more limited or symbolic, which may help explain the lower prevalence of observable power-sharing changes in studies with older learners.

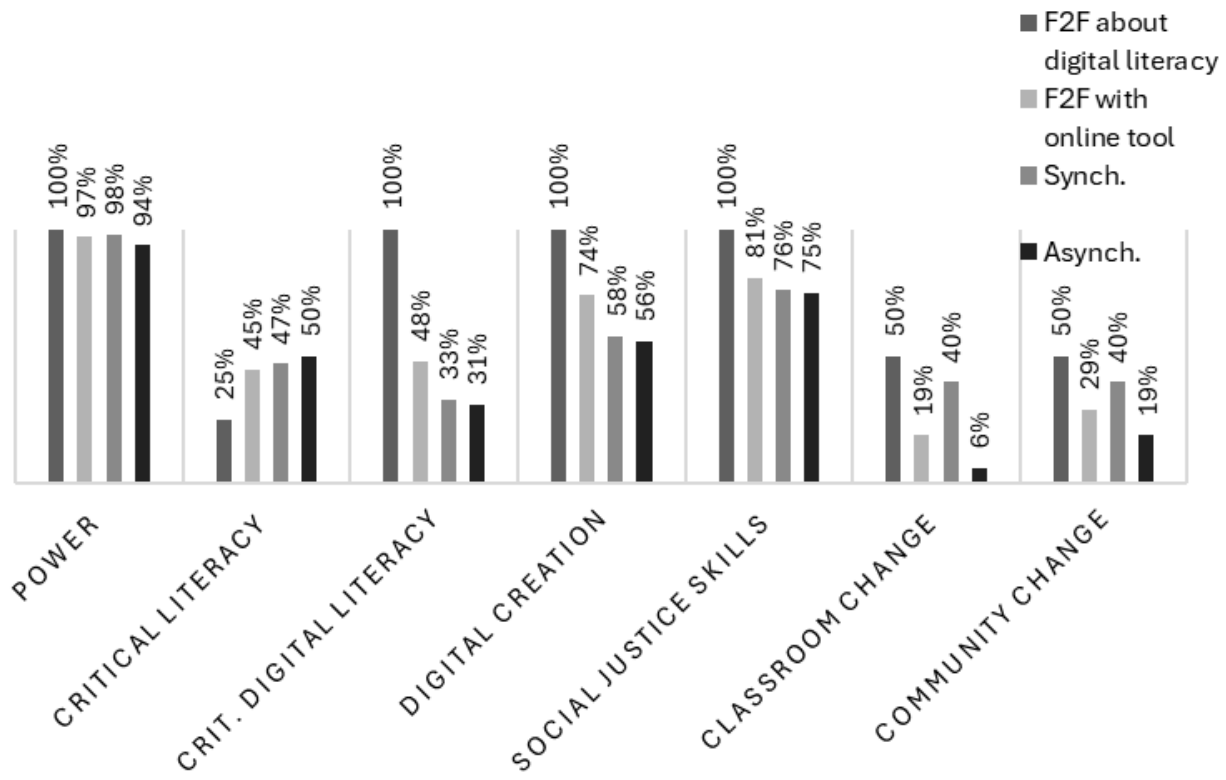
Studies at the elementary level were also slightly more likely to integrate critical literacy and critical digital literacy strategies than studies of upper grades. Importantly, the topic of projects also varied by grade level with critical digital literacy projects in elementary and middle school more often emphasizing awareness of algorithms and civic structures, while critical digital literacy projects in undergraduate settings were more likely to center critical theory, self-assessment, and applied digital justice practices. In contrast, we observed differential patterns of digital content creation between studies at the K-12 and undergraduate level, with substantially lower rates observed at the university level (46 versus 75 to 86 percent).

Online Instructional Type

In contrast, we observed relatively few differences in critical consciousness strategy enactment by online instructional type, as shown in Figure 5. The most notable differences included higher prevalence of change in classroom power dynamics and social justice action outside the classroom in studies examining synchronous online learning contexts (40 percent) or face-to-face classrooms about digital literacy (50 percent). We observed much lower prevalence in face-to-face classrooms using online tools (19 to 29 percent) and asynchronous online contexts (6 to 19 percent).

Figure 5

Strategies to Foster Critical Consciousness by Online Instructional Type



In some ways it makes sense that interventions focusing on online tool integration may be more likely to prioritize novelty or other technology-based advantages over critical action priorities. This lower prevalence in asynchronous environments may reflect the challenges of fostering dialogic exchange and collective reflection—core components of critical consciousness—without real-time interaction or community-building moments. In asynchronous settings, opportunities for immediacy, shared vulnerability, and mutual accountability are reduced, which can make it more difficult to translate critical reflection into collective or outward-facing critical action. Nevertheless, several asynchronous interventions used innovative tactics to engage students in critical action, such as sustained reflective blogging (Chamberlain, 2017), threaded discussions that scaffolded critical dialogue over time (Myers, 2021), collaborative digital storytelling projects (Garcia & Mirra, 2021; Snyder & González, 2021), and the creation of multimodal artifacts (e.g., podcasts, visual posters, advocacy websites; Dias-Fonseca & Potter, 2016; Price-Dennis et al., 2015; Price-Dennis & Sealey-Ruiz, 2021; Rippy & Muñoz, 2019) that allowed students to voice concerns and connect their learning to broader civic contexts even without synchronous interaction.

Limitations

As a synthesis, this study is inherently limited by the available academic literature. We cannot speak to strategies to foster critical consciousness online that have not yet been studied. There is also the possibility that our search strategy excluded relevant literature, although we followed best practices through the inclusion of studies across the three major databases, by including high-quality unpublished as well as published studies, and using a comprehensive set of search terms. We also supplemented our initial search by reviewing cited studies for inclusion in our review. Importantly, our findings are not comprehensive of critical consciousness raising across all open, distance, and digital education contexts. We focus specifically on fully online or technology-mediated instructional environments since the introduction of widely accessible, classroom-ready technologies. Lastly, all findings are descriptive and not intended to imply causality. Replicability of the strategies highlighted depends on local context, including orientation, support, and expertise.

Discussion

Through our systematic review process, we identified prevalent and exceptional strategies for critical consciousness raising in online spaces and using online tools, including unique affordances of these spaces and tools. In addition, we examined variation in the strategies enacted across time, grade span, and online instruction type. This analysis allowed us to speak to potential advantages and limitations of each approach as well as suggest promising strategies for broader enactment. Findings have implications for future research and practice within the fields of online learning, critical theory, educational policy, educator preparation, instructional design, and curriculum and instruction. Policy-makers can use findings to inform online education legislation, curriculum standards, accountability incentives, and education technology funding programs.

Grounded in prior theory and literature, and responsive to emerging trends in our data, we ultimately identified five overlapping strategies employed to foster critical consciousness online. Engaging with power and fostering critical (digital) literacy supported the critical reflection component of critical consciousness, while students developed political efficacy by cultivating skills to engage in social justice work. Finally, students participated in critical action by enacting community change and/or shifting classroom power dynamics.

Across interventions, students engaged in critical reflection online through critical analysis (often using a sociological, historical, political, or critical literacy lens), the discussion of complex and/or controversial topics, thinking critically about normative beliefs, and self-reflecting on identities and practices. Some of the most powerful strategies for accomplishing these aims involved digital storytelling and other forms of art-based expression. Oral history, storytelling, and counter-narrative projects emphasized that students' own testimonies and those of their community were valid data, challenging hierarchies privileging institutional or statistical knowledge (Garcia et al., 2021). This positioned youth and community voices as central to documenting histories of oppression and resistance. Additionally, students were encouraged in several studies to reconstruct narratives to reflect marginalized voices as a way of identifying and pushing back against normative assumptions and power dynamics (Catena, 2020; Garcia & Mirra, 2021). Unique affordances of online tools and spaces directly supported students' critical reflection, including access to digital cultural repositories and the ability to create liberatory content in oral and visual digital mediums.

Interventions cultivated political efficacy several ways, including by positioning students as activists and social innovators, participating in cross-cultural dialogue, embodying the role of (digital) creator, leveraging simulations and role-play as civics training, and engaging in creative expression. Unique affordances and enhancements to enacting these strategies online include learning digital skills to engage in social justice work. Depending on context, learning how to create digital content, leverage online platforms for sharing, and even how to code can cultivate political efficacy. Virtual simulations can also be a helpful platform for allowing students to explore and start building confidence with what critical action looks like in their professions and daily lives (Huang, 2014; Kaye & Spurgin, 2011; Levett-Jones et al., 2015; Shliakhovchuk, 2023). Online tools also reduce barriers to cross-cultural dialogue and collaboration, which support students in developing the skills and confidence to work globally (Dawson & Robinson, 2021; Kerkhoff et al., 2024; Senbayrak & Ortaçtepe Hart, 2024). By transforming students' relationships with technology from consumer to creator, they begin to see themselves as capable of enacting change while developing the confidence and skills to assert their identities and beliefs outside the classroom.

Lastly, students engaged in critical action by enacting community change and shifting classroom power dynamics. In these interventions, participation in political and community movements is shown not as an "add-on" but as integral to classroom instruction. Whether through service learning, participating in political or community movements, or producing knowledge that centered cultural histories and community voice, students' instructional activities had real world relevance and impact outside of the classroom. Students at all grade levels were positioned as civic actors capable of contributing to collective social change. The use of online tools and settings enabled new forms of critical action, including digital advocacy through social media and digital campaign/petition platforms (Harris et al., 2025; Beck et al., 2021; DeWaard & Roberts, 2021; Dias-Fonseca & Potter, 2016; Giunco, 2020; Mills et al., 2011; Ordem, 2023). Students were also able to join in larger social movements regardless of geographic proximity. In many instances, students engaged in critical action by publicly sharing multimodal content, such as digital storytelling, documentaries, and research findings.

Shifting classroom power dynamics positioned students as collaborators rather than passive recipients, transforming classrooms into democratic and participatory communities. Strategies included providing space for students to lead instructional activities, fostering student choice and

inquiry-driven learning, and co-designing classroom structures and assessment practices. By shifting authority, educators fostered classrooms that modeled equity and justice, preparing students for active civic participation both within and beyond school. These efforts not only expanded the scope of online learning beyond the classroom but also disrupted traditional hierarchies of political voice, showing that young people can, and do, play meaningful roles in shaping civic life.

We identified similar trends in strategies to foster critical consciousness by timeframe, grade span, and online instructional type, with a few exceptions. We noticed lower critical action outside of the classroom, complemented by an increase in shifts in classroom power dynamics, after the COVID-19 pandemic, which may be a response to increased censorship and political backlash against any practices related to diversity, equity, and inclusion. Unfortunately, we are unable to examine within our data whether the shifts of similar magnitude occurred in non-online classrooms during the same time. Shifts in classroom power dynamics were observed across all grade-spans, with the highest prevalence observed in elementary grades.

Most relevant, specifically, to online learning, was the observation that critical action (versus solely critical reflection and/or political efficacy) was more likely to occur in synchronous online classrooms than in face-to-face classrooms integrating online tools or asynchronous online classrooms. Nevertheless, we identified several successful strategies for integrating critical action in asynchronous classrooms that could be expanded more broadly, such as blogging, digital storytelling, and digital content creation. Additional research is needed to document the extent to which lower rates of fostering critical consciousness asynchronously are due to inherent limitations of the instructional mode versus the philosophies of learning within institutions providing asynchronous learning. Doing so will provide educators with insight into how to support the integration of liberatory practices across all online modes.

Implications for Future Research, Practice, and Policy

In conducting this study, we systematically reviewed and integrated insights across numerous disciplinary and theoretical approaches. Literature on critical digital literacy contributed concrete examples of critiquing and creating digital content to engage with power and enact social change (Beck et al., 2021; Darling-Aduana et al., 2024; Price-Dennis & Sealey-Ruiz, 2021), while critical theory and liberatory pedagogy deepened sociohistorical grounding, motivations, and implications for critical consciousness raising (Freire, 1994, 2000; Jemal, 2017; Watts et al., 2011). Importantly, while CRP is one of the most common equity-based pedagogies, we identified substantially more studies fostering critical consciousness within other frameworks (Viano & Darling-Aduana, 2025). Those working within CRP frameworks may benefit from drawing upon the liberatory work being conducted in overlapping fields to deepen critical consciousness raising, the least frequently enacted tenet (Ladson-Billings, 2014). Continuing to draw on insights and collaborate across these fields will only strengthen the work of critical scholars and practitioners working in online spaces.

In particular, an entire field exists around critical digital literacy that is unknown to many other critical scholars and educators but has been remarkably successful in not just fostering critical consciousness but also preparing students to engage in an empowered, informed manner in a technology-based world. Curricula ask students to interrogate representation and ethics and analyze how platforms shape meaning (Adams-Grigorief, 2023; Huang, 2014; Nash, 2021; Pangrazio et al., 2022; Wittebols, 2016). Additionally, the emphasis on digital creation fosters

critical consciousness by positioning them as civic actors. Creating multimodal products (digital stories, art, journalism, simulations, oral histories) extends learning beyond the classroom—making justice tangible (DeWaard & Roberts, 2021; Chappell & Hetherington, 2024; Thompson et al., 2022; Murray et al., 2024; Price-Dennis et al., 2017; Rapp, 2023). This body of work positions critical digital literacy as civic practice, rather than a neutral or purely technical skill. Through educating others, developing agency, participating in movements, and engaging in service learning, students learn to critique power, challenge inequities, and contribute to their communities. Importantly, these studies suggest that critical digital pedagogy can foster not only academic growth but also democratic participation and systemic change.

Similarly, shifting power dynamics in the classroom (Amgott, 2018; Blue & Redick, 2022; Giunco, 2020; Little et al., 2024) was a strategy that received a lot of emphasis in some theoretical traditions and none in others. As a result, we as a research team spent time considering but could not fully answer within our research questions and data the interplay and importance of classroom versus community change under the critical action header. Future research could examine how they are related to (and potentially support) each other in enhancing student experiences, outcomes, and critical consciousness. Another important question is whether they can both be equally impactful or whether shifting power dynamics might sometimes be more akin to developing political efficacy, staging or practicing skills they can use subsequently in broader society.

Another major contribution of this study is the documentation and theorizing of what fostering critical consciousness online looks like for younger students. Historically, most enactment has occurred at the higher education level with critical consciousness raising generally (but especially online) rarely observed at the elementary and even middle school level (Viano & Darling-Aduana, 2025). This systematic review documents several successful strategies enacted at younger grades and highlights several others traditionally enacted at the higher education level – such as service learning and participating in political and community movements – that could be integrated with minor modifications within younger grades. Importantly, students as young as elementary school are more than capable of engaging with power, with shifting classroom power dynamics a common strategy for empowering students to identify, learn skills to, and push back against hierarchical systems (Diller & Hays, 2022; Price-Dennis, 2016; Price-Dennis et al., 2015; Price-Dennis & Sealey-Ruiz, 2021; Little et al., 2023). Some of the strongest enactments of critical consciousness raising in elementary classrooms included Global Read Aloud (GRA) - a literacy project that connects students and teachers globally through shared reading experiences (Kerkhoff et al., 2024). This – and similar programs for elementary students like a virtual Freedom School (Amoako Kayser et al., 2024; Harris et al., 2025) – provided texts and prompts to encourage critical reflection, supplemented by digital content creation and community projects and advocacy work. Despite its promise, we did not identify any students enacting critical consciousness online with students younger than third grade.

Additionally, we were struck in reviewing studies how often and important a role creative expression and arts-based pedagogy played. Its prevalence is hardly surprising, as the conceptualization of lived experience and story as epistemology is central to liberatory pedagogy. The arts and storytelling have also been used throughout history to confront norms and surface hidden assumptions (Murray et al., 2024; Price-Dennis et al., 2015), while counternarratives are central to several critical theoretical traditions (i.e., Solórzano & Yosso, 2002). Nonetheless, we emphasize the frequent use of the arts in fostering critical consciousness to inspire others to take up these essential tools for delving into individual identity, society, and the human condition.

At an organizational level, many strong examples of critical consciousness raising occurred by engaging in programs developed by foundations and other not-for-profit organizations (i.e., CDF Freedom Schools, COMPUGIRLS). Expanding the capacity of organizations to provide easily integratable activities that foster critical consciousness, as well as potentially developing these programs within governmental organizations, could greatly expand access to and use of these resources. The development and funding of these free or low-cost programs remove the often unrealistic and unsustainable pressure on individual educators to fully understand, develop, and implement liberatory pedagogy within their classroom. Partnership in this manner can also expand the impact of student action beyond the classroom, connecting them with peers, experts, and policy-makers outside their local communities.

From a policy perspective, findings provide insight into the types of interdisciplinary, student-directed, community-grounded instructional activities most likely to foster informed, engaged citizens. Encouraging (or at minimum making space for) these practices in curriculum standards, legislation, and funding decisions is likely to have the greatest positive impact. Additionally, this study highlights the types of technological tools, modes, and enactments most likely to foster critical consciousness. Prioritizing educational technology and research funding in alignment with these strategies supports student, community, and societal improvement goals.

Conclusion

In the studies reviewed, fostering critical consciousness cultivated not only critical awareness but also the belief that students' voices and actions could matter in the public sphere. By empowering students to critique social inequity and work towards more just systems, critical consciousness raising acts as a bridge between individual learning and collective transformation. Students gain individual skills and learn how to become — through writing, storytelling, gaming, and advocacy — agents of social change. With this outcome in mind, we hope the findings from this systematic review can be used to create a framework and extend best practices within online teacher training, professional development, instructional design, and policy to support the enactment of liberatory pedagogy online that both inspires the mind and transforms communities.

Ethical approval and informed consent: This study involved secondary analysis of publicly available research. Therefore, ethical approval and informed consent were not required, as no new human participants were involved.

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

AI Statement: No generative AI tools were utilized for this submission.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability: All research documents reviewed in the systematic review are identified within the reference section and are publicly available.

References

* Indicates that this article was included in the systematic review

*Adams-Grigorieff, J. (2023). *Grounded critical digital literacies: Youth countering algorithmic and platform power in school and everyday life*. University of California, Berkeley. https://www.proquest.com/dissertations-theses/grounded-critical-digital-literacies-youth/docview/2867925075/se-2?accountid=11226

Ahn, J., & McEachin, A. (2017). Student enrollment patterns and achievement in Ohio's online charter schools. Educational Researcher, 46(1), 44–57. https://doi.org/10.3102/0013189X17692999

*Amgott, N. (2018). Critical literacy in# DigitalActivism: Collaborative choice and action. *The International Journal of Information and Learning Technology*, *35*(5), 329-341. https://doi.org/10.1108/IJILT-05-2018-0060

*Amoako Kayser, A. A., Darling-Aduana, J., Harris, J., Edwards, C., & Alridge, D. (2024). Virtual connections: Teacher beliefs and practices enacting culturally relevant practices in a virtual freedom school. *The Urban Review*, *56*(2), 285-312. https://doi.org/10.1007/s11256-023-00676-3

Apple Inc. (2010, January 27). *Apple launches iPad*. Apple Newsroom. https://www.apple.com/newsroom/2010/01/27Apple-Launches-iPad/

Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. Review of Educational Research, 86(1), 163–206. https://doi.org/10.3102/0034654315582066

*Basile, A. E. (2021). *YPAR: exploring new literacies and critical literacy pedagogy in a virtual teaching and learning environment*. University of Rochester. https://www.proquest.com/dissertations-theses/ypar-exploring-new-literacies-critical-literacy/docview/2572537031/se-2?accountid=11226

*Beck, E., Goin, M. E., Ho, A., Parks, A., & Rowe, S. (2021). Critical digital literacy as method for teaching tactics of response to online surveillance and privacy erosion. *Computers and Composition*, *61*, 102654. https://doi.org/10.1016/j.compcom.2021.102654

*Bell, P. E., & Liu, L. (2015). Social Justice Reasoning of Education Undergraduates: Effect of Instruction in Moral Development Theory and Dilemma Discussion in the Asynchronous Online Classroom. *International journal of technology in teaching and learning*, *11*(1), 60-75. https://eric.ed.gov/?id=EJ1213479

*Bell, P., Coates, R., Colombo, E., Dolgon, C., Hernandez, S., Margulis, M. E., ... & Romo, H. (2015). Globalizing the classroom: Innovative approach to national and international Learning. *Humanity & Society*, *39*(2), 236-253. https://doi.org/10.1177/0160597615574742

Benjamin R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity. file:///Users/qurratainrasheed/Downloads/htarkkala,+07_LONGWORTH-1.pdf

*Blue, L., & Redick, S. (2022). Science for All: Socially Transformative Teaching. *Science Scope*, *45*(4), 48-55. https://doi.org/10.1080/08872376.2022.12291470

*Boetto, H., & Bell, K. (2015). Environmental sustainability in social work education: An online initiative to encourage global citizenship. *International Social Work*, *58*(3), 448-462. https://doi.org/10.1177/0020872815570073

*Bozdağ, Ç. (2022). Inclusive media education in the diverse classroom: A participatory action research in Germany. *Media and Communication*, *10*(4), 305-316. https://doi.org/10.17645/mac.v10i4.5640

*Brennan, C., Packard, M., & Newman, J. (2022). Building a critically transformative community of inquiry in an online undergraduate program. *The Internet and Higher Education*, *52*, 100830. https://doi.org/10.1016/j.iheduc.2021.100830

*Brown, E. A., White, B. M., & Gregory, A. (2021). Approaches to teaching social determinants of health to undergraduate health care students. *Journal of Allied Health*, *50*(1), 31E-36E. https://www.ingentaconnect.com/content/asahp/jah/2021/00000050/00000001/art00018#

*Burrough, X., & Starnaman, S. (2017). A Digital Korl Woman: Students and Workers Recover the Spirit of Life in the Iron Mills from the Digital Factory to the Classroom. *Transformations*, *27*(2), 121-141. https://doi.org/10.5325/trajincschped.27.2.0121

*Call-Cummings, M., & Hauber-Özer, M. (2021). Virtual photovoice: Methodological lessons and cautions. *Qualitative Report*, *26*(10). https://doi.org/10.46743/2160-3715/2021.4971

*Catena, E. C. P. (2020). *Journalism and activism anew: Participatory movements with adolescents writing for change* (Doctoral dissertation, University of Pennsylvania). https://www.proquest.com/dissertations-theses/journalism-activism-anew-participatory-movements/docview/2451112701/se-2?accountid=11226

*Chamberlain, E. (2017). Extending the classroom walls: using academic blogging as an intervention strategy to improve critical literacy skills with elementary students. *Education 3-13*, *45*(2), 243-257. https://doi.org/10.1080/03004279.2015.1078831

*Chappell, K., & Hetherington, L. (2024). Creative pedagogies in digital STEAM practices: natural, technological and cultural entanglements for powerful learning and activism. *Cultural Studies of Science Education*, *19*(1), 77-116. https://doi.org/10.1007/s11422-023-10200-4

Cheng, A., & Hamlin, D. (2023). Contemporary homeschooling arrangements: An analysis of three waves of nationally representative data. *Educational Policy*, *37*(5), 1444-1466. https://doi.org/10.1177/08959048221103795

*da Silva, I., Rogers, D., & Arnett, A. E. (2024). "Welcome to my backyard": an intersectional approach to inclusive teaching in the asynchronous learning environment. *Distance Education*, *45*(3), 446-451. https://doi.org/10.1080/01587919.2024.2355538

*Dancis, J. (2024). Being fractal: Embodying antiracism values in course-based participatory action research. *American Journal of Community Psychology*, *73*(1-2), 234-249. https://doi.org/10.1002/ajcp.12714

Darling-Aduana, J. (2021). Authenticity, engagement, and performance in online high school courses: Insights from micro-interactional data. Computers & Education, 167. https://doi.org/10.1016/j.compedu.2021.104175

Darling-Aduana, J., Good, A. G., & Geraghty, E. (2022). The culture of power online: An analysis of cultural responsiveness and relevance in online high school courses. *Urban Education, 57*(4), 714-742. https://doi.org/10.1177/0042085920972169

Darling-Aduana, J., Asim, S., Heath, M. K., Stewart, A., Viano, S., Garcia, A., Langran, E., Heinrich, C., & Woodley, X. (2024). Digital ethics and equity in K-12 blended and online learning spaces, In Dimitri Christakis and Lauren Hale (Eds.), *Children and screens: A handbook on digital media and child and adolescent development, health, and well-being.* Springer.

*Dawson, L., & Robinson, J. (2021). Constructing ethical learning spaces through interdisciplinary dialogue. *Journal of Intercultural Studies, 42*(3), 300-316. https://doi.org/10.1080/07256868.2021.1930725

De Los Ríos, C. V. (2020). Writing oneself into the curriculum: Photovoice journaling in a secondary ethnic studies course. *Written Communication*, *37*(4), 487-511.

https://doi.org/10.1177/0741088320938794

Delgado, A.J., Wardlow, L., McKnight, K., O'Malley, K. (2015). Educational technology: A review of the integration, resources, and effectiveness of technology in k-12 classrooms. *Journal of Information Technology Education, 1*(14). http://www.jite.org/documents/Vol14/JITEv14ResearchP397-416Delgado1829.pdf

*DeWaard, H., & Roberts, V. (2021). Revisioning the potential of Freire's principles of assessment: Influences on the art of assessment in open and online learning through blogging. *Distance Education, 42*(2), 310-326. https://doi.org/10.1080/01587919.2021.1910494

*Dias-Fonseca, T., & Potter, J. (2016). Media education as a strategy for online civic participation in Portuguese schools. *Comunicar, 24*(49), 9-17. https://doi.org/10.3916/C49-2016-01

Diemer, M. A., Hsieh, C., & Pan, T. (2009). School and parental influences on sociopolitical development among poor adolescents of color. *The Counseling Psychologist, 37*(2), 317–344. https://doi.org/10.1177/0011000008315971

Diemer, M. A., Pinedo, A., Banales, J., Mathews, C. J., Frisby, M. B., Harris, E. M., & McAlister, S. (2021). Recentering action in critical consciousness. *Child Development Perspectives*, *15*(1), 12-17. https://doi.org/10.1111/cdep.12393

*Diller, M., & Hays, A. (2023). Reading, Research and Relationships in a Third-Grade Virtual Book Club. *The Reading Teacher, 76*(5), 535-544. https://doi.org/10.1002/trtr.2147

*Doerr-Stevens, C. (2017). Multimodal exploration of civic stance in the English language arts classroom. *English Teaching: Practice & Critique*, *16*(1), 6-28. https://doi.org/10.1108/ETPC-11-2016-0142

*Donovan, E. L., & Barcus, M. J. (2021). An example of functional integration and application of diversity, inequities, and social determinants of health in an applied physiology course. *Advances in Physiology Education*, *45*(4), 841-848. https://doi.org/10.1152/advan.00085.2021

*Early, J., & Lasker, G. A. (2018). Strengthening communities of inquiry online and offline: Exploring the benefits and challenges of including service-learning in a fully online women's global health course. *Pedagogy in Health Promotion, 4*(3), 218-226. https://doi.org/10.1177/2373379917730843

Esmonde, I., & Booker, A. N. (2017). Power and privilege in the learning sciences. New York, NY: Routledge. https://doi.org/10.4324/9781315685762

*Eudey, B. (2012). Civic engagement, cyberfeminism, and online learning: Activism and service learning in Women's and Gender Studies courses. *Feminist Teacher*, *22*(3), 233-250. https://doi.org/10.5406/femteacher.22.3.0233

Federal Communications Commission. (2014). *Modernizing the E-rate program for schools and libraries* (Report and Order, WC Docket No. 13-184). https://docs.fcc.gov/public/attachments/FCC-14-99A1.pdf

Fitzpatrick, B., Berends, M., Ferrare, J. J., & Waddington, R. J. (2020). Virtual illusion: Comparing student achievement and teacher characteristics in online and brick-and-mortar charter schools in Indiana. Educational Researcher, 29(3), 161–175. https://doi.org/10.3102/0013189X20909814

Flores, E. (2024). *From classroom to community: Cultivating critical consciousness in K-12 ethnic studies for civic and community engagement*. CGU Theses & Dissertations, 779. https://scholarship.claremont.edu/cgu_etd/779

Freire, P. (1994) *Pedagogy of Hope.* New York: Continuum. https://doi.org/10.1080/00131911.2020.1766207

Freire, P. (2000). *Pedagogy of the Oppressed.* New York: Continuum (Original work published 1970). https://doi.org/10.2307/30023905

*Fuchs, D., Zeiser, P., & Engelkamp, S. (2010). *Internationalizing course discussions: round two of a German-American partnership.* Münster: Universität Münster, FB Erziehungswissenschaft und Sozialwissenschaften, Institut für Politikwissenschaft. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-257386

Fütterer, T., Hoch, E., Lachner, A., Scheiter, K., & Stürmer, K. (2023). High-quality digital distance teaching during COVID-19 school closures: Does familiarity with technology matter? *Computers & Education, 199*. https://doi.org/10.1016/j.compedu.2023.104788

*Garcia, A., & Mirra, N. (2021). Writing toward justice: Youth speculative civic literacies in online policy discourse. *Urban Education, 56*(4), 640-669. https://doi.org/10.1177/0042085920953881

Giroux, H. A. (2010). Rethinking education as the practice of freedom: Paulo Freire and the promise of critical pedagogy. *Policy Futures in Education, 8*(6), 715-721. https://doi.org/10.2304/pfie.2010.8.6.715

*Giunco, K. (2020). A Teacher's Reflection on Catholic Social Teachings and Hopeful Curriculum During COVID-19. *Journal of Catholic Education, 23*(1), 175-183. http://doi.org/10.15365/joce.2301132020

Goldhaber, D., Kane, T. J., McEachin, A., & Morton, E. (2022). A Comprehensive Picture of Achievement across the COVID-19 Pandemic Years: Examining Variation in Test Levels and Growth across Districts, Schools, Grades, and Students. Working Paper No. 266-0522. *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*. https://eric.ed.gov/?id=ED620384

Google. (2021, June 16). *10 years of Chromebooks in education.* The Keyword. https://blog.google/outreach-initiatives/education/10thbirthday-chromebook-education/

*Guo, Y. (2019). Mind the Gap: Investigating the Design and Learning Affordances of Agent-Based Participatory Simulations in Changing People's Perceptions and Beliefs toward Economic Inequality. *ProQuest LLC*. https://eric.ed.gov/?id=ED608709

*Guthrie, K. L., & McCracken, H. (2010). Teaching and learning social justice through online service-learning courses. *The International Review of Research in Open and Distributed Learning*, *11*(3), 78-94. https://doi.org/10.19173/irrodl.v11i3.894

Gutstein, E. R. (2016). "Our issues, our people—Math as our weapon": Critical mathematics in a Chicago neighborhood high school. *Journal for Research in Mathematics Education, 47*(5), 454–504. https://doi.org/10.5951/jresematheduc.47.5.0454

*Han, H. (2021). *Middle School Students' Use of Cognitive and Sociocultural Resources During an Examination of a Contested Topic in a Digital Space* (Doctoral dissertation, University of Pittsburgh). https://www.proquest.com/docview/2607559522?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

*Harris, J., Darling-Aduana, J., Amoako Kayser, A., Edwards, C., & P. Alridge, D. (2025). Examining the impacts of a developmentally appropriate, culturally relevant virtual school on student outcomes. *Urban Education*, *60*(4), 896-924. https://doi.org/10.1177/00420859231180139

Heinrich, C. J., Darling-Aduana, J., & Good, A. (2020). Equity and quality in digital learning: Realizing the promise in K–12 education. Harvard Education Press.

*Hicks, E. T. (2022). *Teaching Multicultural Psychology as a Cultural Competence Intervention: An Empirical Evaluation of Course Components*. Utah State University. https://www.proquest.com/docview/2691409740?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

*Hicks, E. T., Alvarez, M. D. L. C., & Domenech Rodríguez, M. M. (2023). Impact of difficult dialogues on social justice attitudes during a multicultural psychology course. *Teaching of Psychology*, *50*(2), 175-183. https://doi.org/10.1177/00986283221104057

*Hilton, J. T. (2013). Digital critical dialogue: A process for implementing transformative discussion practices within online courses in higher education. *MERLOT Journal of Online Learning and Teaching*, *9*(4), 602-614. https://jolt.merlot.org/vol9no4/hilton_1213.pdf

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn?. Educational Psychology Review, 16(3), 235-266.

Horrigan, J. B. (2010). *Broadband adoption and use in America* (Federal Communications Commission Omnibus Broadband Initiative Working Paper No. 1). Federal Communications Commission. https://transition.fcc.gov/national-broadband-plan/broadband-adoption-in-america-paper.pdf

*Hostetler, K. (2021). *The tale of two pandemics: Culturally relevant pedagogy in a blended learning environment amidst COVID-19 and racism pandemics* (Doctoral dissertation, University of South Carolina). https://www.proquest.com/docview/2673637324?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

*Huang, H. K. (2014). Examining youth attitudes toward civic engagement: A study of collaborative social entrepreneurship using a multi-user virtual environment. https://etda.libraries.psu.edu/catalog/23558

Jemal, A. (2017). Critical consciousness: A critique and critical analysis of the literature. *The Urban Review, 49*, 602–626. https://doi.org/10.1007/s11256-017-0411-3

*Jones, L. C. (2021). 'Justice for Victims is Possible': Student Reflections on# MeToo in an Online Gender and Justice Course. *Journal of Criminal Justice Education, 32*(3), 344-360. https://doi.org/10.1080/10511253.2021.1958886

*Jones, T. C. (2020). *"Why have I never learned this?": Exploring health disparities through a critical consciousness lens* (Doctoral dissertation, Iowa State University). https://www.proquest.com/docview/2485145294?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

Junco, E., Vaughn, R., Karimova, G., & Duffy, M. (2023). Constructivist learning practices & culturally responsive teaching: A post-secondary developmental literacy program's multimodal response to the pandemic. *Alanyazın, 4*, 93–104. https://doi.org/10.59320/alanyazin.12401

*Kaler-Jones, C. J. (2021). *" You Can't See Me By Looking at Me": Black Girls' Arts-based Practices as Mechanisms for Identity Construction and Resistance* (Doctoral dissertation, University of Maryland, College Park). https://www.proquest.com/docview/2583583285?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

*Kaye, S., & Spurgin, E. (2011). Using the Internet Platform Second Life to Teach Social Justice. *Teaching Philosophy*, *34*(1), 17-32. https://doi.org/10.5840/teachphil20113412

*Kerkhoff, S., Carpenter, J., Yang, Q., & Dong, Y. (2024). Part of Something Bigger: Critical, Digital, and Global Literacies in the Global Read Aloud. *Literacy Research and Instruction*, *63*(1), 17-41. https://doi.org/10.1080/19388071.2022.2154084

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher*, *51*(7), 500-506. https://doi.org/10.3102/0013189X221109178

*Kumar, K. (2024). Critical peace education and critical hope: co-researching with middle school students in India to imagine possibilities for social change. *Journal of Peace Education*, *21*(3), 313-335. https://doi.org/10.1080/17400201.2024.2347207

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*, 465–491. https://doi.org/10.3102/00028312032003465

Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: aka the remix. *Harvard Educational Review, 84*(1), 74-84. https://doi.org/10.17763/haer.84.1.p2rj131485484751

*Little, S., Raine, H., Choo, A., Joshi, R., Qarni, S. J., Sukri, A., ... & Pakravesh, S. (2024). 'The power to SAY what I want to and it gets written down': Situating children's and adults' voices and silence in participatory research. *British Educational Research Journal*, *50*(3), 964-980. https://doi.org/10.1002/berj.3966

Luter, D. G., Mitchell, A. M., & Taylor Jr, H. L. (2017). Critical consciousness and schooling: The impact of the community as a classroom program on academic indicators. *Education Sciences, 7*(1), 25. https://doi.org/10.3390/educsci7010025

*Lee, C. D., White, G., & Dong, D. (2021). Educating for Civic Reasoning and Discourse. *National Academy of Education*. https://doi.org/10.31094/2021/2

*Levett-Jones, T., Bowen, L., & Morris, A. (2015). Enhancing nursing students' understanding of threshold concepts through the use of digital stories and a virtual community called 'Wiimali'. *Nurse education in practice*, *15*(2), 91-96. https://doi.org/10.1016/j.nepr.2014.11.014

*Lawrence, N., & O'Brien, J. (2016). An Ongoing Journey to Foster Urban Students' Online "Public Voices". *International Journal of Cyber Ethics in Education (IJCEE)*, *4*(1), 32-45. https://doi.org/10.4018/IJCEE.2016010103

*Lyons, K., & Howarth, M. (2022). The importance of hemispheric perspectives for the environmental humanities: reflections on bilingual digital environmental justice storytelling. *Tapuya: Latin American Science, Technology and Society*, *5*(1), 2098685. https://doi.org/10.1080/25729861.2022.2098685

*Marques, R. R., Malafaia, C., Faria, J. L., & Menezes, I. (2020). Using online tools in participatory research with adolescents to promote civic engagement and environmental mobilization: The WaterCircle (WC) project. *Environmental Education Research*, *26*(7), 1043-1059. https://doi.org/10.1080/13504622.2020.1776845

*Matias, C. E., & Grosland, T. J. (2016). Digital storytelling as racial justice: Digital hopes for deconstructing whiteness in teacher education. *Journal of Teacher Education, 67*(2), 152-164. https://doi.org/10.1177/0022487115624493

*McDaniel, D. S. (2024). Toward culturally digitized pedagogy: Informing theory, research, and practice. *Reading Research Quarterly*, *59*(2), 193-210. https://doi.org/10.1002/rrq.534

*McKenzie, T. (2021). Empowered learning: teaching the video essay in the secondary school environment. *Media Practice and Education*, *22*(1), 5-6. https://doi.org/10.1080/25741136.2021.1844950

*Mills, A., Schechter, S., Lederer, S., & Naeher, R. (2011). Global stories of citizenship: Oral history as historical inquiry and civic engagement. *The Oral History Review*, *38*(1), 34-62. https://doi.org/10.1093/ohr/ohr042

Molnar, A. (2021). Virtual schools in the U.S. 2021. National Education Policy Center. https://nepc.colorado.edu/sites/default/files/publications/RB%20Virtual%20Schools%202021.pdf

Molnar, A. (Ed.), Miron, G., Barbour, M. K., Huerta, L., Shafer, S. R., Rice, J. K., Glover, A., Browning, N., Hagle, S., & Boninger, F. (2021). *Virtual schools in the U.S. 2021*. National Education Policy Center. http://nepc.colorado.edu/publication/virtual-schools-annual-2021

Molnar, A. (Ed.), Miron, G., Hagle, S., Gulosino, C., Mann, B., Huerta, L. A., Rice, J. K., Glover, A., & Bill, K. (2023). *Virtual schools in the U.S. 2023*. National Education Policy Center. http://nepc.colorado.edu/publication/virtual-schools-annual-2023

*Muriel, P., & Acosta, E. (2015). Observing the Evolution of Marginalized Youth's Identity through Critical Literacy, Critical Pedagogy and Online Commenting. *The International Journal of Literacies*, *23*(1), 1. https://doi.org/10.18848/2327-0136/CGP/v23i01/48763

*Murray, M., Slingerland, G., Pantidi, N., & McCarthy, J. (2024). Socially Engaged Art Approaches to CSCW with Young People in Rurban Communities. *Computer Supported Cooperative Work (CSCW)*, *33*(4), 645-684. https://doi.org/10.1007/s10606-024-09494-x?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle

Musaddiq, T., Stange, K., Bacher-Hicks, A., & Goodman, J. (2022). The pandemic's effect on demand for public schools, homeschooling, and private schools. *Journal of Public Economics*, *212*, 104710. https://doi.org/10.1016/j.jpubeco.2022.104710

*Nash, B. L. (2021). Constructing meaning online: Teaching critical reading in a post-truth era. *The Reading Teacher*, *74*(6), 713-722. https://doi.org/10.1002/trtr.1980

National Center for Education Statistics. (2021). The digital divide: Differences in home internet access. https://nces.ed.gov/blogs/nces/post/the-digital-divide-differences-in-home-internet-access

New York State Education Department. (2018). *The culturally responsive-sustaining (CR-S) education framework*. https://www.nysed.gov/sites/default/files/programs/crs/culturally-responsive-sustaining-education-framework.pdf

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*(4), 280–312. https://doi.org/10.1086/444136

*O'Brien, G. (2021). Flipped learning as a tool to enhance digital citizenship: How teachers' experiences of online teaching during the COVID-19 pandemic can encourage participatory and justice-oriented citizenship. *Citizenship Teaching & Learning*, *16*(2), 213-223. https://doi.org/10.1386/ctl_00057_1

O'Connor, C. (1997). Dispositions toward (collective) struggle and educational resilience in the inner city: A case analysis of six African-American high school students. *American*

Educational Research Journal, 34, 593–632. https://doi.org/10.3102/00028312034004593

*Ordem, E. (2023). Black Lives Matter in an EFL speaking class. *ELT Journal, 77*(2), 176-185. https://doi.org/10.1093/elt/ccac048

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(71). https://doi.org/10.1136/bmj.n160

Pancer, S. M., Pratt, M., Hunsberger, B., & Alisat, S. (2007). Community and political involvement in adolescence: What distinguishes the activists from the uninvolved?. *Journal of community psychology*, *35*(6), 741-759. https://doi.org/10.1002/jcop.20176

*Pangrazio, L., Stornaiuolo, A., Nichols, T. P., Garcia, A., & Philip, T. M. (2022). Datafication meets platformization: Materializing data processes in teaching and learning. *Harvard Educational Review*, *92*(2), 257-283. https://doi.org/10.17763/1943-5045-92.2.257

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., ... & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47-61.

*Philpot, R., Ovens, A., & Bennett, B. (2024). Implementing a democratic pedagogy into a service-learning course during COVID-19: A self-study of online practice. *Studying Teacher Education, 20*(3), 282-299. https://doi.org/10.1080/17425964.2023.2295908

*Porto, M. (2018). 'Yo antes no reciclaba y esto me cambio por completo la consiencia': intercultural citizenship education in the English classroom. *Education 3-13*, *46*(3), 317-334. https://doi.org/10.1080/03004279.2016.1259247

*Price-Dennis, D. (2016). Developing curriculum to support Black girls' literacies in digital spaces. *English Education*, *48*(4), 337-361. https://doi.org/10.58680/ee201628671

*Price-Dennis, D., & Muhammad, G. E. (Eds.). (2021). *Black girls' literacies: Transforming lives and literacy practices*. Routledge. https://api.pageplace.de/preview/DT0400.9780429521133_A41590606/preview-9780429521133_A41590606.pdf

*Price-Dennis, D., & Sealey-Ruiz, Y. (2021). Advancing racial literacies in teacher education: Activism for equity in digital spaces. *Teachers College Press.* https://www.tcpress.com/advancing-racial-literacies-in-teacher-education-9780807765500

*Price-Dennis, D., Muhammad, G. E., Womack, E., McArthur, S. A., & Haddix, M. (2017). The multiple identities and literacies of black girlhood: A conversation about creating spaces for black girl voices. *Journal of Language and Literacy Education*, *13*(2), 1-18. https://acrobat.adobe.com/id/urn:aaid:sc:US:4c942edd-2c5e-477b-a3db-8e39bbd4672c

*Price-Dennis, D., Holmes, K. A., & Smith, E. (2015). Exploring digital literacy practices in an inclusive classroom. *The Reading Teacher*, *69*(2), 195-205. https://doi.org/10.1002/trtr.1398

Rafalow, M. H. (2020). *Digital divisions: How schools create inequality in the tech era.* University of Chicago Press. https://doi.org/10.7208/9780226726724

*Rapp, S. M. (2023). *" What Counts" in YPAR?: Mobilities and Immobilities in an Online High School English Classroom During the Pandemic*. University of California, Santa Cruz. https://www.proquest.com/docview/2846786251?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

Reeves, T. C., Herrington, J., & Oliver, R. (2002). Authentic activities and online learning. In A. Goody, J. Herrington, & M. Northcote (Eds.), *Quality conversations: Research and Development in Higher Education* (Vol. 25, pp. 562–567). HERDSA. http://researchrepository.murdoch.edu.au/id/eprint/7034/1/authentic_activities_online_HERDSA_2002.pdf

*Ripero-Muñiz, N., Senabe, T., Maseko, M., & Reigadas, Y. (2021). Decolonising and Africanising the Spanish studies curriculum under emergency online teaching: introducing Equatorial Guinean literature in South African higher education. *Anthropology Southern Africa*, *44*(4), 192-205. https://doi.org/10.1080/23323256.2021.2013121?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle

*Rippy, M., & Munoz, M. (2019). Integration of engaging social justice assignments into online courses. *Ubiquitous Learning*, *12*(2), 25. https://doi.org/10.18848/1835-9795/CGP/v12i02/25-38

*Ross, D. L., & Grant, M. (2022). Studying Collisions: Social Justice in Physics. *The Science Teacher*, *89*(4), 46-52. https://doi.org/10.1080/00368555.2022.12293684

*Rushton, E., Dunlop, L., Atkinson, L., Stubbs, J., Diepen, M. T. V., & Wood, L. (2024). Researching education for environmental sustainability through intergenerational dialogue. *International Research in Geographical and Environmental Education*, *33*(2), 157-172. https://doi.org/10.1080/10382046.2023.2270399

*Şenbayrak, M., & Ortaçtepe Hart, D. (2024). Exploring social justice dialogues in EFL conversation clubs: discursive moves and affordances. *Language, Culture and Curriculum*, *37*(3), 327-342. https://doi.org/10.1080/07908318.2024.2339316

*Sernbo, E., Sjöström, M., & Rademaker, A. L. (2024). Developing international virtual student exchange to enhance theory-practice transfer. *Social Work Education, 43*(4), 1110-1122. https://doi.org/10.1080/02615479.2023.2167199

*Schnurr, M. A., De Santo, E., & Craig, R. (2013). Using a blended learning approach to simulate the negotiation of a multilateral environmental agreement. *International Studies Perspectives*, *14*(2), 109-120. https://doi.org/10.1111/j.1528-3585.2012.00470.x

*Sclater, M., & Lally, V. (2013). Virtual voices: Exploring creative practices to support life skills development among young people working in a virtual world community.

International Journal of Art & Design Education, *32*(3), 331-344. https://doi.org/10.1111/j.1476-8070.2013.12024.x

*Scott, K. A., & Garcia, P. (2016). Techno-social change agents: Fostering activist dispositions among girls of color. *Meridians*, *15*(1), 65-85. https://doi.org/10.2979/meridians.15.1.05

*Scott, K. A., & White, M. A. (2013). COMPUGIRLS'standpoint: Culturally responsive computing and its effect on girls of color. *Urban Education*, *48*(5), 657-681. https://doi.org/10.1177/0042085913491219

*Scott, K. A., Clark, K., Hayes, E., Mruczek, C., & Sheridan, K. (2010, March). Culturally relevant computing programs: Two examples to inform teacher professional development. In *Society for information technology & teacher education international conference* (pp. 1269-1277). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/33532/

*Scott, K. A., Sheridan, K. M., & Clark, K. (2015). Culturally responsive computing: A theory revisited. *Learning, media and technology*, *40*(4), 412-436. https://doi.org/10.1080/17439884.2014.924966

*Shliakhovchuk, O. (2023). Learning about refugee life with educational video games. *Information Technologies and Learning Tools*, *98*(6), 94-106. https://doi.org/10.33407/itlt.v98i6.5168

Sidi, Y., Shamir-Inbal, T., & Eshet-Alkalai, Y. (2023). From face-to-face to online: Teachers' perceived experiences in online distance teaching during the Covid-19 pandemic. *Computers & Education, 201*, 104831. https://doi.org/10.1016/j.compedu.2023.104831

Solórzano, D. G., & Yosso, T. J. (2002). Critical race methodology: Counter-storytelling as an analytical framework for education research. *Qualitative Inquiry*, *8*(1), 23-44. https://doi.org/10.1177/107780040200800103

*Smith-Purviance, A. L., Jackson, S., Harper, B., Merandisse, J., Smith, B., Hussey, K., & Lopez, E. (2022). Toward Black girl futures: Rememorying in Black girlhood studies. *Girlhood Studies*, *15*(3), 67-83. https://doi.org/10.3167/ghs.2022.150307

*Snyder, C. K., & González, S. (2021). Towards a pedagogy of transnational feminism when teaching and activism go online. *The Radical Teacher*, (121), 66-76. https://www.jstor.org/stable/48694859

Thanapornsangsuth, S., & Holbert, N. (2020). Lessons from a teacher: Culturally relevant teaching in a constructionist classroom. *Constructionism*, 387. https://www.researchgate.net/publication/343766211_Culturally_relevant_constructionist_design_exploring_the_role_of_community_in_identity_development

*Tharp, D. S. (2017). Imagining flipped workshops: Considerations for designing online modules for social justice education workshops. *Multicultural Perspectives*, *19*(3), 178-184. https://doi.org/10.1080/15210960.2017.1335081

*Thompson, H. R., Mendelson, J., Zamek, M., Cortez, G., & Schillinger, D. (2022). Impact of an arts-based public health literacy program delivered online to high school students during

the COVID-19 pandemic. *Journal of health communication*, *27*(7), 520-534. https://doi.org/10.1080/10810730.2022.2131942

*Thompson, S. A. (2012). Thrice disabling disability: Enabling inclusive, socially just teacher education. *International Journal of Inclusive Education, 16*(1), 99-117. https://doi.org/10.1080/13603111003671640

Viano, S., & Darling-Aduana, J. (2026). Culturally relevant pedagogy in K–12 online learning: A systematic review. *Review of Educational Research*, *96*(3), 724-760. https://doi.org/10.3102/00346543251320355

*Waghid, Z. (2024). Cultivating critical thinking, social justice awareness and empathy among pre-service teachers through online discussions on global citizenship education. *Journal of Creative Communications*, *19*(1), 74-93. https://doi.org/10.1177/09732586231194438

Watts, R. J., Diemer, M. A., & Voight, A. M. (2011). Critical consciousness: Current status and future directions. *New Directions for Child and Adolescent Development,* 134, 43-57. https://doi.org/10.1002/cd.310

*Weiner, C., Van Demark, K., Doyle, S., Martinez, J., Walklet, F., & Rutstein-Riley, A. (2020). The girlhood project: Pivoting our model with girls during COVID-19. *Girlhood Studies*, *13*(3), 32-47. https://doi.org/10.3167/ghs.2020.130304

*Wittebols, J. H. (2016). Empowering students to make sense of an information saturated world: The evolution of information searching and analysis. *Communications in Information Literacy*, *10*(1), 6. https://doi.org/10.15760/comminfolit.2016.10.1.18

*Yeager, E. P., Benjamin, S., Dillette, A., & Goad, A. (2022). Developing self-efficacy through transformative field experiences in tourism pedagogy. *Journal of Teaching in Travel & Tourism*, *22*(3), 242-249. https://doi.org/10.1080/15313220.2022.2096174

*Zamparutti, L. (2022, July). The "antenarrative" in online-asynchronous technical communication courses: A social justice approach to teaching. In *2022 IEEE International Professional Communication Conference (ProComm)* (pp. 392-396). IEEE. https://doi.org/10.1109/ProComm53155.2022.00079

*Zeleny, R. (2023). What a Trip: Social Justice, Virtual Field Trips, and Lessons Learned from Pandemic Pedagogy. *Pedagogy*, *23*(2), 363-378. https://doi.org/10.1215/15314200-10296094

Zimmerman, M. A., Ramirez-Valles, J., & Maton, K. I. (1999). Resilience among urban African American male adolescents: A study of the protective effects of sociopolitical control on their mental health. *American Journal of Community Psychology, 27*(6), 733–751. https://doi.org/10.1023/A:1022205008237

Zimmerman, M. A., & Zahniser, J. H. (1991). Refinements of sphere-specific measures of perceived control: Development of a sociopolitical control scale. *Journal of Community Psychology*, *19*(2), 189-204. https://doi.org/10.1002/1520-6629(199104)19:2%3C189::AID-JCOP2290190210%3E3.0.CO;2-6

Appendix A: Full Text Coding Manual

CRITICAL CONSCIOUSNESS ONLINE LITERATURE SYNTHESIS CODING MANUAL

1. Study ID

2. Publication year (January 1, 2010 and onward)

3. Language

   1 = English
   2 = Another language **[exclude from synthesis]**

4. Publication type

   1 = Peer-reviewed journal article
   2 = Peer-reviewed conference proceedings
   3 = Doctoral dissertation (NOT master's theses)
   4 = Report from trustworthy organizations - governmental, university-affiliated, or non-partisan (exclude reports from companies, partisan organizations)
   5 = Other, specify **[exclude from synthesis]**

5. Grade-level (when more than one grade-level is present, select the span in which the majority of students fall)

   1 = Early Childhood (PreK - 2)
   2 = Elementary (3-5)
   3 = Middle (6-8)
   4 = High school (9-12)
   5 = Undergraduate (preservice teachers if focused on teachers developing skills/competences, excluding preservice/in-service teacher professional development)
   6 = Graduate (excluding teacher professional development) **[exclude from synthesis]**
   7 = Teacher professional development to support enactment in the classroom for students
   8 = Teacher professional development, not specific to classroom enactment **[exclude]**
   9 = Adult learners (career-based) **[exclude from synthesis]**

6. Instruction type

   1 = formal instruction, in a classroom environment
   2 = formal instruction, in an out-of-school environment
   3 = informal instruction **[exclude from synthesis]**

7. Online instruction type

   1 = Asynchronous online course or learning component

2 = Synchronous online course or learning component
3 = In-person instruction using a fully online tool or resource
4 = other, specify **[exclude from synthesis]**

8. Was this study conducted before, during, or after the COVID-19 switch to emergency virtual learning?

   1 = Before
   2 = During
   3 = After

9. What is the research design of the study?

   1 = Intervention
   2 = Observational
   3 = Conceptual
   4 = Review

10. Could the intervention be enacted in a traditional, face-to-face classroom without modifications?

    1 = Yes
    2 = No

11. [If NO] Could the interventions be enacted in a traditional, face-to-face classroom with modifications?

    1 = Yes
    2 = No

12. Are students asked to engage with power?

    1 = Yes
    2 = No **[screen for inclusion, likely to exclude]**

13. Is there evidence of critical literacy "inviting students to question, examine, and even dispute the power relations between writers and readers" (Morrison et al., 2008 p. 441)? **Select and if yes, explain in the following text box.**

    1 = Yes
    2 = No

14. Is there evidence of critical DIGITAL literacy (i.e., encouraging students to "deconstruct the meanings, uses, and messages of digital artifacts to make judgements about digital tools and content… and/or apply ethical perspectives and interrogate power dynamics

present in the creation and use of digital artifact" (Darling-Aduana, 2024)? **Select and if yes, explain in the following text box.**

1 = Yes
2 = No

15. Is there evidence of digital content creation? **Select and if yes, explain in the following text box.**

1 = Yes
2 = No

16. Is there evidence of developing the skills and sociopoltical awareness to engage in social justice work (i.e., by contextualizing academic content in social justice topics)? **Select and if yes, explain in the following text box.**

1 = Yes
2 = No

17. Is there evidence of students enacting change in their classrooms (i.e., co-designing curriculum, reshaping power dynamics in the classroom)? **Select and if yes, explain in the following text box.**

1 = Yes
2 = No

18. Is there evidence of students enacting change outside their classrooms (i.e., engaging in social justice work or providing services)? **Select and if yes, explain in the following text box.**

1 = Yes
2 = No

ONLINE LEARNING
Volume 30 · Issue SI2 · 2026 · e5809
https://doi.org/10.24059/olj.v30iSI2.5809

Modality Matters to Parents Who Support Adolescent Online Learners: A Literature Synthesis

Bobbie Sandberg, Ph.D.
Richard E. West, Ph.D.
Brigham Young University, Provo, Utah

Abstract

In this literature synthesis we investigated parent perspectives on how instructional modality affects the type and amount of support parents need to provide their primary/secondary online learners. Informed by the Academic Communities of Engagement framework (Borup et al., 2020), we analyzed parents' perceptions of their support roles for primary/secondary online learning and their students' support needs given the context of their instructional modality. An intensified need for parents to support primary/secondary online learning was reflected in the literature during the emergency remote teaching (ERT) of the pandemic. We searched ERIC and Education Full Text databases for articles with intersections of four constructs: student engagement, parental engagement support, adolescent online learning, and instructional modality. After identifying an initial corpus of literature, we searched citations and Semantic Scholar using Elicit, an AI-based tool, to triangulate our findings. We discovered how the design decisions online programs make regarding instructional 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. This study affirmed the importance of purposefully designing teaching modalities to align with desired learning outcomes. A carefully designed blend of synchronous and asynchronous content, defined as bichronous instruction, can alleviate stress associated with parental involvement and improve students' independent engagement. When designing bichronous learning, online programs should consider modality affordances, student engagement, and personal circumstances. This synthesis has implications for how modality decisions affect digital equity given concerns about privacy and discrimination in synchronous learning.

Keywords: virtual classrooms, synchronous communication, asynchronous communication, parent role

Introduction

Full-time primary and secondary online students typically learn from home without the physical presence of a teacher. This makes parental involvement in online learning different from a traditional learning context. Understanding how to provide support in an online context can be confusing for parents because parent and teacher support roles often overlap (Hasler Waters & Leong, 2014). The type and amount of support online students need from teachers and parents is affected by the way online content is delivered. Historically, adolescent instruction has been delivered asynchronously so students could complete their courses any time, any place (Veletsianos & Houlden, 2019). But synchronous online learning is increasingly common, where students learn at the same time but in different places (Martin et al., 2021).

For this literature synthesis, we relied on Borup et al.'s (2020) Academic Communities of Engagement (ACE) framework and Martin et al.'s (2020) explanation of online instructional modalities. The ACE framework affirmed that both institutional and personal communities provide support necessary for the success of primary/secondary K-12 online learners. It provides a theoretical foundation for our research because it explains the central role parents play in online contexts. Additionally, it frames how parents and educators must direct engagement support toward affective, behavioral, and cognitive needs. Martin et al. (2020) introduced

Corresponding author: Dr. Bobbie Sandberg, bobbiesandberg@gmail.com
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

bichronous delivery as a distinct instructional modality that intentionally combined elements of synchronous and asynchronous instruction.

As COVID-19 "shifted the educational paradigm" (Dorji, 2021, p. 90), many schools relied on a combination of modalities to meet the needs of both immediacy and flexibility. As parents assumed increased learning support roles (Garbe et al., 2020), researchers sought to understand their perspectives on providing online learning support. The focus of this research is to uncover how parents saw instructional modality decisions as a means to support both students and parents in online learning. The possibility for schools to support both students and parents through instructional modality decisions became increasingly clear as parents took more responsibility for their children's education (Lee, 2022). Parents became the focus of research studies across the world to understand their experiences providing online learning support.

Despite the increase in both parental support research and synchronous online instruction during the pandemic, a very limited amount of research has been conducted to address this intersection of modality and parental support roles. To date no review of the literature exists to synthesize the effect of modality on parents' supportive roles for online learners. The purpose of this study is to determine how primary/secondary online instructional modality affects the need that students have for parental support to help their academic engagement. Understanding this issue will better prepare parents, teachers, and schools to support learners in the current landscape of online learning choices.

Review Question

How does primary/secondary online instructional modality affect the need that students have for parental support to help their academic engagement?

Methods

To answer the question, how does adolescent online instructional modality affect the need that students have for parental support to help their academic engagement?, we used a literature synthesis approach. West and Martin (2024) defined a literature synthesis as a larger cover term for many different kinds of review synthesis approaches where the goal is to synthesize, rather than report, findings. According to these authors, in a literature synthesis, the focus during the literature search is on finding relevance rather than comprehensiveness in the articles, and PRISMA and other review protocols are often used to document the process for transparency rather than as a firm criterion. The goal of this type of review, then, is to create insight for practitioners and reviewers by mapping a space in the literature and using narrative to argue for interpretation of the studies. In this way, it bears similarities to what others call narrative syntheses (Tricco et al., 2016), purpose-specific reviews (Sutton et al., 2019), or generically a literature review (Grant & Booth, 2009).

We reviewed articles with intersections of four main constructs in our research question, student engagement, parental support, primary/secondary online learning, and instructional modality. We defined student engagement as "the energy exerted towards productive involvement with course learning activities" (Borup et al., 2020, p. 810). Parental support is defined as support for students' affective, behavioral, and cognitive engagement. Online learning

is learning that uses digital technologies to give students access to education despite boundaries such as geographic distance. In addition, instructional modality is defined as either synchronous, asynchronous, or bichronous meaning a blend of the two (Martin et al., 2020).

Literature and Database Searches

Our primary sources for locating records for this review were the education databases, ERIC (EBSCO) and Education Full Text (EBSCO). We began our search in the ERIC database with descriptors representing the following four constructs: student engagement, parental support, K-12 online learning, and instructional modality. We identified descriptors from the thesaurus and key articles to ensure accurate terms. Additionally, we limited our search to peer reviewed journal articles and excluded ERIC documents, reports, and conference papers (see Table 1). We did not constrain the article search by publication years; we knew research on our topic would be limited prior to the pandemic but we were interested to find any work that had been done on the topic prior to our study.

Table 1

ERIC Database Search Descriptors and Results of Intersecting Articles

Search Construct	ERIC Search Descriptors and Boolean Operators	Number of Articles
Parental Engagement Support	(DE "Family Relationship" OR DE "Parent Child Relationship") OR (DE "Parent Role" OR DE "Parents as Teachers") OR (DE "Parent Participation" OR DE "Parents as Teachers") OR (DE "Parents" OR DE "Fathers" OR DE "Grandparents" OR DE "Mothers") OR (DE "Parent Child Relationship" OR DE "Parent Student Relationship")	42,551
Student Engagement	DE "Student Needs" OR (DE "Learner Engagement" OR DE "Student Motivation") OR (DE "Academic Achievement" OR DE "Educational Attainment" OR DE "Student Promotion") OR DE "Academic Support Services"	91,944

Search Construct	ERIC Search Descriptors and Boolean Operators	Number of Articles
Adolescent Online Learning Context	DE "Online Courses" OR DE "Blended Learning" OR (DE "Educational Technology" OR DE "Instructional Systems" OR DE "Performance Technology") OR (DE "Technology Uses in Education" OR DE "Computer Uses in Education" OR DE "Electronic Learning") OR DE "Virtual Schools" OR (DE "Distance Education" OR DE "Correspondence Study") AND (DE "Elementary Secondary Education" OR DE "Elementary Education" OR DE "Secondary Education")	78,027
Instructional Modality	DE "Synchronous Communication" OR DE "Asynchronous Communication" AND (DE "Electronic Learning" OR DE "Teaching Methods" OR DE "Web Based Instruction")	3,539
Combined Search		5

Unfortunately the search only yielded five articles. We realized that many articles were not being assigned descriptors for modality. We decided to search for articles where the other three constructs intersected and then manually scanned the articles for a discussion of modality. The results of the search without instructional modality yielded 159 peer reviewed journal articles; the 5 articles from our original search were included in this second search. We then conducted a similar search in the Education Full Text database and found 16 articles that were not duplicates from the ERIC search, totaling 175 peer reviewed journal articles. With the 175 articles, we reviewed the titles and abstracts of the articles to check them against the inclusion and exclusion criteria in Table 2.

Table 2

Inclusion/Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Parents are the research subjects	Students or teachers are the research subjects

Inclusion Criteria	Exclusion Criteria
Context is Articles describe parental support during primary/secondary K-12 online learning	Context is Articles describe parental support in traditional schools or higher education
Articles are empirical research	Articles are theoretical, or literature reviews, ERIC documents, reports, and conference papers
Articles give information about the instructional modality of the online learning	Articles describe technology integration

We excluded 115 articles by reviewing abstracts based on the criteria in Table 2 (see Appendix A). We then read the 60 remaining articles and coded them for discussion of modality. The first author read the full text of the 60 articles and coded them for a discussion of modality. Twenty-one articles were excluded due to no discussion of modality. The remaining 39 articles were coded based on the type of modality (asynchronous, synchronous, or bichronous), whether the discussion of modality was explicit or implicit, and terms or phrases that exemplified how the article discussed modality. All articles that did not explicitly discuss modality were then reviewed independently by the second author. We reviewed our findings to ensure agreement on the final designations (see Appendix A). After resolving any disagreements, we compiled a list of common terminology used to describe instructional modality (see Table 3). We hope this list may help future researchers who face similar challenges where modality decisions are not described overtly.

Table 3

Terminology Describing Instructional Modality

Synchronous terms	Asynch	
Live session		Prerecorded videos

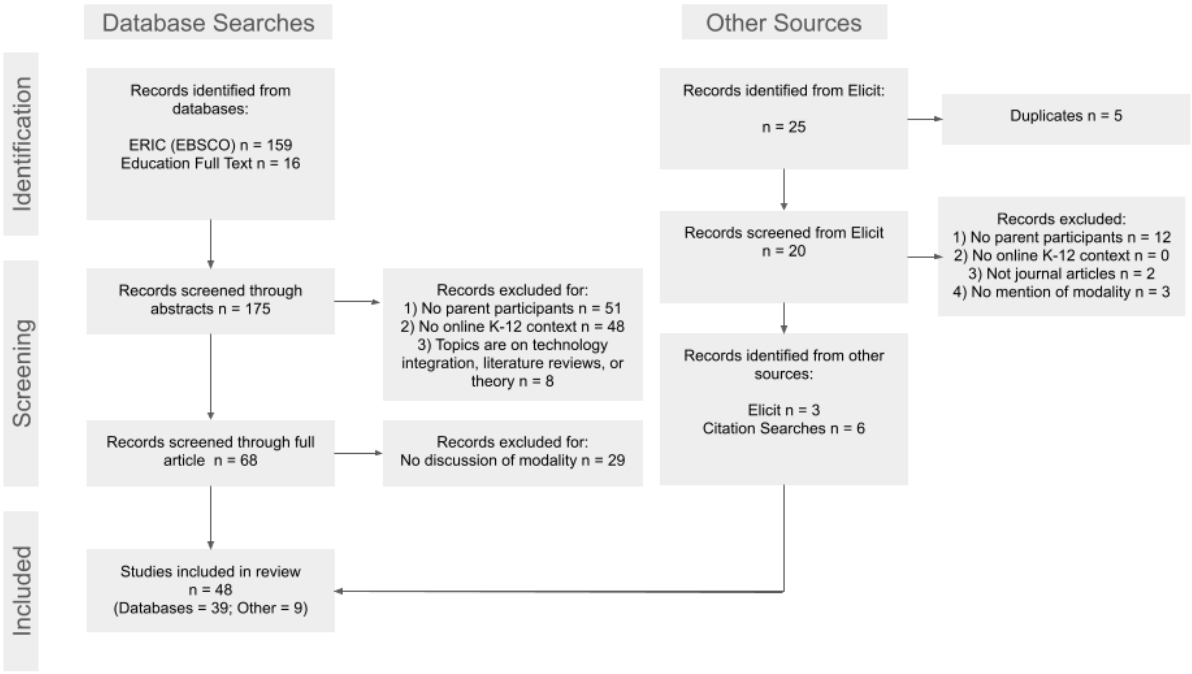
Once we had completed our database search, we reviewed the citations of the included articles and identified six more articles. We also used Elicit, an artificial intelligence tool designed for literature reviews, to triangulate our findings. Based on a semantic analysis of our research question, Elicit conducted a search of Semantic Scholar and OpenAlex. It retrieved 500 papers with semantic relevance to our topic. Elicit then generated a list of exclusion and inclusion criteria. The AI-generated criteria matched ours but added new or more specified criteria in some cases, e.g. rather than searching for parents as the research subjects, Elicit looked for specific evidence of parental support within each article, or Elicit added a new criterion of searching for articles that addressed academic engagement outcomes. The tool then generated a table of the 25 most relevant articles. The characteristics of included studies revealed that defining the instructional modality was difficult for the LLM as well. It was able to identify some synchronous and asynchronous learning explicitly, but other articles were simply categorized as “online learning” or defined by the technological tools they used like Zoom or

Canvas. We reviewed the 25 articles and found that five were duplicates from the database searches and 17 were excluded based on our criteria. We included the three remaining articles in our review.

AI-assisted literature screening has limitations in that searches are not comprehensive, transparent, or reproducible. They require added oversight from researchers to account for potential bias in the AI algorithm retrieval process. We found that queries may misrepresent highly specific concepts, like the types of instructional modality. Knowing these limitations, we intentionally used Elicit as a method of triangulation only after we conducted our original database searches. We hope our paper can provide a guide to others seeking to use AI tools in how they can be used intelligently to triangulate human-based literature synthesis. Figure 1 shows a PRISMA chart, documenting the search process for choosing the 48 articles included in this research.

Figure 1

PRISMA Diagram Chart



Note. Adapted from "The PRISMA 2020 statement: An Updated Guideline for Reporting Systematic Reviews," by M. J. Page et al., 2021, The British Medical Journal, 372. https://doi.org/10.1136/bmj.n71. In the public domain.

Literature Review/Analysis Strategy

To review the articles, we charted information about each study's publication date, nationality of participants, research design, purpose, target population, and modality. We then examined the articles for discussion of instructional modality and noted how parents talked about their parental support roles. We noted the type and amount of support they gave, as well as any

recommendations they made for how online programs' decisions about instructional modality could affect their parental support roles. Although theoretical articles were not part of our database-generated search, we referenced key theoretical articles about parental engagement support and instructional modality along with the findings to give context to the research.

The final list of included articles in this review represented a broad range of geographic locations with parents from 24 different nations. There was a surge of research conducted about parents' experiences supporting online learning during the pandemic, and that is reflected in the articles included in this review. The majority of the included articles used qualitative research methods to capture the emotions parents experienced supporting primary/secondary students during the pandemic (See Table 4).

Table 4

Description of Articles Included in Review

Description of Articles		Number of Articles
Publication Date		
	Before COVID	6
	During COVID	39
	After COVID	3
Research Design		
	Qualitative	29
	Quantitative	13
	Mixed Methods	6
Participants' Country of Origin		
	Afghanistan, Australia, Bhutan, Canada, China, Croatia, Czech Republic, Eritrea, Ethiopia, Greece, Ireland, Libya, Nepal, Pakistan, Palestine, Philippines, Portugal, Qatar, Somalia, Turkey, United Arab Emirates, United Kingdom, United States, Wales	24

Findings

From our review of the 48 articles in our pool, we found that instructional modality decisions can affect the type and level of support primary/secondary parents need to provide their online students, specifically how parents support their students' affective, behavioral, and

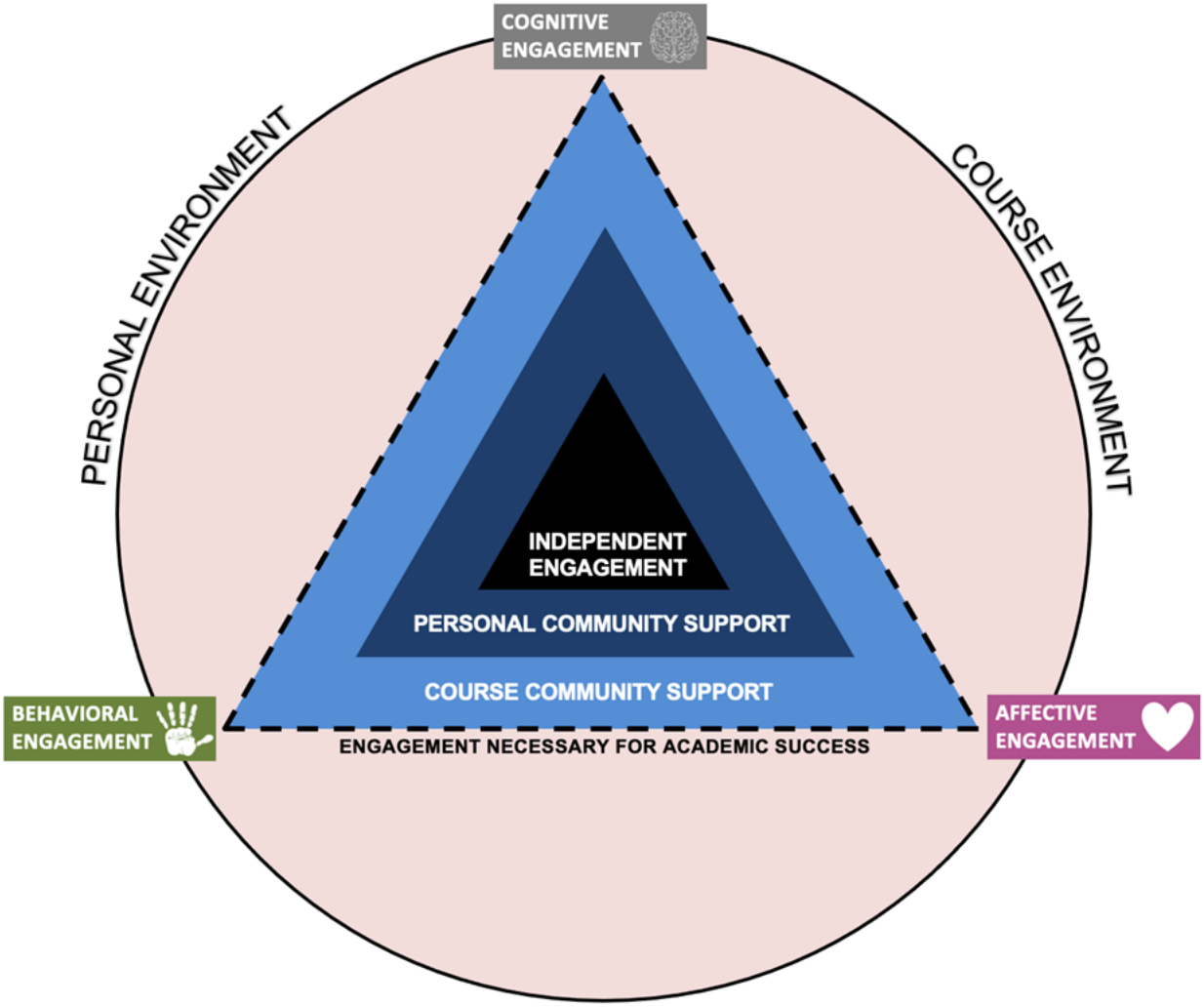
cognitive engagement. Modality choices also influence parent perceptions of school-provided support and impact online students' support needs.

Modality Affects the Type and Amount of Support Parents Provide

When students learn online, they depend on different sources of support to achieve the engagement necessary for academic success. Borup et al. (2020) developed the Academic Communities of Engagement framework to explain academic engagement in blended and online learning contexts. To be successful, students need to engage in learning affectively, behaviorally, and cognitively. Students' ability to engage independently in all three dimensions varies, so school and personal communities need to provide support to fill the gaps in students' engagement (See Figure 2). Parents, who are the primary actors in the personal community, have to construct new support roles in online learning contexts (Sandberg et al., 2026). The decisions the school community makes about instructional modality can affect the type of support roles parents need to assume.

Figure 2

The Dimensions of Engagement and Communities of Support in the ACE Framework



Created by Jered Borup and Charles Graham, CC BY 2.0.

Defining Parents' Online Learning Support Roles

Much of the literature prior to the pandemic addressed instructional modality in the context of defining parental support roles in an online learning environment. Those roles are challenging to define, not only because they differ from in-person education contexts, but also

because in online learning, parents' roles often overlap with teachers (Borup et al., 2014). For example, Downes (2013) explored parental support roles in adolescent distance education in remote regions of Australia. Parent supervisors in her study had both asynchronous and synchronous learning support through learning packets and satellite lessons. Even with support in both types of instructional modality, she found that parents filled roles of instructing, assisting with assignments, monitoring progress, preparing lessons, and motivating–all roles that aligned with those of teachers. She compared parents' experiences to that of new teachers, except that parents had no educational training, no professional development, and no teacher network to rely on. In addition, parents often supported students of multiple age groups simultaneously. Despite these challenges, parents reported significant satisfaction in their role as supervisors due to the closer relationships they developed with their children. Because parent supervisors' roles so closely matched teachers' roles, Downes referred to parents as "an often overlooked cohort of educators" that needs attention in the research (2013, p. 31).

Asynchronous learning contexts seemed to exacerbate the confusion this "overlooked cohort of educators" feel in distinguishing their support roles from the teacher's. To differentiate between the roles of parents and teachers, Hasler Waters and Leong (2014) conducted an ethnographic analysis of a mostly asynchronous online program. They concluded that in their asynchronous context, parents' roles were managers and guides for the curriculum, while teachers' roles were instructors and facilitators. Even though they differentiated parent and teacher roles, they still concluded that parents were "co-educators" with teachers (p. 52). Similarly, Borup et al. (2015) found that in an asynchronous learning context parents were more comfortable filling roles of advising, mentoring, and organizing both time and space, but they felt they needed to monitor and motivate their students because they were working asynchronously. Parents reported discomfort when students wanted their help with instructing because parents wanted students to ask teachers their questions about course content. Curtis and Werth (2015) recommended that adding synchronous instruction might improve parents' role confusion by shifting some responsibility back to the teachers. They postulated that synchronous instruction might also improve transactional distance between students and teachers and build caring support communities.

Research conducted after the pandemic about parental support roles included more direct discussion of how the affordances of each instructional modality affected parents' roles. Radcliffe (2023) explored parent perspectives on the affordances of instructional modalities from survey responses of 43 parents of adolescent distance education students in Australia. While synchronous learning allows for direct instruction, live interaction, collaboration, and immediate feedback, asynchronous instruction makes it possible for students to choose from a broader selection of courses, to have flexibility in time and place of learning, and to choose the pace of progress and depth of study. Noting the demographic shift of distance education students from rural communities to urban centers, Radcliffe (2023) suggested that today's distance education students need specific affordances of both synchronous and asynchronous instruction so they can experience both community and deep learning opportunities in an online context.

Modality Affects Parents' Affective, Behavioral, and Cognitive Support Roles

Several studies have found evidence of the effect of instructional modality as they use the ACE framework to explore parental support roles across affective, behavioral, and cognitive

dimensions (Borup et al., 2020). Guo et al. (2024) explored how parents' online learning support is affected by learning modalities. In open-ended survey responses from 567 parents of adolescent online students, parents expressed a need for more synchronous learning. They reported that their students lost motivation for learning when they did not have interactions with teachers or peers. They felt a sense of inadequacy about providing support, particularly in cognitive areas like math where they did not have subject-matter knowledge. This call for more synchronous instruction is interesting given other findings from the survey that indicate parents of students learning in synchronous modalities reported needing to provide more support than parents of students in asynchronous courses.

This call for more synchronous instruction was repeated in many of the studies we reviewed, but like this study, the research suggested that simply increasing the amount of synchronous instruction does not lead to the outcomes parents desired. For instance, Budhrani et al. (2021) explored how parental support roles of organizing, facilitating, monitoring, motivating, nurturing and supporting learning are differentiated based on instructional modality. Parents needed to support asynchronous students with learning loops, assessment, and feedback. In bichronous courses, parents had to provide just-in-time assistance to answer students' questions. Parents noted that although the time their parental support time decreased with synchronous instruction, but the reduced flexibility and time-constrained deadlines for live classes increased the demands for immediate help and caused stress in the family Parents felt that synchronous instruction was so limited that students still needed their help with both the learning loop and the just-in-time assistance. As a result, Budhrani et al. proposed that online school orientations should prepare parents with accurate expectations for parental support within the context of the school's instructional modality.

Beck (2024) found evidence of a nuanced perspective of how parents see their support change based on instructional modality when he examined how virtual schools helped at-risk students engage affectively, behaviorally, and cognitively. Parents noticed their students' affective engagement improved when teachers used time in live classes to assess their emotional wellness. They also said that giving their students time to interact with classmates in break-out rooms boosted their affective engagement. Parents noted improved behavioral engagement from asynchronous support. A weekly newsletter from the school outlining the learning plan for the week or color coding of tasks in the LMS helped students manage their online work. But students also benefited behaviorally from synchronous support, like when a teacher shared her screen to walk the students through each step of how to submit an assignment online. Finally, parents felt synchronous interactions supported cognitive engagement when students could do small group work in zoom rooms or attend office hours where they could get live, one-on-one help from a teacher. Parents noticed their secondary-age students struggled to get their questions answered in the live sessions because of the number of students. They preferred to get cognitive help from teachers in the form of asynchronous feedback, like a video response to their specific question. Parents in Sandberg et al. (2026) similarly noticed that instructional modality particularly affected their ability to provide cognitive support. When parents needed to help students with cognitive tasks, they often tried to relearn the material. Instructional modality choices affected how parents were able to relearn course content. In settings where content was delivered synchronously, parents often listened to live classes along with their students, whereas in asynchronous courses, parents would study materials in the LMS so they could understand and explain the concepts to their students.

Modality Affects Parents' Perception of School Support

Parents' perception of online school support was affected by the amount of synchronous instruction the school provided. Although they were primarily concerned with how parent perceptions varied across socioeconomic status, one of the features Smetackova and Stech (2021) considered when they administered a questionnaire to 2,528 parents in the Czech Republic was parent satisfaction with school practices during the pandemic. They noted that as students learned online in the Czech Republic, the amount of live, synchronous instruction they received varied. They found that regardless of the amount of live instruction time their students received, parents had positive perceptions of online meetings, viewing them as good substitutes for their traditional classes.

Similarly, Booth et al. (2021) surveyed over 2,000 parents in the UK and Wales about factors that affected learning, including instructional modality and parental support. They found that the instruction time students received in virtual classes varied significantly between state schools and private schools. Where private school students received four hours of live instruction each day, state school students averaged only one hour. Booth et al. (2021) asked parents about their perceptions of students' learning. In contrast to Smetackova and Stech (2021), parents of state school students felt that their students' academic progress had suffered more than parents of private school students. Although the study only measured parent perceptions of outcomes rather than actual academic outcomes, it is interesting to note that just receiving more synchronous instruction time led parents to perceive better academic outcomes.

Modality Decisions Affect School Support for Accommodations

The transition to online learning left many students with disabilities with unmet needs (Ortiz et al., 2021) given the disruption of services and routines (Valcourt-Pearce, 2022). Parents of students with disabilities experienced new and increased support roles as well as a loss in accommodation services with a move to virtual learning (Munsell et al., 2021; Ortiz et al., 2021). Averett (2021) found that families struggled, not only because schools did not provide accommodations or resources at home, but because the students' needs were different when they were learning remotely. Students with disabilities struggled in asynchronous contexts that lacked interaction with teachers and an opportunity to observe visual cues like body language. Maintaining focus was especially challenging in distracting home environments and in lengthy video-based assignments. Synchronous contexts posed other challenges, particularly with behavior, where students needed constant monitoring from parents or one-on-one Zoom meetings with their teacher. Parents reported the most positive outcomes when, regardless of modality, they felt the teachers made added efforts to reach their student's individual needs.

The fact that students with disabilities struggled in both synchronous and asynchronous modalities was echoed by multiple researchers. Parents' in Harkins et al. (2022) had mixed opinions about the effectiveness of each modality. Some believed virtual schools should take advantage of online affordances that would benefit students with disabilities, like recorded class sessions while others disliked synchronous learning because their students struggled to participate appropriately in synchronous settings. Valcourt-Pearce (2022) found that her deaf children benefited from synchronous content delivery because accommodations like closed captions and ASL interpreters in Zoom sessions that helped make content more accessible, but

she also noticed academic losses since other accommodations and services were not available to students learning from home. Yıldırım and Rovshenov (2022) found that parents of students on the autism spectrum struggled in the early weeks of the pandemic when all learning happened asynchronously. But even when courses moved to synchronous delivery, students struggled with distraction and staying caught up with the teacher.

Significantly, parents felt if online learning was the only option then they needed an entirely different system because both the synchronous and asynchronous delivery methods were failing their students with disabilities (Yildirim & Rovshenov, 2022). Parents in Harkins et al. (2022) noted that when students with disabilities transfer to a virtual learning environment, modifications may need to be made to the online instruction and to the IEP. And Munsell et al. (2021) advocated for exploring new, more effective, delivery options for students with disabilities. Specifically, they felt a multimodal approach where schools provide asynchronous materials to parents who need more flexibility with time may reduce parent stress.

Modality Decisions Are Affected by Access to Digital Resources

Many families had limited access to educational content either because they lacked technological tools or proficiency for using those tools. In some cases, language and cultural barriers also prevented access to content provided through the tools. Because many students had insufficient access to technological devices, parents appreciated teachers who made deliberate modality decisions with technology access in mind. Knowing access to digital tools was limited, Libya, Palestine, and Afghanistan each used both synchronous and asynchronous content delivery to ensure a continuity of education during primary/secondary school lockdowns. They created online materials to be completed asynchronously. When they taught in live synchronous sessions, they made recordings for those who could not participate live (Khlaif et al., 2021).

Synchronous content delivery accentuated the problem of limited digital resources, even amongst more privileged families. Bhamani et al.'s (2020) sample of parents came from urban areas of Pakistan, had internet access, and spoke English. Yet even these parents reported a lack of digital resources as one of their challenges. They noted that when schools delivered content synchronously, many families could not participate because they did not have enough devices for multiple students to be in online classes simultaneously.

To account for limited digital access, schools in Bhutan delivered content asynchronously through resources like Google classroom and shared content synchronously through television. However, they found that both modalities had limitations. Synchronous content was ineffective because it did not match what students were learning asynchronously and it was not differentiated for grade level. Students struggled to complete asynchronous learning activities on time given their limited access to digital tools and inexperience with navigating Google classroom. This created significant stress for both parents and students, so much that some students reported parental abuse and cheating because of the pressure they felt to complete assignments (Dorji, 2021).

In locations where technological tools or internet access was even more limited, schools offered fully asynchronous content that required no internet connectivity through self-learning modules (SLMs). Although Filipino students expressed a preference for education content to be delivered through a variety of visual, auditory, and tactile methods (Capinding, 2021), many

schools in the Philippines offered SLMs to ensure that digital access did not determine educational opportunity. Three studies in our review highlight Filipino parents given the integral part they played in supporting learning asynchronously. In Naungayan et al. (2024), parents described the support roles they needed to fill when their children learned in asynchronous SLMs as sustaining motivation, monitoring compliance, and tutoring. Other research assessed how parent demographics affected learning outcomes in SLMs. Capinding (2021) found that parent education and economic status correlated with high mathematics performance from students learning through SLMs in the Philippines. Finally, Collado et al. (2023) assessed how internet access affected students' learning engagement. Interestingly, parents reported that internet access was not a predictor of student engagement. In this case where digital access was a concern, asynchronous activities allowed for high student engagement without internet connectivity.

In other cases, the digital divide was defined less by access to resources and more by an ability to access what they needed from the digital tools. In the United States, parents of English Learners (EL) had difficulty providing the necessary support even when the schools provided devices for each student because they lacked digital literacy skills. Because access, language, and time were all barriers to parental support, the schools offered a variety of both synchronous and asynchronous learning support. Parents responded well to asynchronous family-based learning activities that were translated into Spanish because they met their need for flexibility and comprehension (Delany-Barmann et al., 2021).

Digital access was also determined by cultural barriers for participation in synchronous and asynchronous learning. Due to privacy concerns, parents in Afghanistan, Libya, and Palestine restricted their daughters from showing live images of themselves during synchronous instruction. Consequently, researchers found that women were more likely to participate in online asynchronous discussion boards, while male students participated freely in both synchronous and asynchronous instruction (Khlaif et al., 2021). This example highlights how modality decisions affected equitable access to learning for all students. Racial barriers also affected digital access. Other students hesitated to participate in live classes due to a fear of discrimination. Zhang and Halpern (2021) looked at a case study of a mother living in northern California whose twin students experienced discrimination at school due to their Chinese heritage. During the pandemic, the school implemented district-wide Zoom classes where students were supposed to keep their cameras on. Although they struggled with engagement, the twins kept their cameras off as it shielded them from the microaggressions they experienced in face-to-face classes. Educators ensure access when they make decisions that are sensitive to issues of privacy and equity that arise with diverse student demographics.

Modality Affects Other Aspects of Parents' Lives

Through the articles in this review, we documented how instructional modality either constrained or improved other aspects of their lives. Time and stress were common concerns for parents supporting online learning, and instructional modality had the potential to impact both. As mentioned previously, parents felt that live classes added stress because of time-constrained deadlines (Budhrani et al., 2021). Parents also reported that supporting online learning created significant stress in their lives (Timmon et al., 2021) because it placed heavy demands on their personal time, especially those who were working full-time jobs. When a brick-and-mortar school administered online web courses for students in remote Canadian provinces, they noticed

what an "enormous amount of school-personnel time" it took to do what parents are required to do when they set up, supervise, and support students' online learning (Philpott et al., 2010). Garbe et al. (2020) also detailed the emotional toll on parents of providing online learning support within the first month of the pandemic, explaining that parents were overwhelmed from balancing responsibilities of their employment while supporting online learning.

In a large, quantitative study Ribeiro et al. (2021) documented how the amount of time parents spent supporting online learning differed based on age, student autonomy, and instructional modality. By surveying 21,333 Portuguese parents, researchers found that parent support time was highest when students were in the youngest level of school. Parental support time was also affected by student autonomy. Students with high independent engagement abilities required less parental involvement than their less autonomous peers. Finally, instructional modality also affected the amount of parent support time. Parents reported spending less time supporting private school students who received more synchronous instruction than public school students. One limitation of this large quantitative study is that close-ended responses may have limited parents' ability to share their attitudes about the time they spent providing online learning support.

Social status of students' families was another factor that affected parental support. Demir and Yildizli (2022) documented at-home learning experiences of 20 parents of public and private school students in Turkey. They found that parents' positive and negative feelings about their role in providing support varied according to the socioeconomic status (SES) of the family. Low SES families who lacked technological resources or academic abilities expressed more negative emotions about online learning and worried the pandemic would widen the academic divide. High SES parents were more likely to express positive emotions about online learning and to leverage additional resources to support their students' learning. Demir and Yildizli (2022) argued that improved instructional practices, like building asynchronous e-learning materials and training teachers in better synchronous teaching techniques and communication, would help address the concerns of low SES families. And ultimately, these changes could benefit all students regardless of their SES status.

Although many parents felt the increased time supporting their students added stress, others reported that it strengthened their relationships (Drvodelic & Domović, 2022; Harmon & Arnold, 2023). The parents who responded to Roy et al.'s (2022) survey felt the biggest benefits of remote learning were (a) more family time, (b) flexibility in the child's schedule and ways of learning, and (c) better parent understanding of child's learning style, needs, or curriculum. In open-ended responses, parents asked for both more live synchronous sessions and better asynchronous learning materials because they felt that modality choices affected their relationships with their students. Similar to Budhrani et al.'s (2021) finding, Roy et al. (2022) found that when synchronous classes were too rigid, they created stress instead of support, as students demanded more just-in-time help from parents.

Modality Affects Students' Online Learning Support Needs

As researchers interviewed and surveyed parents about their students' support needs while learning at home, they found that parents worried about social interactions, student engagement, and insufficient teacher support. They were concerned that the lack of preparation

for online teaching during the pandemic led to insufficient teacher support. They felt that online schools could use instructional modality to address some of these student support needs.

Social Interactions

Parents worried about their children's feelings of isolation and social development while learning online (Bhamani et al., 2020). Chances for learning through collaboration and interaction were limited. Researchers concluded that more synchronous learning could improve social challenges in remote learning (Flynn et al., 2021). Looking specifically at the affective, behavioral, cognitive, and psychosocial dimensions of engagement, Flynn et al. (2021) surveyed 2733 parents of adolescent students in Ireland to describe the effects of online learning on social development. Informed by Vygotsky's social constructivist ideas, their study confirmed the need for social collaborative environments and scaffolded learning opportunities with more knowledgeable peers. They recommended additional synchronous learning activities like live online classes and collaborative learning with peers.

Harmon and Arnold (2023) also asserted that learning needed to be engaged, active, meaningful, and socially interactive. They perceived remote learning as inferior to in-person learning partly due to its lack of social interactions. They recommended that online schools consider how they structure and deliver education, including using technology to improve online interaction. Despite their concern over social isolation, parents in this study also noted that although their students had limited peer interactions, they had more time for positive family interactions.

Student Engagement

Parents in multiple studies noticed that a blend of synchronous and asynchronous instruction was beneficial in supporting student engagement. Some research suggested that instructional content needed to be delivered using varied methods so students could engage in learning both on-line and off-line (Lau et al., 2020). Respondents in Lei et al. (2022) described how parents felt their students' engagement was affected by instructional modality. They noticed it was challenging for students to engage affectively in asynchronous learning, and sometimes in synchronous sessions if the teacher did not facilitate good interactions. But even when there were noticeable deficiencies in teachers' online facilitation, parents still felt that synchronous delivery improved engagement since students were able to ask questions, get feedback, and interact with teachers and peers. Eight highly educated Czech mothers of Primary school-aged children also advocated for blending modalities. They noted that well-prepared synchronous instruction had a positive impact on student engagement, technical skills, and independent engagement, but their students' engagement suffered if they had to maintain focus in a synchronous environment all day (Kostelecká et al., 2021). Other parents worried that too much synchronous instruction led to excessive screen time and was damaging to their students' learning engagement (Drvodelic & Domović, 2022; Lau et al., 2020; Saliverou et al., 2021).

Blending modalities in an online course has benefits that can be linked to improved student engagement. Since students prefer a variety of visual, auditory, and tactile learning methods when they are completing online coursework (Capinding, 2021), changing modalities expands the possibilities for how students can engage with their work. Varying modalities may be one way for teachers to adapt learning activities to suit the learners in their course (Burns et

al. 2023). Because teachers understand the needs of the individual students in their course, they can make decisions about how to deliver content most effectively, rather than following district-mandated modality decisions (Timmons et al., 2021).

Parents appreciated that their students gained confidence in their abilities to engage independently as they learned online (Drvodelic & Domović, 2022; Philpott et al., 2010). Wilson-Forsberg et al. (2024) interviewed and conducted focus groups with 10 refugee mothers from the Horn of Africa who were living in Canada during the COVID-19 pandemic. Due to language barriers, the mothers had to rely on interpreters to communicate with the school. When the pandemic restrictions limited any in-person contact, these mothers lost their ability to communicate face-to-face with the school. Consequently, they relied on their students to act as interpreters in their home-school communication. Their students developed significant independence as they managed coursework that required both synchronous and asynchronous interactions and acted as liaisons between their mothers and the online schools.

Interestingly, parents' behavior impacts student engagement in certain modalities. Shishakly (2024) found that parents both negatively and positively affect online student engagement by how they interact with students during their synchronous learning. Parents' interference in online classes negatively affected teachers conducting the classes. Conversely, parents positively impacted student engagement when they encouraged students to have good attitudes and pay attention.

Insufficient Teacher Support

Parents worried that teachers' lack of training in online pedagogy, especially in asynchronous contexts, would lead to poor academic outcomes (Lau et al., 2020; Lei et al., 2022). Ye et al. (2022) argued that increasing student-teacher interactions in both synchronous and asynchronous courses would support students academically and emotionally. As parents noted deficiencies in teachers' online course facilitation, they called for training in online pedagogy and assessment (Demir & Yildizli, 2022). Kaphle and Rana (2023) noticed that teachers in Nepal initially struggled to teach through Zoom, but classes worked better when parents sat with their students during online classes. Teachers realized how crucial parents' participation was if students were going to be successful learning online. They relied on parents to get their students to enroll and participate in classes. They needed them to provide technological support and create an environment at home that was conducive to learning.

Despite the fact that teachers needed training for online course facilitation, parents felt that teachers should be allowed to make their own decisions about content delivery. Parents in Timmons et al. (2021) felt that some synchronous instruction improved their children's ability to engage with learning and receive affective and social support. However, teachers were constrained in deciding which modality to use because of directives from other entities like school administration, school boards, and unions. Given their knowledge of both the students' needs and their access to technological tools, parents felt teachers were best equipped to make modality decisions for their classes.

Discussion

This review synthesized parents' perspectives on how instructional modality affects the need adolescent online students have for help from their parents. Our review indicates a steady move in online learning toward more synchronous instruction, although it is not a call for completely synchronous schedules. Parents recognize that adding synchronous instruction does not absolve them of support responsibilities, but it does alleviate some of the heavy demands of online learning support. The desire to add more synchronous instruction is not a call for completely synchronous schedules. The rigidity of live sessions and time-sensitive assignments adds stress (Budhrani et al., 2021) and may even increase parental support responsibilities (Guo et al., 2024). But a meta-analysis showed synchronous instruction had a small but significant positive effect on cognitive engagement (Martin et al., 2021). And offering more synchronous learning options increased parents' perceptions of support from their school (Booth et al., 2021). We recommend that primary/secondary online programs should support parental involvement by incorporating a blend of synchronous and asynchronous instruction into their course offerings. Most research examining the best practices for designing bichronous instruction is based on teachers' perspectives (Martin et al., 2024). This review synthesis uncovers parents' perspectives on what is important when designing bichronous instruction, including consideration of the affordances of modalities, student engagement, and personal circumstances.

Affordances of Modalities

Just as blended learning research provided direction on how to design effective instruction when combining online and in-person learning contexts (Graham et al., 2024), the pandemic propelled researchers to consider how synchronous and asynchronous learning can be effectively combined (Baba et al., 2021). Blending modalities is most successful when teachers carefully plan instruction (Kostelecká et al., 2021) based on the purpose of the course and the affordances of the modalities (Martin et al., 2023). Teachers' firsthand knowledge of their students' needs should also guide instructional decisions (Timmons et al., 2021). Well-planned bichronous instruction requires considering how each modality supports the desired learning outcome. Changing modalities should expand the possibilities for how students engage with their work (Capinding, 2021) and enable students to experience both community and deep learning opportunities in an online context (Radcliffe 2023). Bichronous instruction may provide the best learning outcomes for adolescent online students (Moorhouse & Wong, 2022).

When planning online instruction, teachers should consider using synchronous activities that encourage human interactions. Synchronous interactions can support emotional wellbeing (Beck, 2024), build community through social interactions (Curtis & Werth, 2015; Flynn et al., 2021), enable immediate feedback from teachers (Lei et al., 2022; Radcliffe, 2023), allow teachers to demonstrate how to use the LMS (Beck, 2024), facilitate one-on-one academic help and small group work (Beck, 2024), and improve the transactional distance between students and teachers (Curtis & Werth, 2015). Live interactions with teachers relieve some pressure from parents who typically are not subject-matter experts (Borup et al., 2015). Online instructors confirmed the parent perspectives in our synthesis with a similar list of affordances for synchronous instruction including one-on-one discussions, explanations of material, group work, answering questions, and conducting assessments (Martin et al., 2023).

Asynchronous learning affordances center around flexibility and choice. Communicating announcements, facilitating online discussions, giving feedback, and providing strong instructional materials are examples of effective asynchronous course components (Martin et al. 2023). The ability to learn any time, any place has been a mainstay on online learning (Veletsianos & Houlden, 2019). Autonomous learning continues to be a draw for students who choose online education (Sandberg et al., 2025). It enables a broader selection of courses, flexibility in time and place of learning, and choice in the pace of progress and depth of study (Radcliffe, 2023). Asynchronous activities support engagement across affective, behavioral, and cognitive domains (Beck, 2024).

Student Engagement

Student engagement necessary for academic success is a concern for many parents whose children enroll in online courses (Flynn et al., 2021; Harmon & Arnold, 2023; Shishakly, 2024). In traditional school settings, students interact with content, teachers, and peers. Students' interactions with content in asynchronous contexts is limited primarily to learner-content interactions (Moore, 1989). We found that parents hoped instructional modality decisions would support better student engagement. Parents focused specifically on using synchronous activities to expand interactions with content and facilitate collaboration with peers and teachers (Flynn et al., 2021; Kostelecká et al., 2021; Lei et al., 2022).

These engagement supports can lead students to develop independent engagement skills, which is one desired outcome for parents who enroll students in online programs (Drvodelic & Domović, 2022; Zhang & Halpern, 2021). Students' engagement abilities vary significantly across individuals (Borup et al., 2020). Young students and students who are new to online learning often struggle with autonomy and locus of control. They require support from either teachers or parents to engage academically (Cavanaugh et al., 2009). How that support is offered can help or hinder students in developing independent engagement (Sandberg et al., 2026). Parents may need to step back from providing support to help their students develop independent self regulation skills (Borup et al., 2019). Teachers also should consider how their course design can support the development of independent engagement, like the color coding that helped students navigate the LMS with less parental supervision (Beck, 2024).

Personal Circumstances

Parents, who have been called "co-educators" (Waters & Leong, 2014, p. 52), typically assume their role without formal training, professional development, or teaching experience (Downes, 2013). This means that the weight of what parents are asked to do can feel heavy, with some parents even reporting experiencing more depression when they had to act as their child's teacher (Burns et al., 2023). In online contexts, parents must construct new roles because their involvement differs from traditional school contexts (Sandberg et al., 2026). Similarly, online schools need to adopt new support roles with a dual purpose of supporting both students and their parents, especially for parents of students with special accommodation needs (Munsell et al., 2021). When making decisions about how to deliver course content, teachers or online programs should consider how the modality will interact with the circumstances of students and their families.

The effect modality decisions have on family relationships is complicated. Limited synchronous instruction can actually increase the stress families feel. Time-constrained deadlines for assignments in live classes increase students' demands for immediate help and reduce flexibility (Budhrani et al., 2021). Heavy demands on parents' time can hinder them from providing adequate support to their students even if they are willing and able to do so (Hoover-Dempsey and Sandler, 2005). However, synchronous instruction can also alleviate stress and improve bi-directional communication, allowing parents to focus on strengthening their parent-child relationships and have insights into their student as a learner (Bhamani et al., 2020; Budhrani et al., 2021; Flynn et al., 2021; Harmon & Arnold, 2023; Roy et al., 2022). If online programs plan bichronous instruction that makes parent support more feasible, then parents can focus more on these positive outcomes of providing online learning support.

Finally, online programs must recognize that engagement and family support are not the only issues affected by modality decisions. In fact, modality decisions are central digital equity. Not only are students with unequal access to digital tools affected by modality decisions, but students of diverse cultural backgrounds may also be affected. Privacy and discrimination concerns were a significant barrier for women and minorities in synchronous modalities (Khlaif et al., 2021; Zhang & Halpern, 2021). Modality choices are not neutral; they have the potential to adversely affect underprivileged students and create additional barriers to learning. As online learning moves toward increased use of synchronous instruction, designers must ensure equitable access by creating policies that are culturally inclusive, particularly those that affect privacy and identity like camera policies and participation expectations.

Review Limitations

This review, as mentioned in the methodology section, followed a narrative synthesis methodology with systematized procedures, but is not a full systematic review methodology. Consequently, the findings are not as generalizable as a systematic review. Specifically, the findings are limited by not completing searches in all databases, although the use of two databases, reference searches, and AI-assisted searching provided a robust collection of papers. In addition, the review is limited by not employing double coding by multiple researchers and calculating interrater reliability, instead relying on coding/screening by the primary author with the second author double coding the articles categorized as implicitly focused on a modality. This was done because the articles with implicit modality definitions were the ones most open to subjective interpretation. Still, the lack of double coding of all articles is a limitation.

There are further limitations in the articles selected for this review. The substantial number of studies conducted during the pandemic is partially due to the limited number of articles that discuss modality prior to the pandemic. The COVID studies likely reflect parental perceptions that are based on emergency remote teaching rather than intentionally planned online learning. The studies selected reflect a global perspective but may be biased in that the databases drew from English language journals which may have excluded international or grey literature. It is also important to acknowledge the potential of algorithmic bias from the AI-generated search, although the number of included articles from Elicit was small.

Implication for Future Research

This synthesis of literature revealed a growing trend toward including more synchronous instruction in primary/secondary online learning. We have advocated for practitioners and online programs to carefully consider the affordances of online learning, student engagement, and personal circumstances as they blend synchronous and asynchronous learning. Instructional modality is one significant way that online programs distinguish themselves from each other so that parents can make decisions about the type of learning experience they feel best matches their student's needs (Sandberg et al., 2025). Online programs help families choose carefully when they make their decisions about content delivery transparent.

Although we know that instructional modality has the potential to impact the type and amount of support parents need to provide adolescent online learners, there is very little research exploring the intersection of parental support roles and instructional modality. We recommend that researchers give more overt consideration to instructional modality. Despite the impact of instructional modality decisions, we found that many researchers do not address the topic directly. Thus, our first recommendation is to make modality an overt part of defining the study context for research into online learning, and that surveys and questionnaires ask parents to describe their instructional context. Second, these online learning contexts are often treated as homogeneous, when in reality, they have significant variation, and so we recommend that researchers analyze data about online learning taking modality specifically into account as they are different learning environments.

Implication for Future Practice

One implication of this review is a call for schools to support parents in adjusting to their new roles as they support their online learners. Part of this comes from recognizing the different stresses placed on the family from modality choices, as synchronous learning may relieve some of the pressure parents feel in co-teaching their children, but may affect issues of equity and access to learning. We recommend that adolescent online programs should utilize a blend of synchronous and asynchronous instruction to support learners and provide choices that families can use in deciding how and whether to learn online. Offering modality choices may provide more equitable opportunities for students to participate regardless of their gender, cultural background, access to technology, or learning accommodations. In addition, many times asynchronous programs can offer synchronous mentoring options for students that need it, which can be helpful for parents. Finally, schools can provide resources and communities of support for parents as they navigate and learn their new roles in supporting their students.

Conclusion

This review of literature extends the ACE framework applications of parental engagement support and shows how instructional modality affects the support parents provide to primary/secondary online learners. Implications from this study suggest that online programs should offer careful blends of asynchronous and synchronous instruction that aim to lighten parents' support load, improve student engagement, and consider the individual circumstances of learners. Research has not fully captured what is happening in primary/secondary online education as online enrollments have skyrocketed since the pandemic. The implications for

families, particularly of young learners, are significant. Research about how to better support the supporters is crucial if we hope to have parents fill the gap of academic engagement needs that is necessary for online students to have academic success.

Declarations

Acknowledgments

The authors acknowledge the guest editors for this special edition for inviting the submission of this manuscript.

Ethics Statement

This study was conducted in accordance with recognized ethical standards. We did not obtain permission from our Institutional Review Board for this research because no human subjects were involved in this project.

Funding Statement

This research received no external funding.

AI Use Statement

Generative artificial intelligence (AI) was used in the preparation of this manuscript for the purpose of triangulating the literature search and editing the final writing to reduce the overall number of words to meet the journal requirements. All content was reviewed, verified, and approved by the authors, who take full responsibility for the accuracy and integrity of the work.

Conflicts of Interest Statement

The authors declare no conflicts of interest related to the research, authorship, or publication of this article.

References

Averett, K. H. (2021). Remote learning, COVID-19, and children with disabilities. *AERA Open, 7*(1), 1-12. https://doi.org/10.1177/23328584211058471

Baba, K., Elfaddouli, N., & Cheimanoff, N. (2021). A comparative study of synchronous and asynchronous learning during COVID-19 crisis. In 4th International Academic Conference on Teaching and Learning (pp. 38-49).

Beck, D. (2024). At-risk and online: parent perceptions of at-risk learner's supports in a fully online school. *Cogent Education, 11*(1). https://doi.org/10.1080/2331186X.2024.2317110

Bhamani, S., Makhdoom, A. Z., Bharuchi, V., Ali, N., Kaleem, S., & Ahmed, D. (2020). Home learning in times of COVID: Experiences of parents. *Journal of Education and Educational Development, 7*(1), 9-26. https://doi.org/10.22555/joeed.v7i1.3260

Booth, C., Villadsen, A., Goodman, A., & Fitzsimons, E. (2021). Parental perceptions of learning loss during COVID-19 school closures in 2020. *British Journal of Educational Studies, 69*(6), 657-673. https://doi.org/10.1080/00071005.2021.1984390

Borup, J., Graham, C. R., West, R. E., Archambault, L., & Spring, K. J. (2020). Academic communities of engagement: An expansive lens for examining support structures in blended and online learning. *Educational Technology Research and Development*, *68*(2), 807-832. https://doi.org/10.1007/s11423-020-09744-x

Borup, J., Stevens, M. A., & Waters, L. H. (2015). Parent and student perceptions of parent engagement at a cyber charter high school. *Online Learning, 19*(5), 69-91. https://doi.org/10.24059/olj.v19i5.699

Borup, J., Walters, S., & Call-Cummings, M. (2019). Examining the complexities of parental engagement at an online charter high school: A narrative analysis approach. *International Review of Research in Open and Distributed Learning, 20*(1). https://doi.org/10.19173/irrodl.v20i1.3605

Borup, J., West, R. E., Graham, C. R., & Davies, R. S. (2014). The adolescent community of engagement framework: A lens for research on K-12 online learning. *Journal of Technology and Teacher Education*, *22*(1), 107-129. Retrieved June 26, 2023 from https://www.learntechlib.org/primary/p/112371/

Budhrani, K., Martin, F., Malabanan, O., & Espiritu, J. L. (2021). How did parents balance it all? Work-from-home parents' engagement in academic and support roles during remote learning. *Journal of Online Learning Research, 7*(2), 153-184.

Burns, S., Jegatheeswaran, C. & Perlman, M. (2023). I felt like I was going crazy: Understanding mother's and young children's educational experiences at home during COVID-19. *Early Childhood Education Journal, 51*, 469–482. https://doi.org/10.1007/s10643-022-01306-5

Capinding, A. T. (2021). Mathematics learning of high school students and the determinants of their performance in the midst of pandemic. *Online Submission Turkish Journal of Computer and Mathematics Education, 12*(12), 2995-3014.

Cavanaugh, C. S., Barbour, M. K., & Clark, T. (2009). Research and practice in K-12 online learning: A review of open access literature. *International Review of Research in Open and Distributed Learning, 10*(1). https://doi.org/10.19173/irrodl.v10i1.607

Collado, Z. C., Rodriguez, V. R., & Dueñas III, Z. D. (2023). Children's engagement in self-learning modules (SLMs) amid the pandemic: A predictive analysis on the role of internet access, household food security, and parental involvement to modular classes. *International Journal of Primary, Elementary and Early Years Education 51*(1), 107-120. https://doi.org/10.1080/03004279.2021.1954969

Curtis, H., & Werth, L. (2015). Fostering student success and engagement in a K-12 online school. *Journal of Online Learning Research*, *1*(2), 163-190. Retrieved June 27, 2023 from https://www.learntechlib.org/primary/p/150962/

Delany-Barmann, G., Paciotto, C., & Meeker, L. (2021). Navigating the educational needs of rural multilingual students during the pandemic: Parents' experiences and perspectives. *Thresholds in Education, 44*(2), 180-205.

Demir, M. R., & Yildizli, H. (2022). Educational processes and learning at home during COVID-19: Parents' experiences with distance education. *International Review of Research in Open and Distributed Learning, 23*(3), 1-20. https://doi.org/10.19173/irrodl.v23i2.5870

Dorji, T. (2021). Problems and challenges faced by students, parents, and teachers in Google classes & television lessons. *Online Submission International Journal of Asian Education, 2*(1), 88-97. https://doi.org/10.46966/ijae.v2i1.98

Downes, N. (2013). The challenges and opportunities experienced by parent supervisors in primary school distance education. *Australian and International Journal of Rural Education, 23*(2), 31-41. https://search.informit.org/doi/10.3316/informit.430349660194263

Drvodelic, M., & Domović, V. (2022). Parents' opinions about their children's distance learning during the first wave of the Covid-19 pandemic. *CEPS Journal, 12*(3), 221-241. https://doi.org/10.25656/01:25507

Flynn, N., Keane, E., Davitt, E., McCauley, V., Heinz, M., & Mac Ruairc, G. (2021). 'Schooling at home' in Ireland during COVID-19': Parents' and students' perspectives on overall impact, continuity of interest, and impact on learning. *Irish Educational Studies, 40*(2), 217-226. https://doi.org/10.1080/03323315.2021.1916558

Garbe, A., Ogurlu, U., Logan, N., & Cook, P. (2020). COVID-19 and remote learning: Experiences of parents with children during the pandemic. *American Journal of Qualitative Research*, *4*(3), 45-65. https://doi.org/10.29333/ajqr/8471

Graham, C. R., Slaugh, B., McMurry, A., Sorensen, S. D., Arnesen, K. T., & Ventura, B. (2024). Blended teaching in higher education: A guide for instructors and course designers. EdTech Books. https://edtechbooks.org/he_blended

Guo, Q., Graham, C. R., Borup, J., Sandberg, B., & West, R. E. (2024). Parental support challenges for K-12 student online engagement. *Distance Education, 45*(4), 579-605. https://doi.org/10.1080/01587919.2024.2397481

Harkins, C., Sadikova, E., Brunt, S., Swanstrom, A., Menezes, M., & Mazurek, M. O. (2022). Caregiver perspectives regarding special education service changes amid COVID-19. *The Journal of Special Education, 57*(2), 63-72. https://doi.org/10.1177/00224669221113528

Harmon, T. M., & Arnold, D. H. (2023). Parent perceptions of remote learning during COVID-19. *The Elementary School Journal, 124*(2), 193-218. https://doi.org/10.1086/727453

Hasler Waters, L., & Leong, P. (2014). Who is teaching? New roles for teachers and parents in cyber charter schools. *Journal of Technology and Teacher Education, 22*(1), 33–56. Retrieved June 27, 2023 from https://www.learntechlib.org/primary/p/112373/

Hoover-Dempsey, K. V., Walker, J. M., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., & Closson, K. (2005). Why do parents become involved? Research findings and implications. *The Elementary School Journal*, *106*(2), 105-130. https://doi.org/10.1086/499194

Kaphle, U., & Rana, K. (2023). Teachers' capability, students' motivation and parents' support in online learning. *International Journal of Technology in Education and Science, 7*(2), 192-210. https://doi.org/10.46328/ijtes.442

Khlaif, Z. N., Salha, S., Fareed, S., & Rashed, H. (2021). The hidden shadow of Coronavirus on education in developing countries. *Online Learning, 25*(1), 269-285. https://doi.org/10.24059/olj.v25i1.2287

Kostelecká, Y., Komárková, T., & Novotná, V. (2021). Remote home-based education as a new phenomenon in the time of the covid-19 pandemic–the experience of Czech families. *Journal of Pedagogy, 12*(1), 141-164. https://doi.org/10.2478/jped-2021-0007

Lau, E. Y. H., & Lee, K. (2020). Parents' views on young children's distance learning and screen time during COVID-19 class suspension in Hong Kong. *Early Education and Development, 32*(6), 863–880. https://doi.org/10.1080/10409289.2020.1843925

Lee, T. T. (2022). Leadership for inclusive online learning in public primary schools during COVID-19: A multiple case study in Hong Kong. *Educational Management Administration & Leadership, 52*(6), 1434-1454. https://doi.org/10.1177/17411432221135310

Lei, J., Liu, Y. & Salim, Z. (2022). A revised curriculum design framework for emergency remote education: lessons from parents' perspectives of K-12 online/hybrid Instruction during

Covid-19. *International Journal of Technology in Teaching and Learning, 18*(2) 101-11. https://doi.org/10.37120/ijttl.2022.18.2.02

Martin, F., Kumar, S., Ritzhaupt, A. D., & Polly, D. (2023). Bichronous online learning: Award-winning online instructor practices of blending asynchronous and synchronous online modalities. *The Internet and Higher Education*, *56*(2023), Article 100879. https://doi.org/10.1016/j.iheduc.2022.100879

Martin, F., Kumar, S., Ritzhaupt, A., & Polly, D. (2024). Bichronous online learning: Perspectives, best practices, benefits, and challenges from award-winning online instructors. *Online Learning, 28*(2), n2. https://doi.org/10.24059/olj.v28i2.3945

Martin, F., Polly, D., & Ritzhaupt, A. (2020). Bichronous online learning: Blending asynchronous and synchronous online learning. *Educause Review, 8*. https://er.educause.edu/articles/2020/9/bichronous-online-learning-blending-asynchronous-and-synchronous-online-learning

Martin, F., Sun, T., Turk, M., & Ritzhaupt, A. (2021). A meta-analysis on the effects of synchronous online learning on cognitive and affective educational outcomes. *International Review of Research on Open and Distributed Learning, 22*(3), 205-242. https://doi.org/10.19173/irrodl.v22i3.5263

Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, 3, 1–7. https://doi.org/10.1080/08923648909526659

Moorhouse, B. L., & Wong, K. M. (2022). Blending asynchronous and synchronous digital technologies and instructional approaches to facilitate remote learning. *Journal of Computers in Education*, 9(1), 51-70. https://doi.org/10.1007/s40692-021-00195-8

Munsell, E. G., Schwartz, A. E., Schmidt, E. K., & Chen, J. (2021). "Everything now falls on me": Parent perspectives on services lost and challenges during virtual learning for youth with disabilities. *Journal of the American Academy of Special Education Professionals, 51*, 67.

Naungayan, R. R., Vivencio, I. J. B., & Caplis, M. M. (2024). Involvement of working parents in students' academic achievement: Basis of support programs. *International Journal of Research in Education and Science*, *10*(1), 62-82. https://doi.org/10.46328/ijres.3281

Ortiz, K. R., Rice, M. F., Curry, T., Mellard, D., & Kennedy, K. (2021). Parent perceptions of online school support for children with disabilities. *American Journal of Distance Education, 35*(4), 276–292. https://doi.org/10.1080/08923647.2021.1979343

Philpott, D., Sharpe, D., Neville, R. & Neville, R. (2010). The effectiveness of web-delivered learning with aboriginal students: Findings from a study in coastal Labrador. *Canadian Journal of Learning and Technology / La revue canadienne de l'apprentissage et de la technologie, 35*(3),. Canadian Network for Innovation in Education. Retrieved January 7, 2026 from https://www.learntechlib.org/p/42882/.

Radcliffe, C. (2023). The urbanization of distance education in Australia. *Quarterly Review of Distance Education, 24*(1), 43-54.

Ribeiro, L. M., Cunha, R. S., Silva, M. C. A. E., Carvalho, M., & Vital, M. L. (2021). Parental involvement during pandemic times: Challenges and opportunities. *Education Sciences, 11*(6), 302. https://doi.org/10.3390/educsci11060302

Roy, A. K., Breaux, R., Sciberras, E., Patel, P., Ferrara, E., Shroff, D. M., ... & Becker, S. P. (2022). A preliminary examination of key strategies, challenges, and benefits of remote learning expressed by parents during the COVID-19 pandemic. *School Psychology, 37*(2), 147. https://doi.org/10.1037/spq0000465

Saliverou, M., Georgiadi, M., Tomprou, D. M., Loizidou-Ieridou, N., & Plexousakis, S. (2021). The impact of COVID-19 pandemic containment measures on families and children with moderate and high-functioning ASD (Autism Spectrum Disorder). *Education Sciences, 11*(12), 783. https://doi.org/10.3390/educsci11120783

Sandberg, B., Erickson, H., Graham, C. R., Borup, J., & West, R. (2025). Choosing Virtual: Understanding the Forces that Drive Parents Toward Online K-12 Education. Journal of School Choice, 1–17. https://doi.org/10.1080/15582159.2025.2534005

Sandberg, B. T., Graham, C. R., Borup, J., West, R. E., & Velasquez, R. Q. (2026). Behind the screen: Exploring parental roles in K-12 online education. *Journal of Research on Technology in Education, 58(3)*, 555–570. https://doi.org/10.1080/15391523.2024.2447729

Shishakly, R. (2024). Young children's online homeschooling during Covid-19 closure in the United Arab Emirates: Parents' experiences. *Issues in Educational Research, 34*(1), 261-279. http://www.iier.org.au/iier34/shishakly.pdf

Smetackova, I., & Stech, S. (2021). The first wave of the COVID-19 pandemic in primary schools in the Czech Republic: Parental perspectives. *European Journal of Education, 56*(4), 564-577. https://doi.org/10.1111/ejed.12478

Sutton, A., Clowes, M., Preston, L. et al. (2019) Meeting the review family : exploring review types and associated information retrieval requirements. Health Information and Libraries Journal, 36 (3). pp. 202-222. ISSN: 1471-1834

Timmons, K., Cooper, A., Bozek, E., & Braund, H. (2021). The impacts of COVID-19 on early childhood education: Capturing the unique challenges associated with remote teaching and learning in K-2. Early Childhood Education Journal, 49(5), 887-901. https://doi.org/10.1007/s10643-021-01207-z

Tricco, A. C., Antony, J., Soobiah, C., Kastner, M., Cogo, E., MacDonald, H., ... & Straus, S. E. (2016). Knowledge synthesis methods for generating or refining theory: a scoping review reveals that little guidance is available. Journal of Clinical Epidemiology, 73, 36-42.

Valcourt-Pearce, C. C. (2022). Shifting priorities, technology-centric practices, and an ever-evolving educational experience for our children. *Odyssey: New Directions in Deaf Education*, 22, 56-61.

Veletsianos, G., & Houlden, S. (2019). An analysis of flexible learning and flexibility over the last 40 years of Distance Education. *Distance Education, 40*(4), 454-468. https://doi.org/10.1080/01587919.2019.1681893

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(5), 2443-2476. https://doi.org/10.1007/s11423-023-10233-0

Wilson-Forsberg, S., Kimani-Dupuis, R., & Masakure, O. (2024). Weathering the storm: How mothers with refugee backgrounds helped their children with school during the COVID-19 pandemic. *International Journal of Multicultural Education, 26*(2), 15-37. https://doi.org/10.18251/ijme.v26i2.3853

Ye, Y., Wang, C., Zhu, Q., He, M., Havawala, M., Bai, X., & Wang, T. (2022). Parenting and teacher–student relationship as protective factors for Chinese adolescent adjustment during COVID-19. *School Psychology Review, 51*(2), 187–205. https://doi.org/10.1080/2372966X.2021.1897478

Yıldırım, H. H., & Rovshenov, A. (2022). Exploring problems of moms with children with autism spectrum disorder during COVID-19. *Education Quarterly Reviews, 5*(3). Available at SSRN: https://ssrn.com/abstract=4182320

Zhang, J., & Halpern, C. (2021). The experiences of a Chinese immigrant family during the COVID-19 pandemic. *Educational Research and Development Journal, 24*(2), 52-72.

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5816
https://doi.org/10.24059/olj.v30iSI2.5816

Refining Collaborative AI Literacy: A Systematic Review of Human-AI Collaboration Competency and AI Literacy Measures

Ji Won Yang[1], Junseo Kim[2], Soojeong Jeong[3], and Jaeseo Lim[4]

[1]College of Social Sciences, Seoul National University, Seoul, Republic of Korea
[2]Department of Education, Seoul National University, Seoul, Republic of Korea
[3]Department of Educational Technology, Hanyang University, Seoul, Republic of Korea
[4]Department of Counseling Psychology, Jeonju University, Jeonju, Republic of Korea

Abstract

With the widespread adoption of artificial intelligence (AI) and its growing role in solving complex problems, human-AI interaction has evolved into a new dimension in which AI is increasingly perceived as a collaborative partner. However, current AI literacy frameworks primarily focus on technical knowledge and ethical use, often overlooking the ability to collaborate effectively with AI as a distinct competency. Because conceptual boundaries between human-AI collaboration competency and AI literacy have yet to be clearly established, this systematic review examines how human-AI collaboration competency is represented and measured within existing AI literacy scales. A PRISMA-guided search of five major databases identified peer-reviewed studies published since 2020 that examined AI literacy or human-AI collaboration-related competencies in higher education and reported measurement items or scales. Methodological quality was assessed through reported psychometric evidence, and findings were synthesized narratively. Synthesis of 17 eligible studies indicates that human-AI collaboration is rarely assessed explicitly, let alone as a distinct construct. Some studies included a small number of items that could implicitly refer to collaboration competency or context, but most showed no direct or indirect representation. Based on these findings, the review discusses theoretical implications and measurement considerations for developing collaborative AI literacy measures, with particular attention to three sub-factors of human-AI collaboration competency: coordination, communication, and metacognition. The findings provide a foundation for reconceptualizing human-AI collaboration competency conventionally positioned within AI literacy, and for developing measurement tools that can assess and foster collaborative AI literacy.

Keywords: AI literacy; human-AI collaboration; collaborative AI literacy; measurement; systematic review

Introduction

Many challenges in increasingly complex modern societies resemble “wicked problems” that cannot be addressed through individual effort alone. Traditionally, addressing such problems “alone” implied the absence of collaboration among human counterparts. However, with the widespread adoption of AI, the scope of collaboration has expanded to include human-AI interaction as well. Yet, simply using AI does not necessarily lead to better outcomes. A recent meta-analysis indicates the effectiveness of human-AI combinations varies substantially across empirical studies (Vaccaro et al., 2024), suggesting that outcomes may depend on how humans collaborate with AI rather than on AI use per se. This finding calls for closer examination of what distinguishes effective human-AI collaboration from simply using AI as a tool, and of how it compares with and differs from human-human collaboration.

Compared with AI literacy, for which frameworks and measures have begun to be established, research on human-AI collaboration competency remains scarce. Prior literature suggests human-AI collaboration competency is a multifaceted, complex construct that requires greater conceptual clarity, yet it has not been comprehensively theorized. The absence of a clear conceptual framework poses a major obstacle to the development of such competency,

Corresponding authors: Soojeong Jeong (sjeong315@hanyang.ac.kr) and Jaeseo Lim (jaeseolim@jj.ac.kr)
This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 The Author(s)

particularly for evaluating and eventually designing effective educational interventions (Hattie, 2008).

To date, AI-related competencies have primarily been conceptualized under the umbrella of AI literacy, typically defined in terms of knowledge, critical thinking, and ethical awareness rather than collaborative ability (Chiu, 2025). With growing attention to AI literacy, research on its conceptualization and measurement has expanded. Consequently, items that assess human-AI collaboration competency are often treated as sub-constructs within broader AI literacy measures (Sidra & Mason, 2025). Moreover, even for AI literacy itself, many intervention studies have relied on ad hoc measure development or adaptation, reflecting the limited availability of validated scales for assessing changes in AI competencies (S. C. Kong et al., 2025; S. C. Kong & Yang, 2025). These limitations underscore the need to reconceptualize AI literacy in a way that explicitly accounts for collaboration with AI and to establish a clearer conceptual foundation for measuring collaborative AI literacy.

In this view, "collaborative AI literacy" should be conceptualized and measured as a distinct competency that goes beyond knowledge and understanding of AI technology (Long & Magerko, 2020). To address these gaps, this study presents a systematic review synthesizing research on human-AI collaboration competency and AI literacy, with the aim of clarifying their conceptual boundaries. Drawing on current literature, we examine studies that measure and validate the constructs of AI literacy and human-AI collaboration competency and propose a refined conceptualization of collaborative AI literacy. This review aims to provide a foundation for future research on developing measurement tools and instructional guidelines that can assess and foster collaborative AI literacy.

Literature Review

AI Literacy Frameworks and Measurement Approaches

Literacy, once defined as the ability to read and write, now widely encompasses diverse competencies across educational and technological contexts. As AI becomes increasingly integrated into society, AI literacy has emerged as one such competency. Most notably, Long and Magerko (2020) define AI literacy as a set of competencies that allow individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use these technologies in goal-directed ways. This definition frames AI literacy as a capability that extends beyond technical knowledge or programming skills to include functional understanding, critical evaluation, and practical application.

This multidimensionality is reflected in the factor structures and measurement items proposed in AI literacy scale development studies. The Scale for the Assessment of Non-Experts' AI Literacy (SNAIL; Laupichler et al., 2023a) identifies a three-factor structure consisting of Technical Understanding, Critical Appraisal, and Practical Application. Ng et al. (2021b), whose framework serves as a foundation for many AI literacy scales, conceptualize four sub-constructs of AI literacy: know and understand, use and apply, evaluate and create, and ethical issues. Adapted by Carolus et al. (2023), items in the use and apply AI construct include statements such as "I can operate AI applications," whereas items measuring knowledge and understanding of AI include "I know definitions of artificial intelligence." Items assessing ethical

considerations capture the ability to weigh the risks and responsibilities of AI use and its broader societal consequences (e.g., “I can weigh the consequences of using AI for society”).

Prior studies have also incorporated factors such as motivation, self-efficacy, and behavioral intention, thereby broadening AI literacy beyond cognitive skills (Laupichler et al., 2022). The Meta AI Literacy Scale (MAILS; Carolus et al., 2023) includes items assessing self-efficacy in learning about and solving problems with AI, emotion regulation, and persuasion literacy (the ability to recognize and resist AI's influence on one's own decisions), emphasizing the role of self-regulation in AI use. Ng et al. (2024) integrated affective disposition, behavioral engagement, and ethical reasoning alongside cognitive dimensions of AI literacy, proposing the ABCE (affective, behavioral, cognitive, ethical) framework. Taken together, these efforts suggest that AI literacy comprises a broader and more complex set of competencies.

Reviews of this rapidly growing body of instruments, however, converge on the conclusion that measurement has not kept pace with conceptualization (Almatrafi et al., 2024; Zhang et al., 2025). Synthesizing 22 validation studies across 16 scales, Lintner (2024) found that existing AI literacy scales rely on self-report formats and have yet to accumulate sufficient psychometric evidence beyond structural validity and internal consistency. Performance-based instruments such as the AI Literacy Test and the AI Concept Inventory directly assess learners’ knowledge of what AI can do, how it operates, and its broader social implications through multiple-choice questions or classification tasks (S. C. Kong et al., 2024). However, these assessments primarily target factual understanding or declarative knowledge and do not sufficiently reflect the interaction and communication context of human-AI collaboration (Jin et al., 2025). While literature suggests that the construct of AI literacy is evolving beyond knowledge and functional use of AI, competencies and processes involved in collaborating with AI have yet to be adequately measured, much less conceptualized.

Conceptual and Theoretical Gaps in Measuring Collaborative AI Literacy

Collaboration has traditionally been defined as a process in which two or more individuals interact to achieve a shared goal (Dillenbourg, 1999). Rather than simply serving as tools for computation, AI systems are increasingly positioned as active collaborators, with human counterparts regulating AI-generated outputs and integrating them into the task at hand (Dellermann et al., 2019). This perspective is grounded in distributed cognition theory, which posits that cognition is not confined to the individual but is distributed across interactions among humans, tools, and environments (Hollan et al., 2000). Human-AI collaboration forms a hybrid intelligence system that extends human cognition through a division of cognitive labor, in which generative AI generates and transforms information while humans perform goal setting, critical judgment, and integration (Artman & Garbis, 1998; Lodge et al., 2023; Siemens et al., 2022).

When cognition is understood as being constructed through interaction, the outcomes of human-AI collaboration are shaped not by users’ technical proficiency alone but also by the quality of interaction. Because AI systems do not possess self-awareness or a shared mental model in the human sense, the responsibility for determining the scope and level of information required for collaboration and for continuously monitoring the interaction falls primarily on the human user (Sidra & Mason, 2025). Specifically, communication involves providing AI with appropriate context and information, as well as critically interpreting its outputs, whereas

coordination refers to managing the flow of interaction and the distribution of roles during joint task performance. Thus, human-AI collaboration requires competencies for regulating real-time interaction and co-constructing meaning with AI (Long & Magerko, 2020), offering a conceptual basis for the emerging notion of collaborative AI literacy.

At its core, collaborative AI literacy goes beyond using AI as a tool and focuses on the competencies needed to engage with AI in joint problem-solving and shared meaning-making through interdependent interaction. Prior literature also suggests including metacognitive regulation, defined as the ability to continuously monitor and adjust one's thinking during interaction with AI (Fan et al., 2025; Tankelevitch et al., 2024). In learning contexts involving generative AI (GenAI), some studies conceptualize AI use in terms of self-regulation and interaction, incorporating dimensions of learner autonomy such as goal setting, self-adjustment, self-reflection, and participatory action, as well as reflective and collaborative use of AI (Hou et al., 2025; Xia et al., 2025).

Although recent work has begun to address collaboration competency between humans and AI, it often does so by indirectly reflecting collaborative elements or contexts at the item level within existing AI literacy scales. Rather than fully measuring the capabilities that drive effective human-AI performance as a standalone construct, studies that define and measure constructs related to collaborative AI literacy primarily focus on conventional AI literacy dimensions such as AI evaluation, AI use, and AI ethics. For instance, items typically emphasize the ability to identify AI errors or to direct AI toward desired outputs (Sidra & Mason, 2025). As a result, these measures are largely limited to assessing individuals' knowledge and skills for judging AI performance, using AI in goal-directed ways, and recognizing ethical implications, remaining closer to the conceptual boundaries of AI literacy.

Purpose of the Study

Despite rapidly increasing social and educational demands for human-AI collaboration competency, existing AI literacy frameworks and measures predominantly emphasize knowledge acquisition and functional use, with limited attention to collaboration and interaction with AI (Carolus et al., 2023; S. C. Kong et al., 2024; Laupichler et al., 2022; Long & Magerko, 2020). Within these measures, human-AI collaboration competency is rarely measured as a distinct construct and, when included, tends to be represented indirectly through a few items placed under other AI-related sub-constructs. To address this gap, the present review synthesizes research on existing conceptualizations and measures of AI literacy and human-AI collaboration competency, with a specific focus on whether collaboration-related dimensions are represented explicitly, implicitly, or not at all. Building on these findings, the review aims to propose a refined conceptualization of collaborative AI literacy that can inform future measurement development and empirical investigation.

This review addresses the following research questions:

RQ1. How have AI literacy and human-AI collaboration competency been defined, conceptualized, and measured across existing studies?

RQ2. To what extent are human-AI collaboration competency and related constructs explicitly or implicitly represented within existing AI literacy and related measures?

RQ3. What theoretical and measurement implications emerge for reconceptualizing AI literacy as collaborative AI literacy?

Methods

This review was designed, conducted, and reported in accordance with the updated PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement and guideline (Page et al., 2021). The completed PRISMA checklist is provided in Appendix B.

Search Strategy

Literature was searched across five databases most relevant to our topic: Web of Science, Scopus, ERIC (EBSCO), PsycINFO (EBSCO), and IEEE Xplore, ensuring comprehensive coverage of education, psychology, and computer science literature. All databases were searched in January 2026. Search strings combined terms from four categories, with the categories connected by AND: (1) AI-related keywords ("artificial intelligence" OR AI OR "generative ai"), (2) literacy and collaboration constructs (literac* OR "ai literacy" OR collaborat* OR "ai collaboration"), (3) education/learning and student context ((educat* OR learn*) AND student*), (4) measurement-related terms (survey* OR questionnaire* OR "self-report*" OR instrument* OR scale* OR framework*). The full search queries for each database are provided in Appendix C. We did not perform supplementary citation searching or hand-searching, and study identification relied solely on the five databases listed above. Given the breadth of the search terms and the cross-disciplinary coverage of the databases, we judged the risk of omitting eligible studies to be low. Following the initial search, the results were uploaded to Rayyan (https://www.rayyan.ai/) for data management, duplicate removal, and systematic screening based on the pre-defined inclusion and exclusion criteria. Duplicate records were removed using Rayyan's automated duplicate-detection function (Kim et al., 2026), and records flagged as potential duplicates were manually verified before removal.

Inclusion and Exclusion Criteria

Our inclusion criteria comprised full-text, peer-reviewed articles in English that measured and validated human-AI collaboration competency and/or AI literacy using newly developed scales or items. Studies published before 2020 were excluded, as the conceptual foundations of current AI literacy discourse, most notably the widely adopted definition by Long and Magerko (2020), had not yet been established, and generative AI use contexts had not yet emerged. We excluded studies conducted outside higher education, as terminology and constructs may differ across age groups. We also limited our review to studies assessing students' competencies, excluding studies focused on teachers or on intervention effectiveness. The specific inclusion criteria are presented in Table 1.

Table 1

Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Time period	Published up to January 2026	Published before 2020
Type of article	Peer-reviewed journal publications	Not peer-reviewed journal publications, conference papers/proceedings, reviews, book chapters
Context & Target Group	Students in Higher Education	- Not related to educational context - Only K-12 students or younger - Professionals outside higher education (including lifelong learning) - Focuses on teachers' competencies
Study focus	Studies that measure and/or validate human-AI collaboration competency or AI literacy using newly developed scales/items	- - - - No survey or items (experimental studies without instrument development) - Only back-translated existing scales - Curriculum design studies (without scale/item development or validation) - Focus on intervention effectiveness, learning outcomes, or perceptions/awareness of AI tools - Focused specifically on AR, VR, IoT, or chatbots

Study Selection

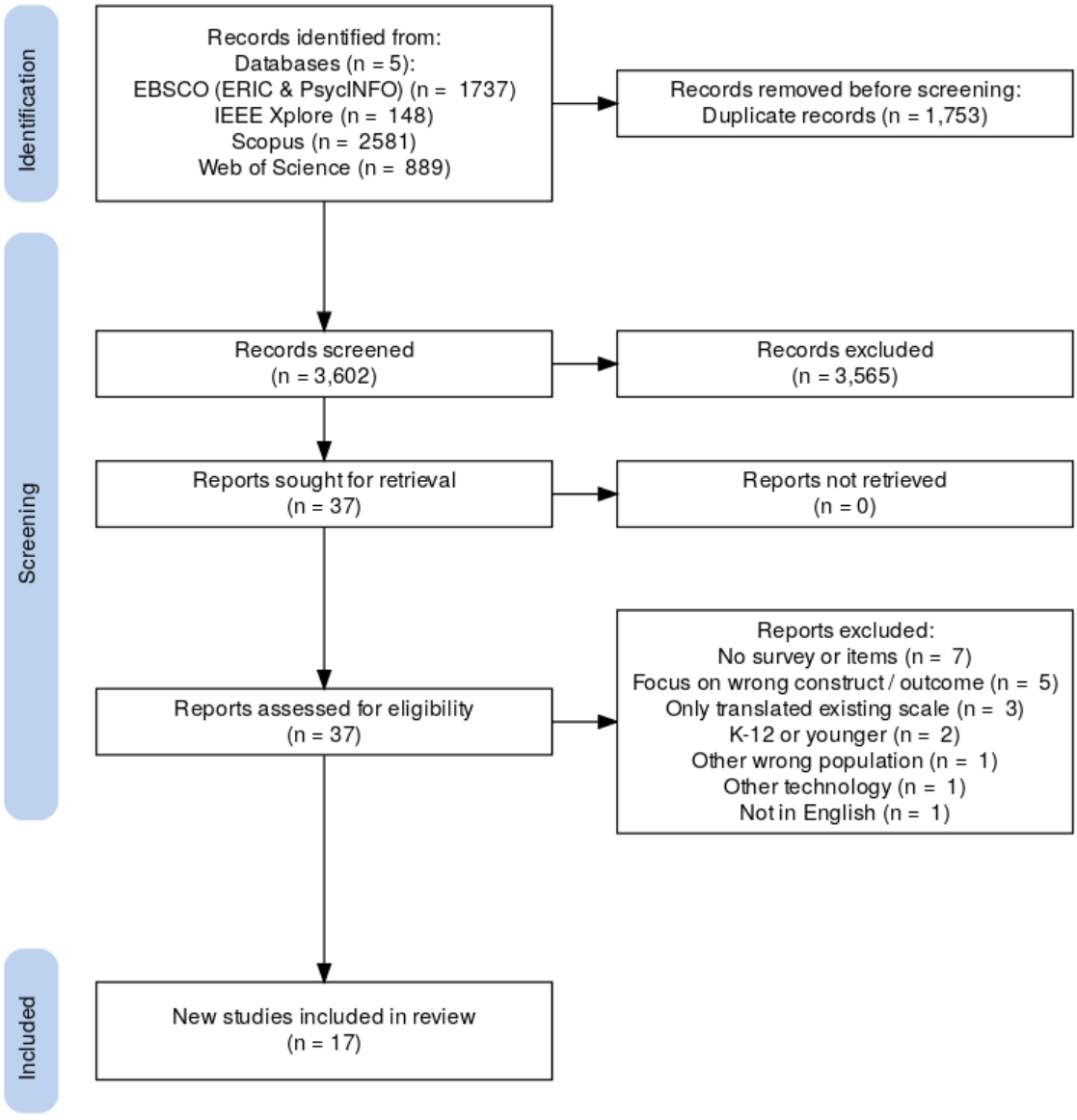
A total of 5,355 records were identified from five academic databases. Retrieved records were equally distributed among three authors, who independently conducted title and abstract screening based on predefined inclusion and exclusion criteria. Inter-rater reliability at the screening stage was assessed using Fleiss' kappa on a randomly selected common subset of records, yielding $\kappa = .78$, indicating substantial agreement (Landis & Koch, 1977). All records judged eligible or uncertain during title and abstract screening were subsequently reviewed by three authors, and disagreements were resolved through discussion to finalize the list of included studies.

The first screening process eliminated 1,753 duplicate records, with 3,602 remaining records relevant for title review. After 2,387 records had been eliminated based on study titles,

1,178 records were excluded from the remaining 1,215 records through abstract review. The remaining 37 reports were retrieved for full-text assessment, and after excluding 20 reports, a total of 17 studies were included in this systematic review. Details of exclusions and reasons at each stage are presented in the PRISMA flow diagram (see Figure 1; Haddaway et al., 2022).

Figure 1

Flow diagram of the study selection process



Data Extraction

Data were extracted from the 17 included studies in line with our research questions, using a standardized extraction form developed a priori by the authors. The form was piloted on

a randomly selected subset of six included studies (approximately 35%) and refined before full extraction. The final form specified the categories and subcategories listed below, and all data were extracted and recorded in a Google Sheets spreadsheet structured accordingly. During this process, each study was assigned a unique identifier (S1–S17), used for cross-referencing throughout the manuscript and the accompanying tables. The blank template and the completed extraction sheet are provided on OSF (https://doi.org/10.17605/OSF.IO/JNH2B).

- Study features and sample information: educational setting and discipline, subject population, sampling institutions, final/total number of participants, and number of sampling institutions
- Conceptual and theoretical features: primary construct measured, theoretical background, (sub-)construct names and definitions, total number of (sub-)constructs, and measurement context
- Measurement features: study and measurement design, data collection method, number of questions developed/validated, psychometric evidence (validity and reliability), and full item list

Full-text review and data extraction were divided equally among the three authors and conducted independently, after which the first author reviewed all extracted data to ensure consistency and completeness. Information not reported in the original studies was coded as not reported, as the absence of certain psychometric indices was itself considered informative for addressing the research questions.

Data Synthesis

Data synthesis was conducted through narrative synthesis. For each included study, collaboration-related competencies were first examined at the theoretical construct level and then at the item level, specifically how relevant items were defined, formulated, and positioned within each scale's sub-constructs. Based on this analysis, studies were then mapped into three categories according to the degree to which human-AI collaboration competency was represented: explicitly operationalized, implicitly embedded, and not represented.

Methodological Quality Assessment

Given that the included studies primarily involved scale development and validation rather than experimental or clinical interventions, conventional risk-of-bias tools such as the Cochrane RoB 2.0 tool were not applicable. While the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) tool offers a framework for evaluating the measurement properties of scales, its full application was not feasible in the present review. COSMIN was originally developed for Patient-Reported Outcome Measures (PROMs) in clinical contexts, and a number of its criteria do not map straightforwardly onto educational measurement contexts. To address this, guided by the criteria outlined in COSMIN (Mokkink et al., 2010), methodological rigor was assessed by examining the reported psychometric properties of each scale, including content validity, structural validity, internal consistency, and reliability indices (see Table 3). This approach follows Lintner (2024), whose assessment of AI literacy

scales indicates that a full COSMIN appraisal would likely yield many incomplete ratings, given the early stage of measurement development in the field.

Results

The extracted data are analyzed in the following section and presented in corresponding tables. For information on basic study features and measurement validity, see Tables 2 and 3. For details on the constructs measured, refer to Table 4. The list of items referenced in the Results and Discussion sections is organized for each study in Appendix A (see Table A2). The list of coded studies (e.g., S1) can be found in the "List of Studies Included" under References.

Features of Included Studies

Participants

The 17 studies represented a total of 13,273 research participants included in the main analyses after screening and exclusion procedures. The sample sizes ranged from 158 to 2,713[1] participants. All studies reported the final number of research participants used in the analyses. Most studies recruited university students (n = 15), while two studies conducted in Hong Kong included both senior secondary school students (N = 341) and undergraduates. In addition, six studies (n = 6) reported including participants with graduate-level education or higher (N = 1,193). Gender distribution, ethnicity, age, and disciplinary breakdowns were not reported with sufficient consistency to support detailed cross-study comparison. Where age was reported, participants were predominantly undergraduates between approximately 18 and 25 years of age.

With respect to sampling sources, 10 studies reported multi-institutional samples, three were based on a single institution, three used online recruitment or online programs for which the number of sampling institutions was not specified, and one did not report its sampling institutions. These multi-institutional samples included studies conducted across multiple universities or institutions in Germany, China, Hong Kong, South Korea, and the Arab region, while the online samples included one cross-national sample from the United States and Canada.

Region

Research conducted in China had the greatest number of published studies included in this review (n = 6), followed by Hong Kong SAR (n = 2) and Germany (n = 2). Singapore, South Korea, Türkiye, Spain, Australia, the United States/Canada (cross-national), and the Arab region each contributed one study. Overall, the majority of studies were conducted in East Asia, particularly mainland China and Hong Kong, with comparatively fewer studies conducted in Europe, North America, or other regions.

Educational Setting and Discipline

Following our inclusion criteria, most studies were conducted in higher education settings (n = 14, 82.4%). Two studies (n = 2, 11.8%) included secondary education settings, and one

[1] In two studies (S9, S14), data were collected from overlapping groups of students at two time points; for these studies, the counts reflect the number of responses analyzed, so the number of unique participants is smaller. Pilot samples used only for item refinement or content validation were not included in these counts.

study focused on health professions education in an interprofessional education (IPE) context within higher education (*n* = 1, 5.9%).

Discipline was open-coded by identifying the primary academic or professional domain targeted by each study. The largest proportion of studies were conducted in general higher education without a single disciplinary focus (*n* = 10, 58.8%). Domain-specific disciplines were also represented, including medical and health professions education (*n* = 3, 17.6%), language education (*n* = 2, 11.8%), and library and information science (LIS) (*n* = 1, 5.9%). Meanwhile, one study focused on interdisciplinary competence itself, rather than a single discipline (*n* = 1, 5.9%).

Study and Measurement Design

All studies employed a quantitative research design. The majority were cross-sectional survey studies (*n* = 15, 88.2%), while two studies (11.8%) incorporated quasi-experimental or pre-post intervention designs alongside scale validation procedures. Regarding measurement format, most studies relied on self-report instruments (*n* = 14, 82.4%), typically using 5- or 6-point Likert scales. A smaller number of studies employed performance-based assessments (*n* = 3, 17.6%), such as multiple-choice knowledge tests, one of which (S7) also included self-assessment items on AI usage.

Data Collection

All studies used surveys as the primary data collection method (*n* = 17, 100%), with one study additionally conducting a focus group interview. Surveys were typically administered online using platforms such as Qualtrics, Wenjuanxing, Google Forms, or institutional systems.

Measurement Validity

Content validity procedures varied substantially across studies. Most studies reported some form of expert review during scale development (*n* = 15, 88.2%). Six studies (*n* = 6, 35.3%) reported conducting pilot studies, and twelve studies (*n* = 12, 70.6%) implemented content or face validity procedures with student participants. In all cases, expert review was described qualitatively without numerical indices such as CVI.

Construct validity was more consistently reported. Exploratory factor analysis (EFA) was conducted in 11 studies (*n* = 11, 64.7%), while confirmatory factor analysis (CFA) was reported in 14 studies (*n* = 14, 82.4%). Two studies (*n* = 2, 11.8%) used CFA specifically to test unidimensionality before fitting IRT models. Nine studies (*n* = 9, 52.9%) employed both EFA and CFA in independent samples.

Convergent and discriminant validity, typically assessed through Average Variance Extracted (AVE), composite reliability (CR), and Fornell–Larcker criteria, were reported in 9 studies (*n* = 9, 52.9%). These procedures were more common in recent SEM-based scale development research, while one study (S15) assessed convergent validity through correlations with a similar existing measure.

Reliability was the most consistently reported psychometric property. Cronbach's alpha was reported in 15 studies (n = 15, 88.2%). A smaller number of studies reported additional reliability indices (n = 2, 11.8%), such as McDonald's omega or test–retest reliability (including kappa or ICC), though these were relatively uncommon.

Table 2

Summary of study features

Study #	Author	Year	Region	Participants (*N*)	Sampling Institutions	Educational Setting & Discipline	Measurement & Study Design	Data Collection
S1	Jin et al.	2025	Australia	438	Online (Prolific)	Higher education (except Study3)	MCQ (performance)	Survey
S2	Laupichler et al.	2024	Germany	377	2	Medical education	Cross-sectional, self-report	Survey
S3	Hornberger et al.	2023	Germany	1286	6	Higher education	MCQ (performance)	Survey
S4	Liu et al.	2025	China	590	Several (unspecified)	English as a Foreign Language (EFL)	Cross-sectional, self-report	Survey
S5	Lee et al.	2026	South Korea	1000	Several (unspecified)	Higher education	Cross-sectional, self-report	Survey
S6	Kong et al.	2025	China	1034	1	Medical education	Cross-sectional, self-report	Survey
S7	Montesi et al.	2025	Spain	163 (100 students, 63 librarians)	1	Library and Information Science (LIS)	MCQ (performance), self-report	Survey, FGI (Focus group interview)
S8	Xu et al.	2025	China	785	3	L2 writing	Cross-sectional, self-report	Survey
S9	Hou et al.	2025	Singapore	2713 (1540 + 1173 responses)	1	Higher education	Cross-sectional, self-report	Survey

Study #	Author	Year	Region	Participants (*N*)	Sampling Institutions	Educational Setting & Discipline	Measurement & Study Design	Data Collection
S10	Gümüş & Kara	2025	Türkiye	773	Not reported	Higher education	Cross-sectional, self-report	Survey
S11	Wu et al.	2026	China	872	Several (unspecified)	Higher education	Cross-sectional, self-report	Survey
S12	Ma & Chen	2024	China	546	5	Higher education	Cross-sectional, self-report	Survey
S13	Kong & Yang	2025	Hong Kong	224	Several (unspecified)	Secondary, Higher education	Quasi-experiment, self-report	Survey
S14	Kong et al.	2025	Hong Kong	886 (477 + 409; 370 matched)	33	Secondary, Higher education	Quasi-experiment, self-report	Survey
S15	Wiss et al.	2025	USA / Canada	158	Online	Medical education (Health profession)	Cross-sectional, self-report	Survey
S16	Lu & Lin	2025	China	276	Online	Higher education	Cross-sectional, self-report	Survey
S17	Hazaimeh & Al-Ansi	2024	Arab	1152	Several (unspecified)	Higher education	Cross-sectional, self-report	Survey

Table 3

Measurement validity of studies included

Study #	Author	Content Validity (Expert panel review)	Content Validity (Face validity)	Construct Validity (EFA)	Construct Validity (CFA)	Convergent & Discriminant Validity	Reliability
S1	Jin et al.	✓ (not reported)	✓	-	✓ (IRT)	-	✓

Study #	Author	Content Validity (Expert panel review)	Content Validity (Face validity)	Construct Validity (EFA)	Construct Validity (CFA)	Convergent & Discriminant Validity	Reliability
S2	Laupichler et al.	✓ (not reported)	-	-	✓	-	✓
S3	Hornberger et al.	✓ (not reported)	✓ (not reported)	-	✓ (IRT)	-	✓
S4	Liu et al.	✓ (not reported)	✓ (not reported)	✓	✓	✓	✓
S5	Lee et al.	✓ (not reported)	✓ (not reported)	✓	✓	✓	✓
S6	Kong et al.	-	-	✓	✓	-	✓
S7	Montesi et al.	✓ (not reported)	✓ (not reported)	-	-	-	✓ (Kappa ICC)
S8	Xu et al.	✓ (not reported)	✓ (not reported)	✓	✓	-	✓
S9	Hou et al.	✓ (not reported)	✓ (not reported)	✓	✓	-	✓
S10	Gümüş & Kara	✓ (not reported)	✓ (not reported)	✓	✓	-	✓
S11	Wu et al.	✓ (not reported)	✓ (not reported)	✓	✓	✓	✓
S12	Ma & Chen	✓ (not reported)	✓ (not reported)	✓	✓	✓	-
S13	Kong & Yang	✓ (not reported)	✓ (not reported)	-	✓	✓	✓
S14	Kong et al.	✓ (not reported)	✓ (not reported)	✓	✓	✓	✓
S15	Wiss et al.	✓ (not reported)	-	✓	-	✓ (convergent only)	✓
S16	Lu & Lin	-	-	-	✓	✓	✓
S17	Hazaimeh & Al-Ansi	✓ (not reported)	-	✓	-	✓	✓

Note. Most reliability indices were reported as Cronbach's alpha.

Definitions, Conceptualizations, and Measurement of Constructs (RQ1)

To answer RQ1, this section provides an overview of how the main constructs of interest, AI literacy and human-AI collaboration competency, were defined, conceptualized, and measured across the included studies.

Definition of Constructs

Most studies drew on Long and Magerko's (2020) definition, which conceptualizes AI literacy as "a set of competencies that enables individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool online, at home, and in the workplace" (*n* = 9; S1, S2, S3, S6, S7, S9, S12, S14, S16). Although this widely cited definition includes keywords "communicate" and "collaborate," the common four sub-constructs proposed by Ng et al. (2021b) do not explicitly measure communication or collaboration competencies. A similar pattern appears in S5, which cites digital "communication and collaboration" competency as "the ability to interact, share information, and collaborate using digital tools (Punie & Ala-Mutka, 2013)," yet operationalizes it in terms of "understanding communication formats, managing online presence, engaging responsibly in online discussions, and recognizing the impact of digital interactions." While prior definitions often include "communication" or "collaboration," the corresponding items rarely measure human-AI collaboration competency or collaborative AI literacy explicitly, as discussed in the following section.

The next most frequently cited definition (*n* = 3; S3, S6, S8) was from Laupichler et al. (2023a, 2023b), who describe AI literacy as competencies including "basic knowledge and analytical evaluation of AI, as well as critical use of AI applications by non-experts." Beyond these recurring definitions, S10 adopted S. C. Kong et al.'s (2023) definition, under which AI literacy comprises "cognitive, affective, and sociocultural competencies, including understanding AI operating principles, problem-solving through ethical AI use, and critical thinking." In the case of GenAI literacy, cited definitions included the "capacity to recognize, apply, and assess GenAI-related products appropriately while maintaining ethical norms (Mondal, 2020; Wang et al., 2023a, 2023b)" in S4 and "the capability to navigate and scrutinize AI-generated content and use it effectively, emphasizing quality and accuracy (Li et al., 2023)" in S5. Lastly, S8, which examined AI literacy in an L2 writing context, cited a more domain-specific definition: "a newly emerged but critical aspect of literacy skills that is necessary for L2 writers to effectively navigate and integrate AI technologies into their writing tasks (Warschauer et al., 2023)." More detailed definitions of each (sub-)construct by study are provided in Appendix A (see Table A1).

Primary Construct Measured

Following the definitions outlined above, most studies (*n* = 11; S1, S2, S3, S4, S5, S6, S7, S8, S10, S12, S16) explicitly measured AI literacy, whereas none measured human-AI collaboration competency as a standalone construct. The study most closely aligned with our research question validated items measuring both constructs (*n* = 1; S11) and named the scale Interdisciplinary Competence in AI-Enabled Learning (ICAIL). The authors highlighted the need to conceptualize "a competence distinct from simply knowing about AI or how to use a specific application in isolation." Accordingly, ICAIL treats AI literacy and human-AI collaboration

competency as proxies for "the sophisticated ability to leverage AI within complex interdisciplinary problem-solving" contexts.

The remaining studies measured related constructs such as AI empowerment (n = 2; S13, S14), AI acceptance (n = 2; S15, S17), and AI reliance (n = 1; S9). AI empowerment is often framed as a key objective of AI literacy programs or interventions, aiming to "empower students to become active and key participants in the digital society (S13)." Although S13 and S14 differed in how they defined AI empowerment, both included impact, self-efficacy, and meaningfulness, with S13 further distinguishing creative self-efficacy in AI. Whereas AI empowerment aligns with the affective domain of AI literacy, AI acceptance is more closely tied to the behavioral domain, capturing attitudes toward AI applications and use (S15). Relatedly, AI reliance refers to "the behavioral patterns of when and how individuals depend on AI suggestions (S9)."

Measurement Context

Conceptualizations of AI literacy and collaboration may differ across contexts (e.g., domains, tasks, and collaboration settings). The measurement context should therefore be clearly specified. Although our inclusion criteria required validation in higher education, many studies used items that referred to AI use in general terms rather than to specific educational contexts. Consistent with widely used AI literacy frameworks (Ng et al., 2021b, 2024), many scales did not include context-specific terms such as "school," "coursework," or "group project" (n = 8; S2, S3, S6, S7, S12, S13, S16, S17). Even in studies situated in medical education (S2) or library and information science (S7), items were not adapted to the disciplinary context.

Two studies substantially tailored items to their target domains (n = 2; S8, S15), focusing on L2 writing (S8) and medical professions (S15). With the rise of GenAI tools, several studies published within the past two years focused on GenAI literacy (n = 4; S1, S4, S5, S10) by specifying use contexts such as creating digital content with GenAI. Notably, three studies (n = 3; S9, S11, S14) situated the measurement scale in problem-solving contexts, such as tasks or curricula in which respondents solve problems with AI. In these contexts, items take on a more collaborative nuance (e.g., "solve problems with AI" in S14), suggesting stronger alignment with human-AI collaboration competencies. However, one study (S13) still measured general AI empowerment despite embedding items in a problem-solving context.

Item Development

As stated in our inclusion and exclusion criteria (see Table 1), we excluded studies that only translated existing measures for validation in a specific language or cultural context. Accordingly, all measures included in this review were either newly developed or integrated existing scales. We labeled studies that combined existing scales as 'integrated,' whereas we labeled studies as 'newly developed' when they drew on multiple prior sources but (1) substantially adapted existing items and (2) added new items to fit specific measurement context (e.g., GenAI) or discipline (see Table 4). Among the six studies with newly developed scales (n = 6; S1, S3, S5, S9, S13, S14), most conducted a scoping review and/or consulted expert panels (see Table 3) during item development. Most studies drawing on prior frameworks (n = 9; S4, S6, S7, S8, S10, S11, S12, S15, S17) integrated more than three existing measures, whereas two

studies (n = 2; S2, S16) integrated two AI literacy scales. The integrated items were largely derived from early-2020s AI literacy scales and frameworks (2020–2023) and therefore closely reflect their definitions and conceptualizations of AI literacy (sub-)constructs.

Table 4

Details on constructs measured

Study #	Author	Primary Construct Measured	Measurement Context	Item Development	Name of instrument	Total number of constructs	Number of final items / Initial items	Availability of full items
S1	Jin et al.	AI literacy	GenAI	Newly developed	GLAT	4	20/25	Yes
S2	Laupichler et al.	AI literacy	General (Medical)	Integrated (2 scales)	-	5	33/36	No
S3	Hornberger et al.	AI literacy	General	Newly developed	-	1 (14 competencies)	30/57	Yes (Table 1 & Appendix)
S4	Liu et al.	AI literacy	GenAI (Digital multimodal composing)	Integrated (3+ scales)	-	4	17/55	Yes
S5	Lee et al.	AI literacy	GenAI (Digital content creation)	Newly developed	Generative AI Competence Scale (GenAIComp)	5	22/22 (63 statements)	Yes
S6	Kong et al.	AI literacy	General	Integrated (3+ scales)	AI literacy conceptual model (AICM)	4	26/29	No (partially available)
S7	Montesi et al.	AI literacy	General (Library and Information Science; LIS)	Integrated (3+ scales)	AILIS	5	35/39	Yes
S8	Xu et al.	AI literacy	Domain specific (L2 writing)	Integrated (3+ scales)	L2 Writing -Student AI Literacy Scale (L2W-SAILS)	4 (Figure 1)	22/35	No (partially available)

Study #	Author	Primary Construct Measured	Measurement Context	Item Development	Name of instrument	Total number of constructs	Number of final items / Initial items	Availability of full items
S9	Hou et al.	AI reliance	Problem-solving	Newly developed	-	4	16/23	Yes (Table 3)
S10	Gümüş & Kara	AI literacy	GenAI	Integrated (3+ scales)	GenAI-LLs	4	29/44	Yes
S11	Wu et al.	AI literacy, Human-AI collaboration competency	Problem-solving	Integrated (3+ scales)	Interdisciplinary Competence in AI-Enabled Learning (ICAIL)	5	16/73	Yes (Table 1)
S12	Ma & Chen	AI literacy	General	Integrated (3+ scales)	AILS-CCS	4	15/23	Yes (Table 1)
S13	Kong & Yang	AI empowerment	General (problem-solving)	Newly developed	-	4	16/17	Yes (Appendix)
S14	Kong et al.	AI empowerment	Problem-solving (PBL)	Newly developed	Empowerment in Using AI for Problem-Solving (EUAIPS)	3	11/13	Yes (Table 2)
S15	Wiss et al.	AI acceptance	Domain-specific (Collaborative practice in health professionals)	Integrated (3+ scales)	AI Acceptance Scale for Interprofessional Education (AAIPE) and Collaborative Practice Settings	1	8/9	Yes (Table 2)
S16	Lu & Lin	AI literacy	General	Integrated (2 scales)	-	6 (Table 1)	37/37	No (Table 3)
S17	Hazaimeh & Al-Ansi	AI acceptance	General	Integrated (3+ scales)	-	5	37/37	No

Representation of Human-AI Collaboration Competency (RQ2)

In connection with RQ2, this section identifies to what extent human-AI collaboration competency and related (sub-)constructs are explicitly or implicitly represented within existing AI literacy and related measures. The study features analyzed in the previous section and the full item lists of the included studies were examined to classify each scale into one of three categories: "explicitly operationalized," "implicitly embedded," and "not represented." Each scale was assigned to the category that best reflected its primary representation of collaboration; studies that also showed features of another category (S4, S7, S9) are additionally discussed in the corresponding section.

Explicitly Operationalized

Only three of the 17 studies (n = 3; S4, S8, S11) directly defined or measured human-AI collaboration competencies. Among them, S11 was the only study that measured both AI literacy and human-AI collaboration competency. The authors of S11 note that "interdisciplinary competence in the AI era is best understood as a dynamic interplay between human critical agency and AI affordances" and emphasize that "this competence extends beyond AI literacy by focusing on leveraging AI in collaborative, cross-disciplinary problem-solving contexts." While many items in S11 resemble conventional AI literacy measures, the five items under the "Collaborative Problem Solving (CPS)" sub-construct more directly reflect human-AI collaboration competency. Specifically, "CPS1: Coordinate roles in human-AI team workflows," "CPS3: Negotiate task ownership in hybrid collaboration," and "CPS5: Adapt communication for AI-mediated cross-domain teams" directly incorporate the terms "coordinate," "collaboration," and "communication" in a human-AI interaction context. Items such as "CPS2: Mediate conflicts using AI consensus-building tools," "CPS3: Negotiate task ownership in hybrid collaboration," and "CPS4: Evaluate AI contributions to team outcomes" view AI more as a partner than as a tool, as indicated by expressions such as "conflicts," "negotiate" in "hybrid" collaboration, and "contributions" to "team (outcomes)."

Study S4 also proposes sub-constructs that include the keywords "interaction," "collaboration," and "communication": social interaction and GAI anthropomorphic interaction under affective literacy, and collaboration and communication under behavioral literacy[2]. In particular, the authors conceptualize anthropomorphic interaction in a way that emphasizes collaboration with AI as similar to that with a human partner, arguing that perceiving AI as a human-like teammate can influence learners' attitudes toward GenAI and their patterns of use. For example, the item measuring GenAI anthropomorphic interaction is "I feel that the way of expression of GAI technologies/products in the interactive text is the same as that of real people when I compose my digital multimodal texts." It can be inferred that such an item has become more feasible and relevant with recent advances in GenAI, particularly LLM-based systems, which enable more naturalistic dialogue between human counterparts and AI.

[2] Of these, only GAI anthropomorphic interaction was retained in the final scale; collaboration and communication were removed following expert review.

Lastly, S8 included one item ("I can communicate effectively with AI tools based on the goals of my L2 writing") that incorporates the term "communicate." However, this competency was validated under the "Use" sub-construct, and because it was not defined or intended to measure human-AI collaboration competency, the item was deleted after the CFA and EFA analyses.

Implicitly Embedded

Consistent with the pattern observed in S8, a few other studies (n = 3; S7, S9, S14) included a small number of items that could be inferred to represent collaboration or related competencies, although not as a standalone sub-construct. For example, the item "In everyday life, I can work together gainfully with GAI in digital multimodal composing" in S4, placed under "Know & Understand GenAI" within "Cognitive literacy," was not positioned as a direct measure of human-AI collaboration competency yet introduces a collaborative context by using expressions such as "together." Similarly, items in S14 contain the term "with," such as "I want to solve problems with AI to make people's lives better." However, these items still tend to emphasize using AI as a tool rather than collaborating with AI as a partner.

Another item from the same study, S14, "Using AI to solve problems is useful to me," illustrates this tendency. Although this should be verified through comparison with measures of collaboration competencies in peer learning contexts is necessary, the term "useful" on its own may not be a reliable indicator of human-AI collaboration competency. Moreover, items within the "use AI" dimension of existing AI literacy frameworks are often phrased in the context of one-way interaction, making it difficult to interpret or perceive them as assessing collaborative competencies. At the same time, the definition of human-AI collaboration cited in S9, "where humans and AI systems work together by leveraging their unique strengths (Dellermann et al., 2019)," suggests that terms such as "together" or "with" can serve as textual cues that an item is intended to measure such competencies provided that it is situated within construct-level definitions or contexts that emphasize human-AI collaboration.

Items that include the term "interaction" can also be considered (S7, S9). However, a more precise definition of interaction is needed before interpreting such items as measures of human-AI collaboration competency, because "interaction" is a broad umbrella term that can, but does not necessarily, entail collaboration. Overall, most measures did not include collaboration-related terms at all, and even where such terms appear at the sub-construct or item level, their presence alone does not justify interpreting a measure as assessing collaboration competencies. Relatedly, several items in S5, S9, S10, and S15 did not assess collaboration with AI. Instead, they asked whether using AI tools or having knowledge about them could improve communication or collaboration among human peers. Although these studies were initially included based on keyword searches, full-text screening revealed that the items were framed around peer collaboration rather than human-AI collaboration. For example, S10 includes items such as "I can use generative AI to improve my dialogue with my teammates in group work" and "I can use generative AI to collaborate with my teammates to solve a problem." Items from S15 also included "I think AI can enhance communication between myself and colleagues from other professions" and "I think AI can enhance health sector teamwork and team-based care."

Not Represented

Consistent with our previous results, most studies (n = 11; S1, S2, S3, S5, S6, S10, S12, S13, S15, S16, S17) did not measure human-AI collaboration competencies either explicitly or implicitly. In particular, measures created by integrating existing AI literacy scales contained no such content (S6, S12, S16), likely because the source scales themselves rarely include collaboration-related competencies. In principle, scales that mainly measure AI empowerment could address human-AI collaboration competencies because they emphasize affective dimensions and therefore extend beyond knowledge and use (S13, S14). However, our item-level analysis indicated that S13 did not include content relevant to collaboration, and S14 did so only implicitly. A similar pattern emerged for AI acceptance measures. S15 included items on collaboration among human colleagues, as noted above, whereas S17 included no collaboration-related items.

Beyond self-report measures, S1 identifies performance-based measures as another way to assess AI literacy. Performance-based measures that primarily assess respondents' knowledge level on AI mechanisms or underlying principles (S1, S3, S7) necessarily fall into the "not represented" category. Greater knowledge about AI may ultimately support effective collaboration. However, as suggested in S4, knowledge categorized as a lower-order cognitive skill in Bloom's taxonomy, which has influenced many AI literacy frameworks, does not necessarily translate into higher-order competencies such as communication or collaboration (Anderson & Krathwohl, 2001; Lodge et al., 2023). Measuring how much a person knows about AI is not the same as measuring their ability to collaborate with AI. Moreover, if collaborative AI literacy is to be defined as a distinct competency, it should be conceptualized and measured separately from knowledge. In the following Discussion section, we further elaborate on why and how collaborative AI literacy should be assessed as a standalone competency.

Discussion

Prior AI literacy research has primarily focused on knowledge, understanding, and use of AI, largely framing AI as an instrument or tool. Our systematic review confirmed a similar pattern, where only one of the 17 included studies (S11) measured human-AI collaboration as a distinct construct. Although a small number of additional studies touched on collaboration and related competencies, they were usually limited to the sub-construct level or only indirectly suggested through the wording of specific items. These findings reflect the possibility that existing AI literacy frameworks do not yet fully represent the interactive and collaborative processes involved in human-AI collaboration.

In summary, our synthesis of prior measurement studies indicates two main issues. First, when human-AI collaboration items are nested within broader AI literacy frameworks, they are often removed during scale development. This pattern was observed in S4, where an expert panel considered collaboration and communication to be unnecessary components of behavioral literacy, and in S8, where the item "I can communicate effectively with AI tools based on the goals of my L2 writing" was judged to be a poor indicator and removed after factor analysis. These results suggest that human-AI collaboration competency may need to be measured as a construct distinct from general AI literacy.

Second, collaboration-related items are inconsistently placed under different sub-constructs across scales, highlighting the lack of a shared conceptual framework. For example, the item from S8 above was categorized under "use," whereas in S10, items relevant to collaborative learning were placed under "autonomous learning" in collaborative tasks. Taken together, the limited and inconsistent representation of human-AI collaboration competencies identified in this review reflects a broader misalignment between recent discussions of human-AI collaboration and existing AI literacy measurement frameworks. Thus, we conclude by delineating theoretical and measurement implications that should be addressed in reconceptualizing collaborative AI literacy.

Theoretical Implications (RQ3)

Viewing AI as a Collaborator or a Tool

Learners inevitably remain the main agents and users of AI. Yet items that position AI solely as a tool may not adequately capture the human-AI collaboration context and competencies required in present-day settings. At the other end of the spectrum, a small number of studies explicitly emphasize an anthropomorphic stance (S4) or more directly situate AI in a peer-like role (S10; "I can use generative AI as a classmate role"). Whether an anthropomorphic perspective is necessary, however, requires caution and further discussion. A focus group interview in S7 noted that "Even though AI is often humanized and perceived as almost human … students are aware that communicating with machines differs from human interactions in several ways."

At the same time, broadly classifying all AI literacy items traditionally framed around tool use as measures of collaborative AI literacy would blur the conceptual boundary. A key consideration, then, is to clearly distinguish whether an item assesses AI as a tool that facilitates human collaboration or assesses collaboration between humans and AI. For example, items that focus on reviewing AI outputs with peers (S9), items on using AI to facilitate communication between human counterparts (S15), or items on the knowledge of AI tools that support teamwork (S5) do not directly capture one's competency to collaborate with AI. Therefore, such context should be considered out of scope for measuring collaborative AI literacy.

Ultimately, when we aim to measure collaborative AI literacy in educational settings, the critical question is how equivalently we conceptualize collaboration with AI relative to collaboration with peers. While collaborative learning is common in higher education (Laal & Laal, 2012), previous literature on computer-assisted learning has largely treated AI as a tool rather than a collaborative partner (Zawacki-Richter et al., 2019). Conversely, as items become more similar to those used in peer learning and collaborative learning research, one could argue that they reflect a stronger form of collaboration. This interpretation is consistent with hybrid intelligence perspectives, which conceptualize effective performance as emerging from complementary contributions of humans and AI rather than from AI use alone (Dellermann et al., 2019; Lodge et al., 2023).

Factors That Influence Human-AI Collaboration

In this review, studies that primarily examined predictors of AI literacy, such as learner characteristics (e.g., gender, self-efficacy) and instructional factors (e.g., teacher intervention, curriculum), were beyond the scope of our research questions. Nonetheless, empirically identifying the predictors of human-AI collaboration competency remains an important direction for future research, particularly given the current absence of a well-defined conceptual framework or measure for this competency. Factors that may support effective human-AI collaboration can be inferred from studies included in this review. For example, S15 emphasizes that team-based learning and performance can be realized when members hold an underlying acceptance of AI tools. Using AI in ways that create synergy may be associated with higher AI empowerment. In S16, recognize AI, know AI, AI ethics, AI empowerment, and AI self-competence each positively predicted apply AI. Given that AI empowerment has been described as "psychological preparation for interaction between humans and artificial intelligence" in S16, it may also facilitate effective human-AI collaboration.

Likewise, S10, which placed items relevant to collaborative learning under the sub-construct of autonomous learning, highlights the importance of agency for effective human-AI collaboration. Examining a related construct from the opposite end, S9 argues that excessive AI reliance can undermine collaboration. From this perspective, some reverse-coded items in the ethical dimension of AI literacy might serve as indirect indicators of collaboration and related competencies. However, as discussed in our previous theoretical implication, items in S9 assessed using AI as a tool that facilitates peer collaboration, leaving room for debate about their relevance to human-AI collaboration competency.

Cognitive, Affective, and Behavioral Dimensions of Collaborative AI Literacy

Several foundational studies on AI literacy measurement, such as the ABCE framework (Ng et al., 2024), distinguish cognitive, behavioral, and affective dimensions of the construct. For example, sub-constructs "knowledge" and "use" are typically classified within the cognitive dimension, and empowerment is often classified as an affective dimension in existing scales (S13, S14). In this context, it is useful to examine how items representing collaborative AI literacy included in this review map onto these dimensions.

Across the reviewed scales, collaboration-related items were placed under different dimensions. S4, one of the studies that explicitly addressed human-AI collaboration competency, positioned collaboration and communication under the behavioral dimension, whereas S14 represented it only implicitly within an AI empowerment scale associated with the affective dimension. At the same time, prior computer-supported collaborative learning (CSCL) literature often classifies the formation of shared problem spaces and the co-construction of knowledge as cognitive components of collaboration (Roschelle & Teasley, 1995; Stahl et al., 2006). A related point appears in S9, which explains that human-AI collaboration, "where humans and AI systems work together by leveraging their unique strengths, can enhance human cognition," thereby emphasizing its cognitive aspect. In the reviewed measures, however, collaboration-related items under the cognitive dimension were framed as knowledge or use of AI (S4, S8) rather than as joint cognitive processes such as building shared understanding with AI. Thus, while all three dimensions must be taken into consideration when measuring

collaborative AI literacy, the cognitive dimension of human-AI collaboration appears to be less clearly articulated than the affective and behavioral dimensions. In conceptualizing collaborative AI literacy, it may also be necessary to consider how the cognitive, affective, and behavioral dimensions relate to one another within human-AI collaboration competency.

Measurement Implications (RQ3)

Without a well-established conceptual framework for collaborative AI literacy, newly developed scales are particularly susceptible to response bias (S1; Lintner, 2024; Ng et al., 2021b). Synthesizing extant literature and our findings, we identify coordination, communication, and metacognition as key factors of collaborative AI literacy. The following section outlines suggestions regarding content and terminology to guide the conceptualization of these factors and the development of corresponding items.

Coordination

In CSCL research, coordination is considered a central mechanism through which learners establish shared understanding and co-construct knowledge (Roschelle & Teasley, 1995; Stahl et al., 2006). The concept of coordination, through which the co-construction of meaning occurs (Long & Magerko, 2020), is sometimes referred to interchangeably as "interaction." Interaction is a broader umbrella term that can encompass coordination, and at the item level its meaning can be ambiguous. For example, S7 includes an item such as "I can interact effectively with AI in my training activities," which does not specify the type of activity or strength of interaction being assessed. As Hornbæk and Oulasvirta (2017) illustrate, interaction with computers (i.e., AI) can take multiple forms, including dialogue, transmission, tool use, and experience. Therefore, when developing measures of collaborative AI literacy, it would be useful to organize various coordination contexts and specify which form of interaction each scale targets. Doing so would help ensure that respondents clearly understand the intended measurement context and reduce bias.

Based on our findings, problem-solving tasks and contexts using GenAI appeared more likely to facilitate co-construction and thus, more suitable for assessing collaborative AI literacy. Collaboration patterns are likely to vary by discipline and must also be taken into consideration, yet much of the reviewed literature did not significantly adapt prior scales or items to fit its domain context. Overall, domain-specific problem-solving contexts may be more likely to elicit collaboration processes, which could make collaborative AI literacy easier to specify and assess.

At the same time, whether an item conveys a collaborative context is likely to depend not only on its content but also on its terminology. For example, although S12 ("I can use artificial intelligence applications or products to help me solve problems in daily life") and S14 ("Using AI to solve problems is useful to me") both refer to problem-solving, terms such as "use" and "useful" can reinforce an instrumental view of AI. Similarly, in S12, items such as "I understand how artificial intelligence technology can aid my learning and daily life" and "I can use artificial intelligence applications or products to assist my learning" emphasize one-way interaction through verbs like "aid" and "assist," making it unclear whether respondents would interpret the context as collaborative.

By contrast, although items that include "with" may not explicitly measure human-AI collaboration competency, they position AI more as a partner in problem-solving. For instance, S14 includes items such as "I want to solve problems with AI to make people's lives better" and "I want to solve problems with AI to make daily life easier." These examples point to the need for careful attention to how items are phrased, as well as to whether and how terms such as "with," "use," and "useful" are used in established measures of (peer) collaborative learning competencies.

That said, because semantic nuance can be difficult to quantify and may vary across translations, it may be more reliable to focus on content that clearly signals coordination or two-way interaction. Measures should make the coordination context salient to respondents rather than implying one-way tool use. For example, as hinted in prior collaborative learning research (Brundage et al., 2023), an item such as "identify errors made by AI" from Sidra and Mason (2025) could be revised to emphasize reciprocal exchange by assessing whether AI also helps identify errors in the learner's work.

Communication

As noted in the focus group interview reported in S7, "Students repeatedly emphasize different aspects of communication with AI, and the theme of human–machine interaction is central in all their accounts." Meanwhile, in S7, "communication" is operationalized as the ability to construct clearer prompts, which reflects a form of one-way interaction rather than reciprocal dialogue. A similar ambiguity appears in S8, where an item "I can communicate effectively with AI tools based on the goals of my L2 writing" uses the term "communicate" without clarifying what communication entails in practice. Although these items use the term "communication," they can at best measure collaborative AI literacy only implicitly.

Prior to the emergence of LLM-based systems, it was difficult to conceptualize "communication" with AI as naturalistic dialogue. As conversational AI systems increasingly support dialogue-based interaction, communication competencies may become a critical component of effective human-AI collaboration. In particular, if human-AI interaction is conceptualized as dialogue (Hornbæk & Oulasvirta, 2017), communication should be validated as a distinct factor of collaborative AI literacy.

Metacognition

Metacognition is conceptualized as a multidimensional construct comprising knowledge of cognition (e.g., declarative, procedural, and conditional knowledge) and regulation of cognition (e.g., planning, monitoring, and evaluating) (Schraw & Dennison, 1994). While prior literature has cautioned against metacognitive laziness, highlighting risks such as automation bias and reduced critical evaluation when interacting with AI systems (Fan et al., 2025), the studies we reviewed similarly emphasized learners' self-regulation or agency in using AI. For example, S9, which addresses AI reliance, suggests that learners need the competency to critically evaluate AI outputs before integrating them into their cognitive structures. As collaborative partners of AI, learners should remain active counterparts in their interaction with AI, making agency an influential factor as in other learning contexts. Accordingly, key

metacognitive processes associated with self-regulation and agency should be operationalized and validated as part of collaborative AI literacy measurement.

Practical and Policy Implications

Governments and educational institutions continue to develop frameworks and guidelines for AI integration in education, such as UNESCO's AI Competency Framework for Students (Miao & Shiohira, 2024). The findings of this review emphasize that educational efforts should move beyond teaching students how AI systems work or how to use AI tools effectively. Literacy for human-AI collaboration should also be regarded as an essential competency that requires systematic training. Recognizing collaborative AI literacy as a distinct competency may help educational institutions better prepare learners to perform effectively in settings that increasingly depend on productive collaboration between humans and AI. This review provides conceptual and measurement foundations to design educational programs and assessments that explicitly measure collaborative AI literacy and support its development, with a particular focus on coordination, communication, and metacognition during human-AI interaction.

Limitations

This review has the following limitations that should be noted. First, the study focused exclusively on newly developed or significantly adapted scales within higher education settings published since 2020, in accordance with our inclusion criteria. That is, widely cited instruments or those validated solely in K–12 or professional contexts were excluded, potentially limiting the comprehensiveness of the AI literacy measurement tools analyzed. Second, the review included a relatively small number of eligible studies (n = 17), and methodological quality was assessed using reported psychometric evidence (i.e., content validity, construct validity, convergent and discriminant validity, and reliability; see Table 3) rather than a full COSMIN appraisal. Third, although multiple databases were systematically searched, the review was limited to English-language publications and did not include gray literature; these decisions may have resulted in the omission of some relevant studies. Finally, the conceptualization of collaborative AI literacy proposed in this review remains theoretical and has not yet been empirically validated through scale development or construct validation studies.

Recommendations for Future Research

Given the scarcity of prior literature, future research could advance the conceptualization of collaborative AI literacy by drawing comparisons with established frameworks in computer-supported collaborative learning (CSCL) and peer collaboration. While current items often borrow surface-level terminology from these fields, future studies should critically examine how collaboration-related competencies converge and diverge across human-AI and human-human contexts. Systematic comparisons between proposed human-AI collaboration items and validated peer collaboration measures may help identify interaction patterns or features that are especially relevant to human-AI settings. Such comparisons may clarify which components of collaboration can be meaningfully transferred to human-AI contexts and which require reconceptualization, thereby guiding the development of measures that go beyond direct adaptations of general collaboration scales.

Conclusion

This systematic review synthesized 17 studies focused on the development and validation of measures for AI literacy and human-AI collaboration competency within higher education. Across the reviewed instruments, human-AI collaboration competency was rarely measured within AI literacy scales predominantly framed around knowledge acquisition and functional tool use. Even in the few studies that reflected collaborative competencies, they were typically represented implicitly at the item level within broader AI-related sub-constructs rather than measured as a distinct construct. Overall, existing measures largely remain within an instrumental view of AI and do not sufficiently reflect the coordination and communication competencies underlying human-AI collaboration.

Building on these findings, this review clarifies the conceptual boundaries between AI literacy and human-AI collaboration competency, providing an initial conceptual basis for collaborative AI literacy. Specifically, we conceptualize collaborative AI literacy in terms of coordination, communication, and metacognition and outline implications for specifying measurement contexts and developing corresponding items. By bringing together extant literature and the studies reviewed, we lay groundwork for future research that can more directly measure the competencies necessary for effective human-AI collaboration.

Declarations

Ethics Statement

Ethics approval was not required for this type of research (systematic review of published literature), which did not involve the collection of primary data from human participants.

Funding Statement

This research received no external funding.

AI Use Statement

Generative artificial intelligence (AI) was used in the preparation of this manuscript for grammar and style refinement, clarity, and reference formatting, using ChatGPT (GPT-5.2 Thinking), Gemini (Gemini 3 Flash), and Claude (Opus 4.8 and Opus 5.5). The AI tools were not used to generate, analyze, or interpret research data, nor to create original scholarly content. All content was reviewed, verified, and approved by the authors, who take full responsibility for the accuracy and integrity of the work.

Conflicts of Interest Statement

The authors declare no conflicts of interest related to the research, authorship, or publication of this article.

Data Availability

The data extraction template and the completed extraction data for this review are publicly available on the Open Science Framework (OSF): https://doi.org/10.17605/OSF.IO/JNH2B.

References

List of Studies Included

[S1] Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, *9*, 100436.

[S2] Laupichler, M. C., Aster, A., Meyerheim, M., Raupach, T., & Mergen, M. (2024). Medical students' AI literacy and attitudes towards AI: a cross-sectional two-center study using pre-validated assessment instruments. *BMC Medical Education*, *24*(1), 401.

[S3] Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, *5*, 100165.

[S4] Liu, M., Zhang, L. J., & Zhang, D. (2025). Enhancing student GAI literacy in digital multimodal composing through development and validation of a scale. *Computers in Human Behavior*, *166*, 108569.

[S5] Lee, S. C., Baby, T., Vongvit, R., Lee, J., Kim, Y. W., Cha, M. C., & Yoon, S. H. (2026). Development and validation of Generative AI Competence Scale (GenAIComp) among university students. *Technology in Society*, *84*, 103059. https://doi.org/10.1016/j.techsoc.2025.103059

[S6] Kong, J., Liu, J., Chen, G., & Shang, W. (2025). Assessing AI literacy in college students: the mediating role of self-efficacy in motivational commitment pathways. *Education and Information Technologies*, *30*(16), 23957-23979.

[S7] Montesi, M., Bornstein, B. Á., Puig, N. B., Ochando, M. B., & Díez, A. S. (2025). AILIS 1.0: A new framework to measure AI literacy in library and information science (LIS). *The Journal of Academic Librarianship*, *51*(5), 103118.

[S8] Xu, L., Zhang, L., Ou, L., & Wang, D. (2025). The development and validation of a scale on student AI literacy in L2 writing: A domain-specific perspective. *Journal of Second Language Writing*, *69*, 101227.

[S9] Hou, C., Zhu, G., Sudarshan, V., Lim, F. S., & Ong, Y. S. (2025). Measuring undergraduate students' reliance on Generative AI during problem-solving: Scale development and validation. *Computers & Education*, *234*, 105329.

[S10] Gümüş, M. M., & Kara, M. (2025). Development and validation of the Generative AI Literacy for Learning Scale (GenAI-LLs). *Australasian Journal of Educational Technology*, *41*(4), 1-16.

[S11] Wu, Y., Lu, X., & Lin, C. (2026). Interdisciplinary competence in AI-enabled learning. *Educational Technology & Society, 29*(1), 133-149.

[S12] Ma, S., & Chen, Z. (2024). The development and validation of the artificial intelligence literacy scale for Chinese college students (AILS-CCS). *IEEE Access*, *12*, 146419-146429.

[S13] Kong, S. C., & Yang, Y. (2025). Developing and validating an artificial intelligent empowerment instrument: evaluating the impact of an artificial intelligent literacy programme for secondary school and university students. *Research and Practice in Technology Enhanced Learning*, *20*, 024. https://doi.org/10.58459/rptel.2025.20024

[S14] Kong, S. C., Zhu, J., & Yang, Y. N. (2025). Developing and validating a scale of empowerment in using artificial intelligence for problem-solving for senior secondary and university students. *Computers and Education: Artificial Intelligence*, *8*, 100359.

[S15] Wiss, A., Joosten-Hagye, D., Pattershall-Geide, J., Showstark, M., Zschaebitz, E., Potter, K., Embry, E., Hageman, H., & Brooks, P. (2025). Development of the AI Acceptance Scale for Interprofessional Education (AAIPE) and collaborative practice settings. *Journal of Interprofessional Education & Practice, 40*, 100752.

[S16] Lu, W., & Lin, C. (2025). How do artificial intelligence literacy constructs work—based on a survey of university non-expert students. *Education and Information Technologies*, *30*(10), 13779-13805.

[S17] Hazaimeh, M., & Al-Ansi, A. M. (2024). Model of AI acceptance in higher education: arguing teaching staff and students perspectives. *The International Journal of Information and Learning Technology*, *41*(4), 371-393.

References

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open, 6*, 100173. https://doi.org/10.1016/j.caeo.2024.100173

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition.* Addison Wesley Longman, Inc.

Artman, H., & Garbis, C. (1998). Situation awareness as distributed cognition. In *Proceedings of the 9th Conference on Cognitive Ergonomics* (pp. 151–156). European Association of Cognitive Ergonomics.

Brundage, M. J., Malespina, A., & Singh, C. (2023). Peer interaction facilitates co-construction of knowledge in quantum mechanics. *Physical Review Physics Education Research, 19*(2), 020133.

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans, 1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

Chiu, T. K. F. (2025). AI literacy and competency: definitions, frameworks, development and future research directions. *Interactive Learning Environments, 33*(5), 3225-3229.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering, 61*, 637–643. https://doi.org/10.1007/s12599-019-00595-2

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Elsevier.

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gasevic, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews, 18*, e1230. https://doi.org/10.1002/cl2.1230

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*(2), 174–196. https://doi.org/10.1145/353485.353487

Hornbæk, K., & Oulasvirta, A. (2017, May 2). What Is Interaction? *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* CHI '17: CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA. https://doi.org/10.1145/3025453.3025765

Kim, M. M., Shanaa, A. M., Hammady, H., Aboelnour, M., & Ayan, R. (2026). Enhancing transparency, auditability, and reproducibility of deduplication in systematic reviews: Tutorial for the Rayyan method and Systematic Auto Resolver. *Interactive Journal of Medical Research, 15*(1), e94317. https://doi.org/10.2196/94317

Kong, S. C., Cheung, M. Y. W., & Tsang, O. (2024). Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence, 6,* 100214.

Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2023). Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society, 26*(1), 16-30.

Laal, M., & Laal, M. (2012). Collaborative learning: what is it? *Procedia-Social and Behavioral Sciences, 31*, 491-495.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence, 3,* 100101.

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023a). Development of the scale for the assessment of non-experts' AI literacy – An exploratory factor analysis. *Computers in Human Behavior Reports, 12,* 100338. https://doi.org/10.1016/j.chbr.2023.100338

Laupichler, M. C., Aster, A., & Raupach, T. (2023b). Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. *Computers and Education Artificial Intelligence, 4*, 100126. https://doi.org/10.1016/j.caeai.2023.100126

Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning, 9*(1), 50. https://doi.org/10.1038/s41539-024-00264-4

Lodge, J. M., Howard, S. K., Bearman, M., & Dawson, P. (2023). Assessment reform for the age of artificial intelligence. *Journal of University Teaching & Learning Practice, 20*(2).

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Miao, F., & Shiohira, K. (2024). *AI competency framework for students*. UNESCO Publishing.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research, 19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021b). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2,* 100031.

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082-1104.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinnrdd, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *71*. https://doi.org/10.1136/bmj.n71

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). Springer.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460-475.

Sidra, S., & Mason, C. (2025). Generative AI in human-AI collaboration: Validation of the Collaborative AI Literacy and Collaborative AI Metacognition scales for effective use. *International Journal of Human-Computer Interaction*. Advance online publication. https://doi.org/10.1080/10447318.2025.2543997

Siemens, G., Marmolejo-Ramos, F., Gabriel, F., Medeiros, K., Marrone, R., Joksimovic, S., & de Laat, M. (2022). Human and artificial cognition: Complementary capabilities and coordination in hybrid human-AI systems. *Computers & Education: Artificial Intelligence, 3*, Article 100107.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge University Press.

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–24).

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*(12), 2293-2303.

Xia, L., Shen, K., Sun, H., An, X., & Dong, Y. (2025). Developing and validating the student learning agency scale in generative artificial intelligence (AI)-supported contexts. *Education and Information Technologies, 30*(10), 13999–14021. https://doi.org/10.1007/s10639-024-13137-5

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39.

Zhang, S., Ganapathy Prasad, P., & Schroeder, N. L. (2025). Learning about AI: A systematic review of reviews on AI literacy. *Journal of Educational Computing Research, 63*(6). https://doi.org/10.1177/07356331251342081

Appendix A. Extracted Definitions and Items by Study

Table A1.

Definition of sub-constructs measured

Study #	Author	Definition of sub-constructs
S1	Jin et al.	Adopts Ng et al.(2021b, 2024) framework and definitions
S2	Laupichler et al.	Adopts "Scale for the assessment of non-experts' AI literacy" (SNAIL) by Laupichler et al. and "Attitudes towards Artificial Intelligence" scale (ATAI scale) by Sindermann et al.
S3	Hornberger et al.	Adopts Long & Magerko's (2020) framework and definitions
S4	Liu et al.	**Affective literacy** comprises two factors that measure students' subjectively experienced feelings GenAI-assisted DMC(digital multimodal composing) in terms of (1) intrinsic motivation, and (2) self-efficacy. **Behavioural literacy** intends to assess behavioral commitment, which consists of two factors: (1) Students' behavioral intention and (2) Students' behavioral engagement of GenAI. **Cognitive literacy** comprises three factors: students' knowledge and skills achievement from (1) lower- (know and understand GenAI), (2) mid- (use and apply GenAI) to (3) high-order thinking skills (evaluate GenAI). **Ethical literacy** comprises a set of three ethical aspects: (1) discrimination and exclusion, (2) misuse, and (3) misinformation and disinformation.
S5	Lee et al.	**Information and Data literacy** emphasizes the ability to locate, retrieve, evaluate, and manage digital information. This involves assessing the credibility and reliability of data sources, distinguishing between information and misinformation, and understanding how data are generated, stored, and used in the digital environment. **Communication and Collaboration** includes the ability to interact, share information, and collaborate using digital tools. This comprises understanding different digital communication formats, managing online presence, engaging in online discussions responsibly, and recognizing the impact of digital interactions. **Digital Content Creation** includes the creation, modification, and publication of digital content while ensuring

Study #	Author	Definition of sub-constructs
		compliance with copyright laws and intellectual property regulations. It comprises skills such as editing digital media, developing creative digital assets, and understanding the ethical and legal implications of AI-generated content. **Safety competence** aims to protect personal data, ensure cybersecurity, and promote responsible digital technology use. This involves understanding cybersecurity threats such as phishing, malware, and identity theft, as well as the ethical implications of data collection and processing. **Problem Solving competence** involves identifying, troubleshooting, and resolving digital challenges. This encompasses applying digital tools for innovation, critically analyzing technological problems, and adapting to evolving digital environments.
S6	Kong et al.	Not specified.
S7	Montesi et al.	Not specified (What is AI / How AI works / Usage of AI / Ethics of AI / Evaluation of AI).
S8	Xu et al.	**'Understanding'** encompasses students' knowledge of and attitudes towards AI tools. Specifically, it involves how students understand the basic functions of AI tools (e.g., knowledge about communication with AI tools based on specific writing needs); It also includes how students understand and perceive their roles in L2 writing (e.g., attitudes towards AI tools' effectiveness in improving different aspects of writing, such as linguistic accuracy, structure, appropriateness, as well as writing efficiency). '**Use**' refers to how students apply AI tools throughout the L2 writing process (e.g., use AI tools to conduct pre-writing planning and to check and correct grammatic mistakes) **'Evaluation'** focuses on how students analyse, select, and critically evaluate AI tools and their outcomes (e.g., compare, analyse and make informed choices regarding AI-generated suggestions). '**AI ethics**', centres on students' ethical and safety considerations when applying AI tools in L2 writing. This may include, for instance, students' awareness of potential risks associated with using AI tools in L2 writing (e.g., privacy and information security), or students' capabilities of distinguishing between ethical and unethical behaviours.
S9	Hou et al.	**Reflective use** includes behaviors where students critically engage with Generative AI. They think about how to frame their questions to get the desired output, critically read and revise AI-generated outputs, and use AI to enhance their initial drafts or solutions.

Study #	Author	Definition of sub-constructs
		Cautious use captures behaviors where students recognize and react to errors produced by Generative AI. They spot inaccuracies, acknowledge the imperfection of AI outputs, and sometimes discard AI suggestions. **Thoughtless use** describes behaviors where students rely heavily on AI without much critical engagement. They might copy and paste task descriptions to Generative AI or AI-generated output directly into their solutions, indicating a passive reliance on AI. **Collaborative use** involves behaviors related to peer communication and collaboration in conjunction with Generative AI use, such as students collaboratively spotting errors, modifying AI output, and supporting each other during problem-solving activities. This collaborative approach highlights the social dimension of learning with AI.
S10	Gümüş & Kara	**Needs analysis** dimension covered in this study refers to how learners select and use the GenAI tools to meet a need or solve a problem, and how they analyse the content generated by these tools. In the same vein, several studies have partially addressed needs analysis by understanding and problem-solving with AI (Kim & Lee, 2022), use (Laupichler et al., 2023; Ma & Chen, 2024) or use and evaluation (Çelebi et al., 2023) and the dimension relevant to information (Biagini et al., 2024). **Prompt and language skills** dimension covered in this study refers to the learners' interactions with GenAI through prompts and their language knowledge and skills during their learning experience. **Autonomous learning** in this study refers to learners' self-regulation skills for individual and collaborative learning. Recent scales on learner agency with GenAI consistently encompassed learner autonomy and collaboration with peers, such as "goal setting", "self-adjustment", "self-reflection" or "participative action" (Xia et al., 2025) and "reflective use" and "collaborative use" in the GenAI reliance scale (Hou et al., 2025). **Critical thinking** refers to learners' skills in questioning and evaluating the content generated by GenAI during their learning experiences, such as inaccuracy and bias, which are required to deal with the potential risks of learning with GenAI (Noroozi et al., 2024).
S11	Wu et al.	**Knowledge Connectivity (KC)**: Utilizing AI tools (e.g., semantic analysis, recommendation systems) to effectively and strategically synthesize fragmented knowledge across diverse domains. **Critical Interdisciplinary Analysis (CIA):** Critically appraising AI-generated information and solutions within interdisciplinary contexts, focusing on bias, ethics, and validity alongside disciplinary perspectives.

Study #	Author	Definition of sub-constructs
		This dimension aligns with the need for vigilant and discerning judgment in human-AI Collaboration. **AI-Driven Innovation (ADI):** Leveraging AI (e.g., generative models, simulations) for iterative prototyping, hypothesis testing, and co-creating novel, integrative solutions across disciplinary boundaries. **Collaborative Problem Solving (CPS):** Effectively coordinating roles, managing tasks, and communicating within hybrid human-AI teams engaged in interdisciplinary work, fostering synergy between human insight and AI capabilities. **Adaptive Transfer (AT):** Metacognitively modifying and transferring AI tool usage and strategies across different interdisciplinary problems and evolving contexts. This dimension is directly informed by Dynamic Capability Theory, highlighting the learner's agility in a rapidly changing technological landscape.
S12	Ma & Chen	**Awareness** refers to the understanding of AI concepts and basic principles. **Usage** denotes the ability to use AI in daily life and learning. **Evaluation** signifies the ability to critically assess AI tools and AI-generated content. **Ethics** involves the ethical and safe use of AI.
S13	Kong & Yang	**Impact** refers to students' perception of the impact of AI and the importance of AI literacy. **AI self-efficacy** refers to students' belief that they have the necessary skills and abilities to use AI. **Creative self-efficacy in AI** refers to students' belief in using AI to produce novel ideas and solutions to real-world problems. **Meaningfulness** refers to students' perceptions of the value of AI.
S14	Kong et al.	**Impact** in using AI for problem-solving refers to students' perception of the extent to which one's behaviour (e.g., using AI) makes a difference to the outcomes of accomplishing tasks (e.g., problem-solving), that is, yielding intended strategic,

Study #	Author	Definition of sub-constructs
		administrative, or operating effects of AI-task environments. **Self-efficacy** in using AI for problem-solving refers to one's belief in his or her competence to use AI for problem-solving, which is analogous to personal mastery. **Meaningfulness** in using AI for problem-solving refers to students' perceived value of using AI for problem-solving based on their beliefs, ideas, and standards.
S15	Wiss et al.	Not specified.
S16	Lu & Lin	**Recognize AI,** also referred to as detect AI, which is seen as a metacognitive ability that refers to the user's cognition of the process of using artificial intelligence tools. **Know AI** includes the underlying logic of artificial intelligence algorithms such as supervised learning (linear regression and K-nearest neighbor algorithms), unsupervised learning (K-means clustering), and neural networks, as well as how to apply these algorithms to solve real-life problems. **AI ethics** refers to the ethical and responsible use of artificial intelligence technology. **AI empowerment** is the psychological preparation for interaction between humans and artificial intelligence. **AI self-competence** is based on the assumption of human-computer interaction, humans always need to dominate to avoid the risk of being controlled by machines in their thoughts, emotions, and behaviors. **Apply AI** (not specified)
S17	Hazaimeh & Al-Ansi	**Adopted attitude** refers to the way a person thinks, feels and behaves toward something or a situation, especially when this attitude is influenced by external factors. The term "adopt an attitude" implies that the individual or group has chosen or accepted a particular stance or approach, often in response to external pressures or influences. **Digital competency** refers to the ability to use digital technologies effectively and efficiently for various purposes such as communication, problem-solving and information management.

Study #	Author	Definition of sub-constructs
		Openness to experience, or simply openness, is a basic personality trait that denotes receptivity to new ideas and new experiences. **Perceived benefits** refer to an individual's beliefs about the positive outcomes or value associated with a specific action or behavior. **AI acceptance** refers to the behavioral intention or willingness of individuals to use, buy or try AI systems or services.

Note. Letters in bold refer to the names of sub-constructs.

Table A2.

List of items referenced in the manuscript

Study #	Author	Items referenced in the Results and Discussion section
S1	Jin et al.	-
S2	Laupichler et al.	-
S3	Hornberger et al.	-
S4	Liu et al.	[Proposed sub-factors] **- Affective Literacy: Social Interaction (not retained)** **- Affective Literacy: GAI anthropomorphic interaction (retained)** **- Behavioral literacy: Collaboration, communication (removed during expert review)** [Specific items] **AL5 (Affective literacy: Self-efficacy-GAI anthropomorphic interaction)** I feel that the way of expression of GAI technologies/products in the interactive text is the same as that of real people when I compose my digital multimodal texts. **CL1 (Cognitive literacy: Know & Understand GAI)** In everyday life, I can work together gainfully with GAI in digital multimodal composing.

Study #	Author	Items referenced in the Results and Discussion section
S5	Lee et al.	**2. Communication and Collaboration** **2.1: Interacting through Generative AI technologies** **2.3: Engaging in citizenship through Generative AI** Q17. I can identify Generative AI services that support social engagement and collaboration. **2.4: Collaborating through Generative AI** Q19. I can choose Generative AI tools that facilitate collaboration in group tasks. Q20. I can select Generative AI technologies that support teamwork and shared goals. Q21. I can identify Generative AI tools that improve efficiency in collaborative processes. **2.5: Netiquette** Q23. I can choose communication modes and strategies adapted to an audience when using Generative AI tools. **3. Digital Content Creation** Q33. I can integrate AI-generated content and ideas to create new and original work.
S6	Kong et al.	-
S7	Montesi et al.	**Interacting with AI to simplify learning tasks.** 19. I can interact effectively with AI in my training activities.
S8	Xu et al.	**Use** I can communicate effectively with AI tools based on the goals of my L2 writing.
S9	Hou et al.	**Thoughtless use** R3: For each activity I did in this class, I spent most of the time interacting with Generative AI. **Collaborative use** R18: My peer(s) helped me spot error in the Generative AI's output. R19: My peer(s) helped me modify Generative AI's output.
S10	Gümüş & Kara	**Autonomous learning** 17. I can use generative AI as a classmate role. 18. I can use generative AI to improve my dialogue with my teammates in group work.

Study #	Author	Items referenced in the Results and Discussion section
		19. I can use generative AI to collaborate with my teammates to solve a problem. 20. I can use generative AI to facilitate group activities to solve a problem.
S11	Wu et al.	**Collaborative Problem Solving (CPS)** **CPS1**. Coordinate roles in human-AI team workflows **CPS2.** Mediate conflicts using AI consensus-building tools **CPS3.** Negotiate task ownership in hybrid collaboration **CPS4.** Evaluate AI contributions to team outcomes **CPS5.** Adapt communication for AI-mediated cross-domain teams
S12	Ma & Chen	**Awareness** "I understand how artificial intelligence technology can aid my learning and daily life." **Usage** "I can use artificial intelligence applications or products to help me solve problems in daily life." "I can use artificial intelligence applications or products to assist my learning."
S13	Kong & Yang	-
S14	Kong et al.	**Impact** 2. I want to solve problems with AI to make people's lives better. 3. I want to solve problems with AI to make daily life easier. **Meaningfulness** 16. Using AI to solve problems is useful to me.
S15	Wiss et al.	6. I think AI can enhance communication between myself and colleagues from other professions. 7. I think AI can enhance health sector teamwork and team-based care.
S16	Lu & Lin	-
S17	Hazaimeh & Al-Ansi	-

Note. Letters in bold refer to the names of sub-constructs.

Appendix B. PRISMA Checklist

PRISMA 2020 Main Checklist

Topic	No.	Item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	p. 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist	p. 1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	pp. 1–4
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	pp. 4–5
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	pp. 5–6 (Table 1)
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	p. 5; Appendix C
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	p. 5; Appendix C
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	pp. 5–6
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	pp. 7–8
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	pp. 7–8

Topic	No.	Item	Location where item is reported
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	pp. 7–8
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	p. 8 (Table 3)
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item 5)).	p. 8
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	p. 8
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	N/A
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	p. 8
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	pp. 6–7 (Figure 1)

Topic	No.	Item	Location where item is reported
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	p. 7 (Figure 1)
Study characteristics	17	Cite each included study and present its characteristics.	pp. 9–12 (Table 2)
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	pp. 10, 13 (Table 3)
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	pp. 10, 13 (Table 3)
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	pp. 20–24
	23b	Discuss any limitations of the evidence included in the review.	pp. 24–25
	23c	Discuss any limitations of the review processes used.	pp. 24–25
	23d	Discuss implications of the results for practice, policy, and future research.	pp. 24–25
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	This review was neither registered nor

Topic	No.	Item	Location where item is reported
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	based on a published protocol.
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	p. 26
Competing interests	26	Declare any competing interests of review authors.	p. 26
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	p. 26

PRISMA Abstract Checklist

Topic	No.	Item	Reported?
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Yes
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Yes
Synthesis of results	6	Specify the methods used to present and synthesize results.	Yes
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes

Topic	No.	Item	Reported?
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Yes
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	N/A
Registration	12	Provide the register name and registration number.	No

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. MetaArXiv. 2020, September 14. https://doi.org/10.31222/osf.io/v7gm2. For more information, visit: www.prisma-statement.org

Appendix C. Database Search Queries

All database searches were conducted in January 2026. The search strategy combined four conceptual blocks: AI-related terms, literacy and collaboration-related terms, education/learning and student-related terms, and measurement-related terms. Terms within each block were combined with OR, and the four blocks were combined with AND. The same Boolean structure was adapted to the syntax and search fields of each database.

Table C1.

Search Queries

Database	Search Queries
Web of Science	TS=("artificial intelligence" OR AI OR "generative ai") AND TS=(literac* OR "ai literacy" OR collaborat* OR "ai collaboration") AND TS=((educat* OR learn*) AND student*) AND TS=(survey* OR questionnaire* OR "self-report*" OR instrument* OR scale* OR framework*)
Scopus	TITLE-ABS-KEY("artificial intelligence" OR AI OR "generative ai") AND TITLE-ABS-KEY(literac* OR "ai literacy" OR collaborat* OR "ai collaboration") AND TITLE-ABS-KEY((educat* OR learn*) AND student*) AND TITLE-ABS-KEY(survey* OR questionnaire* OR "self-report*" OR instrument* OR scale* OR framework*)
ERIC via EBSCO host	("artificial intelligence" OR AI OR "generative ai") AND (literac* OR "ai literacy" OR collaborat* OR "ai collaboration") AND ((educat* OR learn*) AND student*) AND (survey* OR questionnaire* OR "self-report*" OR instrument* OR scale* OR framework*)
PsycINFO via EBSCO host	("artificial intelligence" OR AI OR "generative ai") AND (literac* OR "ai literacy" OR collaborat* OR "ai collaboration") AND ((educat* OR learn*) AND student*) AND (survey* OR questionnaire* OR "self-report*" OR instrument* OR scale* OR framework*)
IEEE Xplore	("artificial intelligence" OR AI OR "generative ai") AND (literac* OR "ai literacy" OR collaborat* OR "ai collaboration") AND ((educat* OR learn*) AND student*) AND (survey* OR questionnaire* OR "self-report*" OR instrument* OR scale* OR framework*)

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5820
https://doi.org/10.24059/olj.v30iSI2.5820

Revisiting Learning Presence in the Community of Inquiry Framework: Insights from a Systematic and Meta-analytic Review

Chungsoo Na
Hanyang University
Fan Yang
University of Arizona
Andrew Walker
Utah State University

Abstract

Learning presence (LP) has been proposed as a "fourth presence" in the Community of Inquiry (CoI) framework and linked to self-regulation and self-efficacy; however, its conceptualization and empirical relationships remain unclear. To address these gaps, this study conducted a systematic review on the definitions of LP (RQ1) and meta-analysis examining LP's associations with the three established CoI presences, satisfaction, and learning outcomes (RQ2). Guided by PRISMA (Page et al., 2021), we identified 51 studies for the systematic review and 18 studies comprising 103 effect sizes (N = 19,827) for the meta-analysis published from 2010 to 2025. Results indicated that the term "learning presence" is widely used to denote LP and that it is most commonly conceptualized as regulation of learning. Meta-analytic findings revealed 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). Meta-regression analyses revealed no significant moderator effects, except for the educational levels. Overall, this synthesis clarifies the conceptual landscape of LP within CoI research and provides robust evidence of its consistent relationships with key presences and student learning. Future research should pursue clearer conceptual definitions and refinements, more standardized measurement approaches, and greater methodological and reporting rigor.

Keywords: Learning Presence, Learner Presence, Community of Inquiry, Meta-analysis, Systematic Review

Introduction

Due to rapid technological developments, increasing demand for flexible learning pathways, and the expansion of distance education, online and blended learning have become significant components of contemporary education. Unlike face-to-face learning settings, where interaction occurs through physical co-presence and immediate verbal and nonverbal communication, online learning relies on digital platforms to serve as the primary medium for interaction among instructors and learners. As a result, researchers have been working on establishing theoretical frameworks to explain how meaningful learning occurs in digitally mediated environments. One of the most influential frameworks is the Community of Inquiry (CoI), developed by Garrison et al. (1999). The CoI framework suggests that effective online learning occurs through the interplay of three core components: cognitive presence, social presence, and teaching presence, all of which work together to create a collaborative and engaging learning experience. It highlights the importance of purposeful course design, active facilitation, and collaborative engagement in fostering student learning and satisfaction in online settings.

In response to the potential limits in the original CoI framework and the ongoing theoretical refinement, some scholars have proposed the addition of a fourth element: learning presence

Corresponding author: Chungsoo Na, chungsoo.na@outlook.com
This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 The Author(s).

(LP) to the existing CoI framework (Shea & Bidjerano, 2010, 2012; Shea et al., 2014, 2022). Learning presence refers to learners' self-regulatory processes, including goal setting, strategic planning, monitoring, and reflection, which influence how they participate in online learning inquiry. Critics of CoI believe that the existing cognitive presence does not fully capture the metacognitive and motivational processes that enable learners to manage learning in online learning environments. As online learning often requires greater autonomy and self-regulation than face-to-face settings, proponents of the fourth presence state that the new component will help clarify the role of self-regulated learning in sustaining cognitive presence and interacting with teaching and social presence (Shea & Bidjerano, 2012). Additionally, the addition of learning presence would also expand the explanatory power of the new CoI framework by explicitly acknowledging learners' self-regulatory role in online learning environments (Dunmoye et al., 2025).

Despite the substantial body of research related to CoI, findings regarding the magnitude of relationships among cognitive, social, and teaching presence have varied across contexts (Annand, 2011; Armah et al., 2023; Martin et al., 2022; Zhang et al., 2023). Additionally, there is a lack of comprehensive review of how LP is operationalized in existing research and how the fourth presence is associated with the other three components. A systematic review and meta-analysis are therefore needed to aggregate empirical evidence, estimate pooled effect sizes, and clarify the structural relationships among the four presences. Such synthesis can contribute to refining the CoI framework and providing valuable insights for instructional design and online teaching practice.

Literature Review

The Community of Inquiry Framework and Its Theoretical Evolution

The Community of Inquiry (CoI) framework, proposed by Garrison et al. (1999), provides a foundation for meaningful learning in online and blended environments by highlighting the triadic interplay among cognitive, social, and teaching presences. As the core of the CoI framework, cognitive presence (CP) is defined as "the extent to which learners are able to construct and confirm understanding through sustained reflection and discourse" (Garrison et al., 2001, p. 11). As suggested by Garrison and Arbaugh (2007), CP is manifested in learners' ability to reflect, construct, and build meaning through their involvement in reflective practices and phases, including triggering events, exploration, integration, and resolution. Through these phases, learners construct and validate meaning.

Social presence (SP) reflects "the ability of participants to identify with the group or course of study, communicate purposefully in a trusting environment, and develop personal and affective relationships by way of projecting their individual personalities" (Garrison, 2011, p. 34). As suggested by Garrison et al. (2001), SP includes three categories: affective expression, open communication, and group cohesion. As a necessary precursor to CP (Swan et al., 2009), SP plays an important role in the CoI framework. The study by Shea and Bidjerano (2008) identified that SP is an important mediator of the relationship between learners' perception of teaching presence and their evaluation of learning. It also establishes the notion that interpersonal trust and a sense of belonging are necessary for sustained collaborative inquiry, which affects learning outcomes.

Teaching presence (TP) is defined as "the design, facilitation, and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes" (Anderson et al., 2001, p. 5). The purpose of TP, therefore, is to promote CP or SP, or both, with a set of instructional activities, including the designing and organizing of instructional materials, the facilitating of discourse, and the assisting of students with questions (Authors, 2024; Garrison et al., 1999).

Despite its empirical robustness, the original CoI model has been critiqued for under-specifying the learner's internal regulatory processes and personal agency in learning processes (Chang et al., 2025). A significant development is the expansion of the CoI model to include emotional presence (Cleveland-Innes & Campbell, 2012), institutional presence (Zhang & Zhu, 2023), and learning presence (Shea & Bidjerano, 2012). The extension of CoI presences reflects a recognition that learning is multidimensional and that the framework must adapt to capture the complexity of modern environments (i.e., technology-mediated environments), while also acknowledging learners as active agents in constructing their own educational experiences (Chang et al., 2025).

Learning Presence: Theoretical Foundations and Conceptual Ambiguity

In response to a conceptual gap regarding the role of learners within the original CoI framework, LP was introduced as a new element to the revised CoI model (RCoI, Shea et al., 2014). Although the CoI framework consists of TP, SP, and CP as core dimensions of the educational experience, the original framework paid limited attention to how learners actively regulate their own learning processes and exercise their agencies (Richardson et al., 2024; Shea & Bidjerano, 2010, 2012)

As online learning increasingly requires self-direction and metacognitive control, researchers have questioned whether the existing presences sufficiently represent learners' proactive roles in both individual's learning processes.

Shea and Bidjerano (2010) articulated learning presence (LP) as grounded in the socio-cognitive motivational model (Zimmerman & Schunk, 2001) and self-regulated learning theory (Zimmerman, 2008), defining it as learners' self-efficacy, effort regulation, and strategic engagement—factors that contribute to meaningful knowledge construction in online discourse. Later, Shea and Bidjerano (2012) conceptualized LP explicitly as self-regulation and positioned it as a moderator of the relations between SP and CP, and between TP and CP, respectively. Building on these extended CoI models, scholarly attention gears towards the conceptual roles and necessity of LP in the CoI framework. Proponents argue that LP addresses a theoretical gap by explicitly accounting for learners' self-regulatory and metacognitive processes, citing its positive associations with the three established presences and with learning outcomes (e.g., Cho et al., 2017, Na et al., 2024a, b). In contrast, critics contend that LP substantially overlaps with elements already embedded within CP and TP, rendering a distinct fourth presence theoretically redundant (Garrison & Akyol, 2013). More recently, Shea et al. (2022) extended LP to include self-, co-, and shared regulation, framing it as a construct that captures both individual and collective regulatory processes in online learning communities.

Despite these theoretical refinements and ongoing debates, questions remain regarding how LP has been conceptualized and operationalized across empirical research. Empirical studies operationalize LP variously as self-regulated learning (Cho et al., 2017), meta-cognitive strategies (Li & Lau, 2025; Zhang, 2018) and shared meta-cognition (Garrison, 2022), self-efficacy (Chen et al., 2025), motivation (Tunceren, 2017), effort regulation (Shea & Bidjerano, 2010), or a general indicator of learner agency. This fragmented conceptualization of LP has blurred its boundaries with TP, CP, and broadly related motivational constructs, leading to inconsistent theoretical positioning. As a result, LP's role within the CoI framework remains diffuse and insufficiently specified. Therefore, research synthesis is needed to clarify how LP has been conceptualized and to examine its empirical associations with other presences and students' learning.

What We Already Know from Prior Research Syntheses

Over the past decade, a growing body of empirical research on the CoI framework has been consolidated through multiple meta-analyses and systematic reviews. (see Table 1). The finding of the meta-analyses consistently indicates that TP, SP, and CP are strongly and positively associated with students' perceived learning and satisfaction, although the strength of these associations varies. TP and SP tend to show stronger links with satisfaction (Caskurlu et al. 2017; Richardson et al., 2017), whereas CP tends to be more strongly related to perceived learning (Martin et al., 2022; Yang et al., 2025). However, when learning is operationalized as objective performance outcomes (i.e., actual learning), the associations are notably weaker (Martin et al., 2022). This suggests that CoI presences are more closely aligned with students' subjective learning experiences. While prior syntheses have solidified the empirical standing of the original triadic model, they largely overlook developments introduced in the revised CoI framework. In particular, the positioning and empirical role of LP remain underexamined.

To address this gap, Chang et al. (2025) conducted the first systematic review focused exclusively on LP within the revised CoI framework (2010–2024), synthesizing 26 quantitative studies. Their review positions LP as a critical presence representing learners' regulatory learning processes and cognitive engagement. Further, LP has been conceptualized as a mediator or moderator in the relations to TP, SP, and TP, and, like these three presences within the CoI framework, shows positive associations with learning outcomes, engagement, and satisfaction. However, the strength of these associations substantially differs across cultural (e.g., individualist vs. collectivist), instructional (e.g., blended vs. fully online), and educational (e.g., K–12 vs. undergraduate) contexts.

Despite its valuable contributions, Chang et al. (2025) included only studies employing standardized survey measures. While this systematic review of LP (Chang et al., 2025) was an important contribution, it excluded a substantial body of CoI research grounded in content analyses of interviews, discourse in online setting, and discussion transcripts. Given the longstanding tradition of qualitative inquiry within CoI research, this methodological restriction offers a limited view of how LP has been conceptualized. Moreover, their conclusions regarding LP's relations with other presences and learning outcomes were based on selected correlations from representative studies rather than a quantitative synthesis of all available effect sizes, limiting the precision and generalizability of their findings.

Table 1

Previous Research Syntheses on the CoI Framework

Publication	Articles	Research Scope and Key Findings
Richardson et al. (2017)	25 articles (1992 ~ 2015)	Meta-analyses on relations between SP and perceived learning and satisfaction ● SP and satisfaction: $r = .56$ ($p < .001$; $k = 26$); course length, disciplinary areas, and measurement scale are significant moderators. ● SP and perceived learning: $r = .51$ ($p < .001$; $k = 26$); course length, disciplinary areas, and target sample are significant moderators.
Caskurlu et al. (2020)	30 articles (1997 ~ 2018)	Meta-analyses on relations between TP and perceived learning and satisfaction ● TP and satisfaction: $r = .59$ ($p < .001$; $k = 26$); course length, disciplinary areas, and scales are significant moderators. ● TP and perceived learning: $r = .60$ ($p < .001$, $k = 23$); target sample and course length are significant moderators.
Martin et al. (2022)	19 articles for meta-analysis and 55 articles for systematic reviews (2000 ~ 2019)	Meta-analyses on relations between three presences (TP, CP, and SP) and actual and perceived learning and satisfaction. ● TP and actual learning: $r = .35$ ($p = .001$, $k = 6$); perceived learning: $r = .39$ ($p < .001$, $k = 10$); satisfaction: $r = .51$ ($p < .001$, $k = 10$) ● SP and actual learning: $r = .20$ ($p = .042$, $k = 6$); perceived learning: $r = .43$ ($p < .001$, $k = 10$); satisfaction: $r = .45$ ($p < .001$, $k = 10$) ● CP and actual learning: $r = .25$ ($p < .001$, $k = 6$); perceived learning: $r = .66$ ($p < .001$, $k = 10$); satisfaction: $r = .59$ ($p < .001$, $k = 10$)
Yang et al. (2025)	32 articles (~ 2024)	Meta-analyses on relations between CP and perceived learning and satisfaction ● CP and satisfaction: $r = .63$ ($p < .001$; $k = 42$); disciplinary areas, course length, and scales are significant moderators. ● CP and perceived learning: $r = .69$ ($p < .001$; $k = 27$); disciplinary areas, course length, and scales are significant moderators.
Chang et al. (2025)	26 articles (2010 ~ 2024)	Synthesize quantitative research on LP within the RCoI framework. ● Focused on demographic and educational context, measurement of LP, methodological aspects of validating LP ● Cultural background, educational level, instructional modalities, and individual differences may affect LP itself, as well as its relationships with learning outcomes and other presences within the CoI framework.
Krebs-Lazendic et al. (2026)	10 articles (2010-2022)	Critical analysis exploring TP (i.e., synchronous instructor presence) and its impact on students' learning ● Course grades and distributions, educational levels, course domains, synchronous components, course delivery. ● Lack of adequate statistical reporting for effect size estimation and limited methodological rigor.

Note. TP = teaching presence, SP = Social presence, CP = Cognitive presence, RCoI = Revised Community of Inquiry framework, *k* = the number of effect sizes. *k* = the number of effect sizes

The Present Study

Throughout the past decade, research on LP within the CoI framework has expanded, yet it remains conceptually fragmented across literature. While a recent systematic review of LP (Chang et al., 2025) has emerged, its restrictive selection of primary studies limits a holistic understanding of how LP has been conceptualized. Moreover, variability in LP has been illustrated through correlations drawn from representative studies rather than synthesized average effect sizes, leaving its overall empirical magnitude unclear. Like other Online, Distance, and Digital Education (ODDE) reviews, there is also a need to improve the transparency of search, screening, coding, and analyses steps as well as account for data that are nested. To address these gaps, the present study conducts a qualitative systematic review and quantitative meta-analysis to clarify LP's theoretical positioning and empirical associations using Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA, Page et al., 2021) guidelines. Specifically, we examine two research questions:

RQ1: How has LP been defined and conceptualized in empirical research grounded in the CoI framework?

RQ2: What is the magnitude of the associations between LP and other presence (i.e., TP, SP, and CP) within the CoI framework, as well as student learning outcomes and satisfaction? To what extent do educational level, research location, and instructional modality moderate the association between LP and other presences?

By synthesizing and quantifying the existing evidence, this study contributes theoretically by sharpening LP's role within the CoI nomological network and methodologically by estimating pooled effect sizes and heterogeneity. Furthermore, the findings are expected to help identify which learning processes need more scaffolding in online and blended learning settings, thereby informing evidence-based design decisions.

Methods

This study followed the PRISMA guidelines to ensure transparency in search, screening, coding, and analytic procedures. Where relevant, and unless specified otherwise, we used a random effects model and set statistical significance thresholds to $p < .05$. Our choice of random effects is driven by the heterogeneity in samples and measurement approaches used to assess LP. This approach assumes that true effect sizes vary across studies and allows for generalization beyond a single underlying population effect.

Literature Search

In the literature search phase, we conducted systematic searches for conceptualization and meta-analyses of learning presence within the Community of Inquiry (CoI) framework. We searched multiple databases, including Education Source, ERIC, and APA PsycINFO (via EBSCOhost), as well as the Education Database and Dissertations & Theses (via ProQuest). The literature search was conducted in November 2025 and used the following search terms in the

title, abstract, and keywords: “Community of Inquiry” AND (“learning presence” OR “learner presence” OR self-regulate* OR co-regulate* OR effort regulation OR self-efficacy OR self-direct*). Further, given that the concept of learning (or learner) presence was proposed in 2010 (Shea & Bidjerano, 2010), we set the search period from 2010 to 2025. All detailed DB quires are available in Supplementary Material.

As part of an extended search strategy, we conducted a backward snowball search of recent reviews (e.g., Chang et al., 2025) to manually identify additional studies. A total of 408 records were identified through the database search and 8 records through the hand search, yielding a total of 416 records.

Screening of the Studies

We used Covidence to import the searching results and then removed duplicate. After removing all duplicates (n = 217), 199 records remained. During the screening phase, we employed a two-step process to assess study eligibility using Excel: (a) title and abstract screening, followed by (b) full-text screening. In the title and abstract screening phase, we excluded 46 records that did not meet the inclusion criteria (see Table 2): (a) written in English; (b) reporting empirical evidence; (c) examining learning (or learner) presence as a focal variable or topic; (d) targeting learners as the sample; and (e) conducted in online learning settings. The full text of two studies could not be retrieved. During the full-text screening phase, we excluded an additional 100 articles for the following reasons: lack of topic relevance (i.e., LP was not examined as a focal variable; n = 87), inappropriate target sample (e.g., instructional designers or faculty rather than learners; n = 6), and in-person learning settings (e.g., VR, n = 3), the same study published in different formats (n = 3), no empirical study (n = 1). Consequently, 51 studies were included in the final analysis for the systematic review (RQ 1).

For the meta-analysis (RQ 2), we further restricted the sample to studies reporting correlations between LP and other presences (i.e., TP, SP, and CP), as well as learning outcomes and satisfaction, resulting in 24 eligible studies. However, several studies did not report sufficient statistical information required for effect size calculation. Accordingly, we contacted corresponding authors (n = 8) to request the necessary information when only a single element (e.g., sample size or non-significant correlation values) was missing, and we received responses from two authors (Li & Lau, 2025; Usmani, 2021). If the authors did not respond or were unable to provide sufficient information, the corresponding outcomes with missing data were excluded from meta-analysis. The final meta-analytic dataset comprised 103 effect sizes across 18 studies (see Figure 1).

Table 2

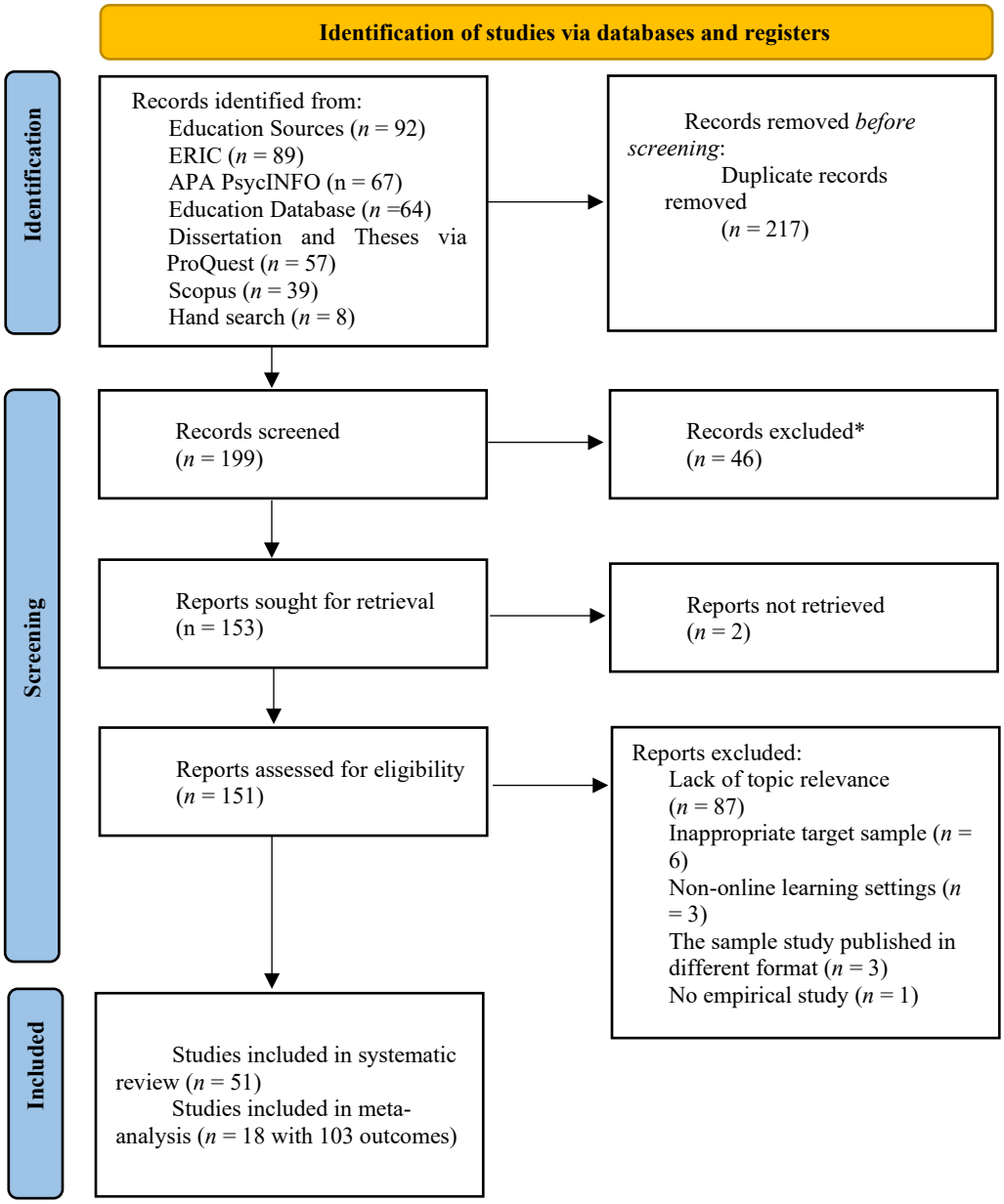
Inclusion and Exclusion Criteria

	Inclusion Criteria	Exclusion Criteria
Language	The study was written in English.	The study was not written in English.
Publication year	The study was published between 2010 and November 2025 (the period during which the research team conducted the search)	The study was published before 2010.

	Inclusion Criteria	Exclusion Criteria
Research type	The study included empirical evidence (e.g., quantitative, qualitative, or mixed-methods research) examining LP	The study did not include empirical evidence (e.g., systematic reviews such as Chang et al., 2025, or position papers such as Shea et al., 2022).
Topic relevance	The study explicitly mentioned LP was used as focal variables or topics in the study.	The study focused on LP-related variables (e.g., self-regulation, motivation, self-efficacy) but did not frame them explicitly as LP.
Target sample	The study targeted learners as the sample (e.g., undergraduates in online settings or pre-service and in-service teachers in professional development contexts).	The study did not target learners as the sample (e.g., instructors or other stakeholders of online courses).
Research Context	The study was conducted in online learning settings, including fully online, blended (or hybrid), flipped, or MOOC environments.	The study was not conducted in online learning settings (e.g., 100% face-to-face or VR settings).

Figure 1

Flow Diagram of the Search and Selection Procedure for Studies



Coding Procedures

After selecting the eligible studies, we extracted information from each study based on the following categories: (a) publication information, (b) LP conceptualization, (c) methodological characteristics, and (d) research context. Prior to coding, the first and second authors developed the coding scheme deductively based on information extracted from the included papers. Both raters independently double-coded all studies. For variables that could be coded nominally (e.g.,

research methods), interrater reliability (IRR) was calculated using Cohen's kappa. Coding disagreements were resolved through discussion and consultation with the third author.

For categories requiring qualitative coding (e.g., LP conceptualization), we adopted a multi-stage qualitative analysis approach. Qualitative data, such as authors' definitions and descriptions of LP, were analyzed using an inductive thematic coding process. First, the first author extracted relevant text segments (vignettes) describing LP from each study. Two researchers then independently reviewed these excerpts and generated preliminary codes based on recurring patterns in how LP was conceptualized. Through iterative discussions, similar codes were refined, merged, and organized into broader conceptual categories.

For example, studies defining LP as learners' ability to monitor, regulate, and manage their own learning processes were coded as self-regulation (e.g., Na et al., 2024b), whereas studies emphasizing shared regulation among team members were coded as "co-regulation" (e.g., Kim et al., 2022). Likewise, studies describing LP as learners' confidence in their ability to engage in learning activities were coded as self-efficacy (e.g., Alsayer, 2025). The resulting coding framework consisted of multiple codes organized into broader conceptual categories. After consensus on the coding framework was built, two researchers (first and second authors) independently coded a subset of studies and discussed discrepancies until agreement was reached. The final categories reported in Table 3 represent themes that emerged from the qualitative analysis. Inter-rater reliability was assessed using percent agreement, and coding disagreements were resolved through discussion and consultation with the third author. Table 3 provides a description of the coding schemes and the agreement levels for each coding category. To understand the landscape of empirical research on LP, we synthesized the included studies in narrative ways.

As variables associated with LP, we included the three presences of the CoI framework (i.e., TP, SP, and CP), as well as learner-related outcomes that have frequently been used in prior CoI meta-analyses (e.g., Caskurlu et al., 2020; Richardson et al., 2017), namely satisfaction and learning outcomes. Drawing on Martin et al. (2022), we further distinguished learning outcomes into perceived learning (e.g., perceived progress in Zhang, 2018) and actual learning (e.g., GPA in Alotebi, 2017) for the supplementary analyses.

Meta-Analysis Procedures

Calculation of Effect Sizes

To answer RQ2, we coded for effect sizes between (a) LP and other presences (i.e., TP, SP, and CP) within the CoI framework, including others that some authors expanded into and (b) LP and learning outcomes. Zero-order correlations (i.e., Pearson's r) were used as the primary effect size measure. When a study reported Spearman's ρ as the correlation coefficient, we converted it to Pearson's r using Moran's (1948) approach (e.g., Ma et al., 2017; Shea et al., 2014). We derived model-implied correlations from the standardized loadings and included them in the analysis (e.g., ElSayad, 2023; Wertz, 2022).

Pearson correlation coefficients were converted to Fisher's z values (i.e., r-to-z transformation) to stabilize variance and make the distribution normal (Cohn & Becker, 2003).

The first and third authors independently coded the statistical information necessary for calculating effect sizes from 30% of the included studies. Inter-rater reliability was substantially high (ICC = .82) and the first author proceeded with solo coding for the remaining 70%. Guided by Cohen (1988), we considered $r = .10$ as a small effect, $r = .30$ as a medium effect, and $r = .50$ or above as a large effect.

Model Selections

We estimated separate meta-analytic models for each variable (i.e., TP, SP, CP, satisfaction, and learning outcomes). This is because these constructs are theoretically distinct and may vary in their associations with LP, heterogeneity, and effect size distributions. When a study reported multiple correlations derived from the same sample, statistical dependency among effect sizes arose. Ignoring such dependency can lead to biased standard errors and distorted parameter estimates. To address this issue, we applied *correlated and hierarchical effects models*, combining multilevel meta-analysis with robust variance estimation (RVE), which accounts for within-study dependence among effect sizes (Pustejovsky & Tipton, 2022). To impute covariance structure for RVE, we assumed an intercorrelation of $\rho = .70$ among effect sizes within studies. After estimating the pooled effects, we back-transformed Fisher's *z* values to Pearson's *r* for interpretation.

Table 3

Coding Information and Inter-Rater reliability

Coding	Descriptions	Agreement Levels
Publication information		
We extracted author names, title, publication years, publication status (i.e., published or gray) from each study.		100% ($\kappa = 1.00$)
LP conceptualization		
How the paper referred to LP	We coded the terminology used in each study to refer to LP (e.g., Learning presence, Learner presence or others)	82.3%
How the paper conceptualized LP	We coded conceptual framing of LP (e.g., SRL, co-regulation, self-efficacy, motivation) in each study.	80.3%
Methodological characteristics		
Research methods	We coded the research methods used in each study. The research methods were coded into three categories: quantitative, qualitative, and mixed methods.	94.1% ($\kappa = .91$)
Data source	We coded the primary data source(s) used in each study, including surveys, interviews, and discussion transcripts.	78.4%
Measures	We coded the measurement approach used in each study, specifying the coding scheme for qualitative studies (e.g., Shea et al., 2012) and the instrument name or survey tool for quantitative studies (e.g., OSLQ in Barnard et al., 2009).	82.3%
Research Context		

Coding	Descriptions	Agreement Levels
Sample size	We coded the sample size for each study, including the number of participants in quantitative studies and the number of postings analyzed in qualitative studies.	86.3%
Educational level	We coded the educational level of the participants. ● Informal learning: participants were not enrolled in formal education programs, even though learning was delivered in an online setting ● Undergraduate: participants enrolled in undergraduate programs (or community college program) ● Graduate: participants enrolled in master's or doctoral program ● Post-secondary: participants were from higher education settings but did not clearly distinguish between undergraduate and graduate levels. ● K-12: participants enrolled in K-12.	86.3% (κ = .78)
Instructional modality	We coded the educational setting of each study, specifying whether it was conducted fully online, blended, and mixed. ● Online: Instruction delivered fully through online platforms ● Mixed: Instruction that included both online and blended in the study ● Blended: Instruction intentionally designed to integrate online and face-to-face components (e.g., hybrid formats).	96.1% (κ = .93)
Research location	We coded the geographic location of each study and categorized it into Asia, North America, or Africa.	100% (κ = 1.00)

Note. IRR = inter-rater reliability (Cohen's κ) for categorical variables and agreement levels (%) for qualitative variables.

Heterogeneity of Effect Sizes

We assessed heterogeneity among effect sizes using Cochran's Q test and the τ^2 statistic. The Q test evaluated whether the observed variability in effect sizes exceeded what would be expected from sampling error alone, with a significant Q indicating the presence of heterogeneity. Given that the Q statistics are sensitive to the number of studies included, we also reported τ^2 to quantify the within- and between-study variance in true effect sizes, which means the extent of heterogeneity that cannot be explained by sampling error alone (Borenstein, 2023). Further, I^2 was reported to indicate the ratio of true heterogeneity to the observed total heterogeneity (Higgins & Thompson, 2002).

To examine potential sources of heterogeneity, we conducted mixed-effects meta-regression analyses in which educational level, research location, and instructional modality were entered simultaneously as moderators in the same model. Drawing on prior systematic reviews suggesting that the magnitude of LP's associations with other presences may vary across cultural, educational, and instructional contexts (Chang et al., 2025), these three variables were included as theoretically informed moderators. Instead of traditional sub-group analyses (i.e.,

multiple single regressions), entering them concurrently (i.e., a single multiple regression) allowed us to estimate their unique effects while controlling the influence of the other moderators. For research location, only one study was classified as originating from Africa (ElSayad, 2023). To avoid unstable parameter estimates and potential overinterpretation based on a single case, this effect size was excluded from the meta-regression analyses.

Publication Bias

Lastly, we employed multiple approaches to evaluate potential publication bias. First, we visually inspected funnel plots, where symmetry suggests the absence of publication bias. Next, we performed Egger's regression test to statistically evaluate funnel plot asymmetry, with a significant result indicating potential small-study effects or publication bias. All analyses were conducted in R using the *metafor* package with restricted maximum likelihood (REML) estimation (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2026) packages in R (version 4.4.2, R core team, 2025).

Results

Overview of Selected Studies

All 51 studies identified and included in this review study were included in the systematic review, and 18 of them were further analyzed in the meta-analysis because of the quantitative elements reported in these studies. With respect to research design, quantitative studies constituted the largest proportion of the included literature (n = 27, 52.9%), followed by qualitative studies (n = 13, 25.5%) and mixed-methods studies (n = 11, 21.6%).

An analysis of the publication years reveals that these 51 studies were published across 14 distinct years between 2010 and 2025. Between 2010 and 2015, publication frequency was initially low and sporadic (1–3 studies per year), reflecting the early development of the field. A sharp increase occurred in 2017 (n = 7), representing the highest output during the early expansion phase. Although publication counts fluctuated between 2018 and 2020, research activity remained consistently higher than in the initial period. From 2023 to 2025, publication numbers stabilized at six studies per year, suggesting sustained scholarly interest and increasing consolidation of CoI.

Among the 51 studies included, two were conducted in multi-country contexts. Specifically, Xu (2019) reported a mixed participant sample drawn from several countries. Similarly, Kurek (2019) collected data from participants in both Germany and Poland. The remaining 49 studies were conducted within a single country/region. Among these, the majority were conducted in the United States (n = 19), followed by China (n = 9). South Africa contributed three studies, while Ghana, Turkey, Greece, the United Kingdom, and Egypt were each represented by two studies. All remaining studies were conducted within a single county/region, including Germany, Italy, Korea, Pakistan, Romania, Saudi Arabia, Singapore, and Hong Kong. Overall, the evidence base is heavily concentrated in the United States and China, with comparatively limited representation from other regions.

With respect to participant educational levels, the majority of studies focused on undergraduate populations (*n* = 26), accounting for approximately 51% of the sample. A smaller number examined mixed populations across levels (*n* = 11) or graduate students (*n* = 9). In contrast, relatively few studies were conducted in K–12 settings (*n* = 3) or informal learning contexts (*n* = 2). Overall, the distribution indicates a strong emphasis on higher education contexts, particularly undergraduate students, with limited representation of K–12 and informal learning environments.

RQ1: Conceptualization of LP

As a way of conceptual analysis, we examined (a) the terminology used to refer to LP, (b) how LP was conceptualized, and (c) which instruments or coding scheme were used to analyze LP across empirical studies. We first analyzed the terminology used to refer to this construct. As shown in Figure 2, the majority of studies explicitly adopted the term *learning presence* (n = 43) and *learner presence* (n = 6). Moreover, a range of alternative terms were used to describe conceptually related phenomena, including *language learning presence* (*n* = 1, Fayram, 2017) and *self-regulated learning presence* (*n* = 1, Shu et al., 2025).

While this variation in terminology suggests surface-level differences, a deeper examination reveals that much of the conceptual divergence centers on whether and how LP is anchored in frameworks of learning regulation at the individual level. Of the total corpus, six studies did not mention any form of regulation, whereas the majority (*n* = 45) explicitly referred to self-regulated learning (SRL). However, even among those 45 studies, a few studies (Chandler, 2021; Hayes et al., 2015; Kim et al., 2014; Usmani, 2021; Zheng et al., 2023; Zou & Jiang, 2025) explicitly conceptualized LP as a form of co- and socially shared regulation.

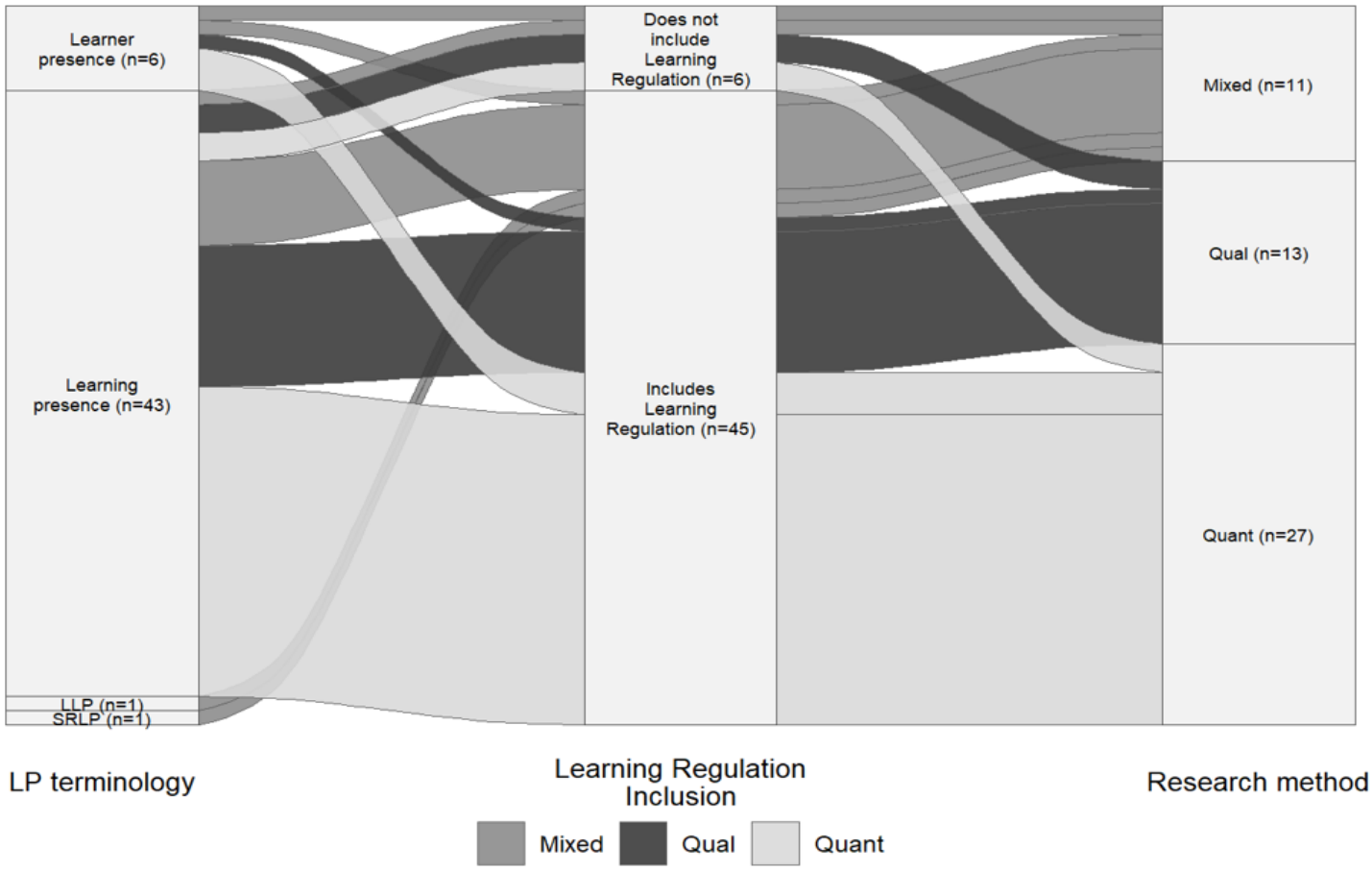
The remaining studies that framed LP beyond SRL perspectives adopted diverse conceptualizations. For example, Xu (2019) and Bosch (2022) explicitly used the term *self-directed learning* to describe LP. Alsayer (2025) conceptualized LP primarily in terms of self-efficacy. Similarly, Gunbatar (2021) interpreted LP as reflecting motivational dimensions, including self-efficacy, general motivation, and changes in study habits, based on coded student interview data. Finally, Volchok et al. (2018) operationalized LP using two behavioral indicators, including scores on extra credit assignments and scores on the first online quiz. Together, these findings illustrate somewhat conceptual heterogeneity in how LP has been defined. Nevertheless, even these varied conceptualizations largely remain at the individual level.

To further examine how LP has been measured and coded from a learning regulation perspective, we categorized the 45 studies into research paradigms for further analysis (see Figure 2). In quantitative research, most studies employed surveys to capture LP through various instruments such as the MSLQ (Pintrich et al., 1993), the OSLQ (Barnard et al., 2009), learning presence measures used by Shea and Bidjerano (2010), WebTalk (Wertz, 2022), or researcher-developed measures. In qualitative research, many studies examined students' online interaction logs, discussion posts, and interview data, drawing on theoretical perspectives of SRL or co-regulation. The majority of coding schemes of LP were rooted in Zimmerman's SRL framework, operationalizing regulation through four phases such as forethought and planning, monitoring, strategy use, and reflection. Several studies adopted the coding definitions introduced by Shea et al.'s (2014) operational definitions, whereas others refined and modified them to align with their

particular research contexts (e.g., Blain 2019). All detailed coding results are available in Supplementary Material.

Figure 2

Distribution of LP Terminology, Learning Regulation Inclusion, and Research Methods Across Included Studies



Note. LLP = language learning presence, SRLP = self-regulated learning presence, Mixed = mixed method design, Qual = qualitative research design, and Quant = quantitative research design.

RQ2. Meta-Analysis of LP

Whereas RQ1 examined LP from a conceptual perspective, RQ2 quantified the extent to which LP is associated with other presences, as well as its relations with learning outcomes and student satisfaction. The 18 studies included in the current meta-analysis comprised 103 correlation coefficients based on 19,827 individual students. Compared to learning outcomes and satisfaction, TP, SP, and CP were more frequently examined in relation to LP. Given the limited number of effect sizes available for satisfaction ($k = 6$) and learning outcomes ($k = 9$), meta-regression analyses were restricted to the relations between LP and the other presences within the CoI framework, with educational level, research location, and instructional modality entered as moderators.

Outliers

In the present analysis, we identified one extremely high correlation coefficient. Notably, ElSayad (2023) reported a correlation of $r = .95$ between the latent variables of LP and CP. Sensitivity analyses indicated that excluding this data point resulted in a meaningful change in the magnitude of the CP–LP relations ($\Delta r = .04$). Therefore, we excluded this effect size from the final analysis.

Publication Bias

Publication bias was examined using visual inspections of funnel plots (see Figure 3) and Egger's regression tests with cluster-robust variance estimation (see Table 4) with mixed results. Visual inspection of the funnel plot showed that the distribution of effect sizes was slightly asymmetric across all five models, indicating possible publication bias. However, as shown in Table 4, the regression coefficients via Egger's test (SE as a predictor) were not statistically significant across the five models, indicting no publication bias. Our conclusion is that the findings should be interpreted cautiously.

Table 4

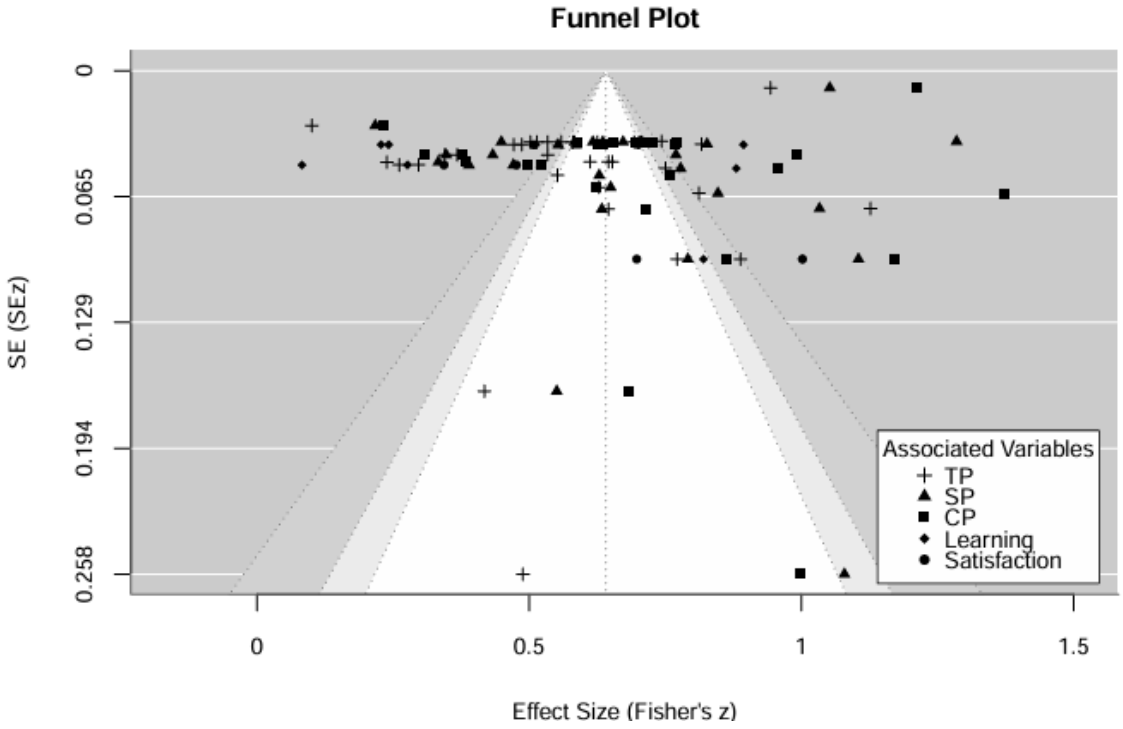
Egger's Regression Test Results

					95% CI	
	β	*SE*	*satt.df*	*p*	*LL*	*UL*
TP	0.02	1.20	2.55	.986	-4.20	4.25
SP	0.93	1.26	2.23	.532	-4.00	5.86
CP	1.91	1.74	2.58	.363	-4.16	7.89
Learning outcomes	8.59	2.09	1.08	.137	-13.56	30.74
Satisfaction	5.35	4.09	1.00	.416	-46.65	57.4

Note. TP = teaching presence, SP = social presence, CP = cognitive presence, *β* = regression coefficients of precision (SE) in Egger's regression test. *satt.df* = approximate degrees of freedom using Welch-Satterthwaite Formular.

Figure 3

Funnel Plots of Fisher's z-transformed Correlations (n = 103)



Note. TP = teaching presence, SP = social presence, CP = cognitive presence, Learning = learning outcomes.

Overall Effect Sizes Between Learning Presence and Other Presences

The average effect sizes indicated significant, large, and positive associations between LP and TP (r = .55, t(16.6) = 10.5, p <. 001, 95% CI[.46, .63]), SP (r = .62, t(15.7) = 10.8, p < .001, 95% CI[.53, .71]), and CP (r = .66, t(13.8) = 9.76, p < .001, 95% CI[.55, .74]). As indicated by significant Q statistics and large τ^2 in Table 5, substantial heterogeneity was observed across effect sizes in TP ($\tau^2_{between}$ = .056, τ^2_{within} = .003, I^2 = 97.7%), SP ($\tau^2_{between}$ = .068, τ^2_{within} = .007, I^2 = 98.2%), and CP ($\tau^2_{between}$ = .087, τ^2_{within} = .004, I^2 = 98.6%). The main sources of variability were primarily attributable to between-study differences rather than to within-study sampling error.

Table 5

Average Effect Sizes on the Relations between LP and Other Presences within CoI framework

				95% CI		Heterogeneity		
	k	*m*	*r*	*LL*	*UL*	*Q*(*df*)	$\tau^2_{between}$	τ^2_{within}
LP-TP	32	18	.55***	.46	.63	*Q*(31) = 1476.08***	.056	.003
LP-SP	29	17	.62***	.53	.71	*Q*(28) = 1476.08***	.068	.007
LP-CP	27	16	.66***	.55	.74	*Q*(26) = 2231.48***	.087	.004

Note. LP = Learning Presence, TP = Teaching Presence, CP = Cognitive presence, *k* = the numbers of effect sizes, *m* = the numbers of primary studies, *r* = average correlation coefficient, LL = lower limit of 95% CI, UL = upper limit of 95% CI, τ^2_{within} = heterogeneity of within-study, $\tau^2_{between}$ = heterogeneity from between-study.

*$p < .05$, **$p < .01$, ***$p < .001$.

Overall Effect Sizes Between LP and Learning Outcomes and Satisfaction

Similar to the relationships between LP and the three presences, the average effect sizes indicated significant, large, and positive associations between LP and learning outcomes ($r = .53$, $t(2.94) = 3.45$, $p = .042$, 95% CI [.04, .81]) and satisfaction ($r = .53$, $t(1.99) = 5.22$, $p = .035$, 95% CI [.10, .80]). Notably, we differentiated learning outcomes into actual learning (e.g., GPA) and perceived learning. The two categories yielded different effect sizes: actual learning ($r = .22$, based on 2 studies with 4 effect sizes, $p = .09$) and perceived learning ($r = .66$, based on 3 studies with 5 effect sizes, $p = .04$). As indicated by significant Q statistics and large τ^2 in Table 6, substantial heterogeneity was observed across effect sizes in learning outcomes ($\tau^2_{between} = .084$, $\tau^2_{within} = .056$, $I^2 = 98.57\%$). Table 6 indicated substantial heterogeneity was observed across effect sizes in learning outcomes ($\tau^2_{between} = .084$, $\tau^2_{within} = .056$, $I^2 = 98.5\%$) and satisfaction ($\tau^2_{between} = .028$, $\tau^2_{within} = .018$, $I^2 = 94.6\%$). Contrary to the relations between LP and other presences, the main sources of heterogeneity were both between- and within-study differences.

Table 6

Average Effect Sizes on the Relations between LP and Learning Outcomes

				95% CI		Heterogeneity		
	k	*m*	*r*	*LL*	*UL*	*Q*(*df*)	$\tau^2_{between}$	τ^2_{within}
LP-Learning outcomes	9	4	.53*	.04	.81	*Q*(8) = 735.99***	.084	.056
Perceived learning	5	3	.66*	.36	.84	*Q*(4) = 187.37***	.000	.035
Actual learning	4	2	.22	-.16	.55	*Q*(3) = 18.56***	.000	.004
LP-Satisfaction	6	3	.54*	.10	.80	*Q*(5) = 77.70***	.028	.018

Note. LP = Learning Presence, *k* = the numbers of effect sizes, *m* = the numbers of primary studies, *r* = average correlation coefficient, LL = lower limit of 95% CI, UL = upper limit of 95% CI, τ^2_{within} = within-study variation, $\tau^2_{between}$ = between-study variation.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Meta Regressions

Table 7 presents the results of separate meta-regression models examining the relations between LP and the other presences, simultaneously entering three moderators: educational level, cultural context, and instructional modality. Undergraduate students, North America, and online learning were specified as the reference groups for each moderator, respectively. To satisfy the linearity assumption of the regression model, all results were interpreted and reported in Fisher's z-transformed units.

Across the three meta-regression models examining the relations between LP and the other presences, most moderators did not demonstrate significant effects. Educational level and instructional modality were generally not associated with variability across effect sizes. One exception was identified for SP in which studies conducted with samples in graduate showed significantly stronger associations compared to undergraduate samples (β = 0.50, p = .029). For research locations, no significant differences were observed for studies conducted in Asia.

Table 7

Results of Meta-Regressions on the Relations Between LP and the Three Presences (Fisher's z-Transformed)

	TP				SP				CP			
	β	*SE*	95% CI		*β*	*SE*	95% CI		*β*	*SE*	95% CI	
			LL	*UL*			*LL*	*UL*			*LL*	*UL*
Intercept	0.60**	0.12	0.29	0.92	0.58*	0.16	0.16	1.00	0.78**	0.17	0.33	1.23
Educational Levels (ref: undergraduate)												
Graduate	-0.13	0.12	-0.45	0.18	0.50*	0.16	0.08	0.92	0.20	0.17	-0.25	0.65
K-12	-0.13	0.09	-0.46	0.19	-0.03	0.11	-0.44	0.37	-0.01	0.20	-0.73	0.70
Post-secondary	0.19	0.10	-0.10	0.47	0.24	0.10	-0.07	0.56	0.31	0.13	-0.09	0.71
Research Location (ref: North America)												
Asia	0.06	0.11	-0.21	0.32	0.21	0.16	-0.18	0.60	0.07	0.14	-0.28	0.41
Instructional Modality (ref: online)												
Blended	-0.17	0.10	-0.46	0.11	-0.21	0.11	-0.54	0.12	-0.23	0.13	-0.63	0.16
Mixed	-0.36	0.29	-1.64	0.91	-0.16	0.44	-1.90	1.57	-0.80	0.29	-1.70	0.10

Note. TP = teaching presence, SP = social presence, CP = cognitive presence, *β* = z-transformed regression coefficients.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Discussion

As LP has increasingly been regarded as a fourth presence within the CoI framework, research on this topic has steadily grown. However, prior studies differ in how LP is conceptualized and operationalized, and even the terminology used to describe the construct varies across the literature. Moreover, to our knowledge, no quantitative synthesis has examined how LP relates to the three established CoI presences as well as to student learning outcomes. The present review investigated how LP has been conceptualized in empirical research using qualitative synthesis (RQ1) and meta-analyzed its quantitative associations with TP, SP, CP, as well as satisfaction and learning outcomes (RQ2).

RQ1: Conceptualization of LP

Despite terminological variation (e.g., learning presence or learner presence), the conceptual analysis of 51 empirical studies indicates that a substantial proportion of LP research (n = 45, 88.2%) remains deeply rooted in the regulation of learning, particularly SRL. This indicates that LP is primarily operationalized as an individual learner's capacity to plan, monitor, and evaluate their own learning in online or blended contexts. Moreover, even studies (n = 6, 11.8%) that do not explicitly frame LP within a regulatory learning perspective tend to conceptualize it in closely related individual-level terms, such as self-directed learning (Bosch & Laubscher, 2022; Xu, 2019), the use of metacognitive strategies (Honig & Salmon, 2021), learning habits (Günbatar, 2021), or motivational constructs (e.g., self-efficacy). This pattern suggests that LP has been predominantly interpreted through an individualistic lens, even when regulatory terminology is not directly invoked.

This orientation is also reflected in how LP has been measured in empirical research. Quantitative studies commonly rely on self-report instruments capturing SRL strategies, such as the MSLQ (Pintrich et al., 1993), OSLQ (Bernard et al., 2009), WebTalk (Wertz, 2022) or the LP scale validated by Shea and Bidjerano (2010). Likewise, in qualitative research, learner artifacts (e.g., discussion postings, transactions, or interviews on learning experiences) are frequently analyzed according to coding schemes used in Shea et al., (2014) which is based on the four phases of Zimmerman's cyclical model of self-regulation (Zimmerman, 2008). Collectively, these approaches indicate that LP has been predominantly conceptualized and assessed as an individual-level regulatory construct, with relatively less attention to regulatory processes in collaborative settings.

However, recent theoretical developments have suggested that conceptualizing LP solely as an individual regulatory process may not fully reflect the collaborative dimensions of learning emphasized in online inquiry environments (Garrison, 2022; Shea et al., 2022). Building on recent theoretical developments, particularly those positioned by Shea et al. (2022), LP has recently been conceptualized by some scholars as a broader regulatory construct encompassing self-, co-, and socially shared regulation of learning (DiDonato, 2013; Hadwin & Oshige, 2011). From this perspective, LP reflects not merely the use of metacognitive or self-regulated strategies in individual learning phases, but the collaborative learning processes through which learners negotiate shared goals (Järvelä & Hadwin, 2013), monitor collaborative learning progress (Näykki et al., 2017), regulate motivation and emotion within group contexts (Järvenoja et al., 2020).

Furthermore, conceptualizing LP solely as an individual-level regulatory learning process may weaken its conceptual distinctiveness. Although CP is partially a collaborative construct within the CoI framework, its operationalization often reflects learners' individual inquiry and meaning-making processes, thereby increasing potential conceptual overlap between LP and CP. In fact, SRL processes such as planning, reflection, and monitoring are conceptually similar to the phases emphasized in CP, which may yield insufficient conceptual distinctions between the two constructs. Such overlap is reflected in ongoing debates regarding the distinctive role of LP within CoI (Garrison & Akyol, 2013). Therefore, reconceptualizing LP as a comprehensive learning regulation construct-overarching self-, co-, and socially shared regulation does more than just evolve the concept of LP; it provides the clarity needed to strengthen the CoI framework.

Based on the findings of this review, we propose that LP can be conceptualized as the regulatory processes through which learners individually and collectively plan, monitor, evaluate, and adapt their learning in online inquiry environments. This definition encompasses self-regulation, co-regulation, and socially shared regulation of learning (Shea et al., 2022) while distinguishing LP from cognitive presence by emphasizing the regulation of learning processes rather than the construction of knowledge itself.

RQ 2: Meta-Analysis of LP

Regarding RQ2, the results of the present meta-analysis partially offer empirical support for these conceptual concerns of LP. Consistent with prior research (e.g., Shea & Bidjerano, 2010, 2012), LP demonstrated strong positive associations with the three core presences of the CoI framework (TP, SP, and CP). The magnitude of these correlations was substantial and positive, with CP (r = .66), SP (r = .62), and TP (r = .55). Interpreted in terms of shared variance, approximately 43.6% of the variance in LP overlaps with CP, and about 38% overlaps with SP. Such substantial overlap in variance implies that LP is not only closely intertwined with the other presences within the CoI framework, but also raises conceptual questions about the unique role of LP.

Given that a large number of the studies in this review defined LP primarily from an individual regulatory perspective, these findings reinforce the notion that regulatory processes are deeply embedded within the broader role of CP in online learning environments. At the same time, they raise important questions about the distinctiveness of LP as a separate presence. Future research should examine how LP relates to the other presences in CoI framework when it is defined more comprehensively to include self-, co-, and socially shared regulation. Such investigations could help enhance the theoretical and observed distinctiveness of LP.

Next, LP demonstrated substantial positive, strong correlations with both learning outcomes (r = .53) and satisfaction (r = .54). The magnitude of these associations is comparable to prior meta-analytic findings examining the relationships between the established CoI presences and learning outcomes (Caskurlu et al., 2020; Richardson et al., 2017; Martin et al., 2022). These results suggest that, similar to TP, SP and CP, LP is meaningfully related to students’ satisfaction and learning in online and blended learning environments.

An especially noteworthy finding emerged when distinguishing between actual and perceived learning outcomes. The association between LP and perceived learning (r = .66) was stronger than that with objectively measured learning (r = .22). This pattern aligns with Martin et al.'s (2022) meta-analysis, which reported that three presences were more strongly related to perceived learning than to objective performance measures. One possible reason is that LP is typically measured using self-reported questionnaires, which are methodologically aligned with perceived learning rather than objective performance. Such self-report measures tend to inflate effect sizes, thereby increasing their association with subjective outcomes (Cote & Buckley, 1987; Podsakoff et al., 2003).

To explore the substantial heterogeneity in the associations between LP and the three CoI presences, we conducted meta-regression analyses. Contrary to Chang et al.'s (2025) expectation that educational level, cultural variation, and instructional modality would meaningfully moderate effect sizes, the results revealed relatively limited moderator effects. Additionally, educational level significantly moderates the relations between LP and SP, with stronger associations observed at the graduate level. One possible explanation is that, compared to undergraduate education, graduate contexts require greater learner agency and more advanced self-regulatory capacity (Artino & Stephens, 2009; Xu et al., 2023). Accordingly, regulatory processes may be more activated through socially mediated learning activities, such as peer-review and scholarly critique. This distinctive nature of graduate education may lead LP and SP to be more tightly intertwined.

Despite the substantial heterogeneity observed across study-level effect sizes, only a limited number of moderators reached statistical significance. This pattern suggests the potential presence of unexamined moderators. Future research should further investigate additional contextual, methodological, and theoretical factors that may account for this heterogeneity.

Lastly, our findings also have implications for online teaching and learning. Across studies, LP demonstrated consistently strong associations with TP, SP, CP, learning outcomes, and satisfaction. While these findings do not argue for causal relations, they suggest that learners' regulatory processes are closely linked to meaningful online learning experiences. Accordingly, online course designers and instructors may benefit from incorporating learning activities that make regulatory processes more visible and support learners' planning, monitoring, and reflection throughout the learning process. At the practical level, this can be exemplified as structured reflection prompts, self-monitoring activities, progress tracking tools through learning analytical tools, and opportunities for peer feedback. These approaches may help create learning environments that better support their regulatory processes and active engagement within online communities of inquiry. At the policy level, universities and online education institutions should consider developing systematic support structures, including orientation programs, learning analytics dashboards, and learner support services, to strengthen students' self-, co-, and socially shared regulation capacities.

Limitations and Future Directions

Despite the contributions of our findings, there are still limitations that should be acknowledged when interpreting the findings of this review. First, although we attempted to identify as many studies on LP as possible through multiple database searches and hand search

processes (see Figure 1), it is unlikely that we identified and included all related literature in this review. In particular, varying terminology (e.g., language learner presence, self-regulated learning presence) and inconsistencies in indexing practices may have resulted in the potential omission of relevant work. Every effort was made to capture the population of LP studies including a robust search, screening, and referral from existing reviews.

Although approximately 27 studies were identified as quantitative designs, only 18 were ultimately included in the meta-analysis due to insufficient statistical information or methodological constraints. This reduced sample size potentially limits the generalizability of the meta-analytic findings. The exclusion of quantitatively designed studies highlights issues in current reporting practices in ODDE literature. Future studies should provide more complete statistical information (e.g., comprehensive descriptive statistics and effect size data, especially sample sizes) and ensure greater methodological transparency to support cumulative knowledge building on LP.

Regarding our findings of RQ2, the strong associations between LP and the three CoI presences suggest that online instructors should intentionally design learning environments that support learners' learning regulations. Rather than assuming that students will naturally self-regulate in online settings, instructional activities should incorporate opportunities for goal setting, progress monitoring, reflection, and feedback. Given recent conceptualizations of LP that extend beyond individual level of learning regulation, collaborative activities such as peer feedback, group planning, and shared goal settings and reflection may also help foster co-regulation and socially shared regulation of learning. Designing online courses that explicitly support these regulatory processes may strengthen learners' engagement and contribute to more meaningful inquiry within the CoI framework.

A quality evaluation across included studies was conducted. Nearly half of the reviewed studies employed qualitative (n = 13) or mixed methods designs (n = 11). Given the heterogeneity in research paradigms across the LP-related scholarship, applying a unified quality appraisal framework was methodologically challenging. Consequently, the quality evaluation (e.g., STROBE Checklist, Von Elm et al., 2007) was limited and may not fully capture methodological rigor across diverse study types. In order to capture literature as diverse as CoI, quality frameworks robust to the full range of methodologies are needed.

Finally, the conceptualization of LP remains heterogeneous and, at the same time, insufficiently specified. Future research should examine whether the strength and pattern of LP's associations with other CoI presences vary across different conceptualizations of LP (e.g., self-regulation, co-regulation, socially shared regulation, self-efficacy, or self-directed learning). Such work would help determine whether these conceptual distinctions reflect substantively different constructs or alternative operationalizations of a common underlying phenomenon. Relatedly, to enable more precise theoretical advancement, future research would benefit from critical analyses grounded in clearly defined and rigorously applied conceptual criteria. For example, Chang et al. (2025) restricted their review to quantitative studies to enhance conceptual consistency, whereas the present study adopted a more inclusive approach to capture the broader range of LP conceptualizations. This represents a trade-off: while such inclusiveness allowed for greater coverage in literature identification, it also led to lower conceptual coherence. Other approaches include establishing more stringent inclusion criteria (e.g., studies in synchronous

online learning settings) and conducting theory-informed critical analyses (e.g., synchronous instructor presence in Krebs-Lazendic et al., 2026) to clarify construct boundaries and strengthen conceptual definitions. Both approaches are needed in order to move the ODDE literature as it intersections with CoI forward.

Conclusions

Since Shea and Bidjerano (2010) proposed *learning presence* to address the under-specification of learners' regulatory agency in the original CoI framework, research on LP has gained increasing scholarly attention and invoked debates about its distinctive role within the CoI framework (Garrison & Akyol, 2013). This review synthesized 51 empirical studies and conducted meta-analyses of 18 quantitative studies (k = 103 effect sizes) to clarify how LP has been conceptualized, measured, and empirically related to key CoI constructs. LP research is largely grounded in regulation of learning, particularly at the individual levels; however, its operationalization has not yet achieved conceptual consensus in literature. Meta-analytic findings indicate strong, positive associations between LP and TP, SP, and CP, as well as satisfaction and learning outcomes, with stronger links to perceived than to objective learning. However, the substantial overlap between LP and existing presences, particularly CP (r = .66, p < .001), raises questions about whether LP constitutes a distinct presence or functions as a metacognitive mechanism within CP. These findings highlight the need for clearer construct delineation and conceptual refinement (Shea et al., 2022). Furthermore, to build a more robust knowledge base on LP and CoI research, greater methodological rigor and more consistent reporting practices are needed. Improving reporting quality as well as the volume of literature will allow scholars to find ways to assure LP is making an important and unique contribution to CoI outcomes, either through critical/conceptual review or meta-analytics structural equation modeling.

Ethics board approval or exemption statement

This study is a systematic and meta-analytic review based on previously published studies and did not involve human subjects. Therefore, Institutional Review Board approval was not required.

Conflicts of interest statement

The authors declare no conflicts of interest.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or non-profit sectors.

Acknowledgement of how AI was /was not used in the paper/study

This paper was reviewed and refined with the assistance of generative AI tool (ChatGPT 5.2, www.chat.openai.com), complementing the human editorial process.

References

Alotebi, H. R. (2017). *Student satisfaction and achievement with blended learning at the first Saudi Electronic University* (Publication No. 28186337) [Doctoral dissertation, Cleveland State University]. ProQuest Dissertations & Theses Global.

Alsayer, A. A., Templin, J., Niileksela, C., & Frey, B. B. (2025). Examining the structure of the revised community of inquiry framework: a multi-level approach. *Education and Information Technologies*, *30*(5), 6785-6807. https://doi.org/10.1007/s10639-024-13090-3

Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing context. *Journal of Asynchronous Learning Networks*, *5*(2). https://doi.org/10.24059/olj.v5i2.1875

Annand, D. (2011). Social presence within the community of inquiry framework. *International Review of Research in Open and Distributed Learning*, *12*(5), 40-56. https://doi.org/10.19173/irrodl.v12i5.924

Artino Jr, A. R., & Stephens, J. M. (2009). Academic motivation and self-regulation: A comparative analysis of undergraduate and graduate students learning online. *The Internet and Higher Education*, *12*(3-4), 146-151. https://doi.org/10.1016/j.iheduc.2009.02.001

Armah, J. K., Bervell, B., & Bonsu, N. O. (2023). Modelling the role of learner presence within the community of inquiry framework to determine online course satisfaction in distance education. *Heliyon*, *9*(5), e15803. https://doi.org/10.1016/j.heliyon.2023.e15803

Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education*, *12*(1), 1-6. https://doi.org/10.1016/j.iheduc.2008.10.005

Blaine, A. M. (2019). Interaction and presence in the virtual classroom: An analysis of the perceptions of students and teachers in online and blended Advanced Placement courses. *Computers & Education*, *132*, 31-43. https://doi.org/10.1016/j.compedu.2019.01.004

Borenstein, M. (2023). How to understand and report heterogeneity in a meta-analysis: The difference between I-squared and prediction intervals. *Integrative Medicine Research*, *12*(4), 101014. https://doi.org/10.1016/j.imr.2023.101014

Bosch, C., & Laubscher, D. J. (2022). Promoting self-directed learning as learning presence through cooperative blended learning. *International Journal of Learning, Teaching and Educational Research*, *21*(9), 17-34. https://doi.org/10.26803/ijlter.21.9.2

Caskurlu, S., Maeda, Y., Richardson, J. C., & Lv, J. (2020). A meta-analysis addressing the relationship between teaching presence and students' satisfaction and learning. *Computers & Education*, *157*, 103966. https://doi.org/10.1016/j.compedu.2020.103966

Chandler, K. (2021). *Students' experiences of synchronous online tuition in health and social care* [Doctoral dissertation, The Open University (United Kingdom)].

Chang, C. C., Yen, W. H., Liu, S. C., & Kuo, C. T. (2025). Learning presence in the revised community of inquiry framework: A systematic review of empirical research (2010–2024). *The Internet and Higher Education*, 101058. https://doi.org/10.1016/j.iheduc.2025.101058

Chen, Z., Jalaludin, N. A., & Rasul, M. S. (2025). Exploring the mediating effect of academic self-efficacy and self-regulated learning in the relationship between teaching, social presences and cognitive presence. *Journal of Pedagogical Research*, *9*(4), 157-172. https://doi.org/10.33902/JPR.202533616

Cho, M. H., Kim, Y., & Choi, D. (2017). The effect of self-regulated learning on college students' perceptions of community of inquiry and affective outcomes in online learning. *The Internet and Higher Education*, 34, 10-17. https://doi.org/10.1016/j.iheduc.2017.04.001

Cleveland-Innes, M., & Campbell, P. (2012). Emotional presence, learning, and the online learning environment. *The International Review of Research in Open and Distributed Learning*, *13*(4), 269-292. https://doi.org/10.19173/irrodl.v13i4.1234

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, 8, 243–253. https://doi.org/10.1037/1082-989X.8.3.243

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research*, *24*(3), 315-318. https://www.jstor.org/stable/3151642

DiDonato, N. C. (2013). Effective self-and co-regulation in collaborative learning groups: An analysis of how students regulate problem solving of authentic interdisciplinary tasks. *Instructional Science, 41*(1), 25-47. https://doi.org/10.1007/s11251-012-9206-9

Dunmoye, I. D., Martin, J. P., Brown, J. S., Lu, L. Z., Hunsu, N., & May, D. (2025). Examining the mediating role of learning presence in the predictive relationships between social and teaching presences and cognitive presence in collaborative virtual reality learning environments. *Computers & Education: X Reality*, *6*, 100101. https://doi.org/10.1016/j.cexr.2025.100101

ElSayad, G. (2023). Can learning presence be the fourth community of inquiry presence? Examining the extended community of inquiry framework in blended learning using confirmatory factor analysis. *Education and Information Technologies, 28*(6), 7291-7316. https://doi.org/10.1007/s10639-022-11480-z

Fayram, J. (2017). *The nature and role of social presence in audiographic, synchronous online language learning contexts* [Doctoral dissertation, The Open University (United Kingdom)]

Garrison, D. (2011). *E-learning in the 21st century: a framework for research and practice*. 2nd ed. New York: Routledge.

Garrison, R. (2022). Shared metacognition in a community of inquiry. *Online Learning, 26*(1), 6-18. https://doi.org/10.24059/olj.v26i1.3023

Garrison, D. R., & Akyol, Z. (2013). The community of inquiry theoretical framework. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 104–120). Routledge.

Garrison, D. R., & Akyol, Z. (2015). Toward the development of a metacognition construct for communities of inquiry. *The Internet and Higher Education*, *24*, 66-71. https://doi.org/10.1016/j.iheduc.2014.10.001

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2-3), 87-105. https://doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R., Anderson, T. & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7-23. https://doi.org/10.1080/08923640109527071

Garrison, D. R., & Arbaugh, J. B. (2007). Researching the Community of Inquiry framework: Review, issues and future directions. *The Internet and Higher Education*, 10(3), 157-172. https://doi.org/10.1016/j.iheduc.2007.04.001

Garrison, D. R., Cleveland-Innes, M., & Fung, T. S. (2010). Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. *The Internet and Higher Education*, *13*(1-2), 31-36. https://doi.org/10.1016/j.iheduc.2009.10.002

Günbatar, M. S. (2021). Evaluation of the problem based flipped classroom instruction process in the framework of Community of Inquiry. *Journal of Problem Based Learning in Higher Education, 9*(2), 46-68. https://doi.org/10.5278/ojs.jpblhe.v9i2.6690

Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*(2), 240-264. https://doi.org/10.1177/016146811111300204

Hayes, S., Smith, S. U., & Shea, P. (2015). Expanding learning presence to account for the direction of regulative intent: self-, co-and shared regulation in online learning. *Online Learning, 19*(3), 15-31. https://eric.ed.gov/?id=EJ1067475

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Honig, C. A., & Salmon, D. (2021). Learner presence matters: A learner-centered exploration into the community of Inquiry Framework. *Online Learning, 25*(2), 95-119. https://doi.org/10.24059/olj.v25i2.2237

Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods*, *25*(4), 430–455. https://doi.org/10.1037/met0000245

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, *48*(1), 25-39. https://doi.org/10.1080/00461520.2012.748006

Järvenoja, H., Järvelä, S., & Malmberg, J. (2020). Supporting groups' emotion and motivation regulation during collaborative learning. *Learning and Instruction*, *70*, 101090. https://doi.org/10.1016/j.learninstruc.2017.11.004

Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *The Internet and Higher Education*, *22*, 37-50. https://doi.org/10.1016/j.iheduc.2014.04.003

Krebs-Lazendic, L., Dello Stritto, M. E., & Park, E. (2026). A critical review of research on synchronous instructor presence in online learning: Elusive evidence and mixed results. *Quarterly Review of Distance Education*, 1–19. https://doi.org/10.1108/QRDE-09-2025-0016

Kurek, M., & Müller-Hartmann, A. (2019). The formative role of teaching presence in blended Virtual Exchange. *Language Learning & Technology, 23*(3), 52–73. http://hdl.handle.net/10125/44696

Ma, Z., Wang, J., Wang, Q., Kong, L., Wu, Y., & Yang, H. (2017). Verifying causal relationships among the presences of the community of inquiry framework in the Chinese context. *The International Review of Research in Open and Distance Learning, 18*(6), 213–230. https://doi.org/10.19173/irrodl.v18i6.3197

Martin, F., Wu, T., Wan, L., & Xie, K. (2022). A meta-analysis on the community of inquiry presences and learning outcomes in online and blended learning environments. *Online Learning*, *26*(1), 325-359. https://doi.org/10.24059/olj.v26i1.2604

Moran, P. A. P. (1948). Rank correlation and product-moment correlation. *Biometrika, 35*(1/2), 203–206. https://doi.org/10.2307/233264

Na, C., Lee, D., Moon, J., & Shin, Y. (2024a). Modeling undergraduate students' learning dynamics between self-regulated learning patterns and community of inquiry. *Education and Information Technologies*, *29*(15), 19621-19648. https://doi.org/10.1007/s10639-024-12527-z

Na, C., Jeong, S., Clarke-Midura, J., & Shin, Y. (2024b). Linking self-regulated learning to community of inquiry in online undergraduate courses: A person-centered approach.

Educational Technology Research and Development, *72*(6), 2895-2920. https://doi.org/10.1007/s11423-024-10380-y

Näykki, P., Järvenoja, H., Järvelä, S., & Kirschner, P. (2017). Monitoring makes a difference: Quality and temporal variation in teacher education students' collaborative learning. *Scandinavian Journal of Educational Research*, *61*(1), 31-46. https://doi.org/10.1080/00313831.2015.1066440

Li, N., & Lau, K. L. (2025). A two-staged SEM-ANN approach to predict learning presence in online foreign language education: The role of teaching presence and online interaction. *The Internet and Higher Education, 66*, 101012. https://doi.org/10.1016/j.iheduc.2025.101012

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinnrdd, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *71*. https://doi.org/10.1136/bmj.n71

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801-813. https://doi.org/10.1177/0013164493053003024

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879 –903. https://doi.org/10.1037/0021-9010.88.5.879

Pustejovsky, J. E. (2026). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections* (R package version 0.6.2) [Computer software]. CRAN. https://CRAN.R-project.org/package=clubSandwich

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23*(3), 425-438. https://doi.org/10.1007/s11121-021-01246-3

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.4.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Richardson, J., Fiock, H., Olesova, L., & Stenbom, S. (2024). Evidence-based practices for developing a community of inquiry. In T. Martindale, T. B. Amankwatia, L.D. Cifuentes, & A.A. Piña (Eds.), *Handbook of Research in Online Learning* (pp. 276-306). Leiden, The Netherlands: Koninklijke Brill NV. https://doi.org/10.1163/9789004702813_012

Richardson, J. C., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior*, *71*, 402-417. https://doi.org/10.1016/j.chb.2017.02.001

Shea, P., & Bidjerano, T. (2008). Measures of quality in online education: An investigation of the community of inquiry model and the net generation. *Journal of Educational Computing Research*, *39*(4), 339-361. https://doi.org/10.2190/EC.39.4.b

Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education*, *55*(4), 1721–1731. https://doi.org/10.1016/j.compedu.2010.07.017

Shea, P., & Bidjerano, T. (2012). Learning presence as a moderator in the community of inquiry model. *Computers & Education*, *59*(2), 316-326. https://doi.org/10.1016/j.compedu.2012.01.011

Shea, P., Hayes, S., Uzuner-Smith, S., Gozza-Cohen, M., Vickers, J., & Bidjerano, T. (2014). Reconceptualizing the community of inquiry framework: An exploratory analysis. *The Internet and Higher Education*, *23*, 9–17. https://doi.org/10.1016/j.iheduc.2014.05.002

Shea, P., Richardson, J., & Swan, K. (2022). Building bridges to advance the community of inquiry framework for online learning. *Educational Psychologist*, *57*(3), 148-161. https://doi.org/10.1080/00461520.2022.2089989

Shu, F., Liu, Q., Wang, Q., Tu, F., Li, H., & Zhao, C. (2025). Examining factors influencing online adult learners' satisfaction with blended synchronous learning. *Education and Information Technologies*, *30*(8), 10257-10280. https://doi.org/10.1007/s10639-024-13265-y

Swan, K., Garrison, D. R., & Richardson, J. C. (2009). A constructivist approach to online learning: The Community of Inquiry framework. In C. R. Payne (Ed.), *Information technology and constructivism in higher education: Progressive learning frameworks* (pp. 43–57). IGI Global.

Tunceren, L. L. (2017). *Community college second language students' perspectives of online learning: A phenomenological case study* [Doctoral dissertation, University of South Florida].

Usmani, S. (2021). COVID-19 pandemic and blended learning: A quantitative assessment of Revised Community of Inquiry (RCoI) framework. *Journal of Education and Educational Development*, *8*(2), 338-358. https://files.eric.ed.gov/fulltext/EJ1334964.pdf

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03

Volchok, E. (2018). Toward a predictive model of community college student success in blended classes. *Community College Journal of Research and Practice*, *42*(4), 274-288. https://doi.org/10.1080/10668926.2017.1287607

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*, *370*(9596), 1453-1457. https://doi.org/10.1016/S0140-6736(07)61602-X

Wertz, R. E. (2022). Learning presence within the Community of Inquiry framework: An alternative measurement survey for a four-factor model. *The Internet and Higher Education, 52*, 100832. https://doi.org/10.1016/j.iheduc.2021.100832

Xu, S. (2019). *Investigating an online community of self-directed language learners at The Mixxer* [Doctoral dissertation, Indiana University].

Xu, Z., Zhao, Y., Liew, J., Zhou, X., & Kogut, A. (2023). Synthesizing research evidence on self-regulated learning and academic achievement in online and blended learning environments: A scoping review. *Educational Research Review*, *39*, 100510. https://doi.org/10.1016/j.edurev.2023.100510

Yang, D., Wang, S., & Zhao, L. (2025). Relationships between cognitive presence and students' learning outcomes in online higher education: a meta-analysis. *Distance Education, 46*(4), 669–690. https://doi.org/10.1080/01587919.2025.2450645

Zhang, Y. (2018). *The effects of community of inquiry, learning presence, and mentor presence on learning outcomes: A revised community of inquiry model for K–12 online learners* [Doctoral dissertation, Michigan State University].

Zhang, Y., Huang, J., Hussain, S., & Dong, Y. (2023). Investigating the impact of the community of inquiry presence on online learning satisfaction: A Chinese college student perspective. *Psychology Research and Behavior Management*, *16*, 1883-1896. https://doi.org/10.2147/PRBM.S409229

Zhang, W., & Zhu, C. (2023). Institutional presence: Toward a further developed Community of Inquiry model integrating institutional functions in online and blended learning environment. *Frontiers in Psychology*, *14*, 1132204. https://doi.org/10.3389/fpsyg.2023.1132204

Zheng, B., Ganotice, F. A., Lin, C. H., & Tipoe, G. L. (2023). From self-regulation to co-regulation: refining learning presence in a community of inquiry in interprofessional education. *Medical Education Online*, *28*(1), 2217549. https://doi.org/10.1080/10872981.2023.2217549

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183. https://doi.org/10.3102/0002831207312909

Zimmerman, B. J., & Schunk, D. H. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 1–37). Lawrence Erlbaum Associates.

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5822
https://doi.org/10.24059/olj.5822

Research quality moderates the relationship between student motivation and academic performance in online and blended courses: A meta-analysis

Andrew Walker https://orcid.org/0000-0001-8693-9921
Kaelyn N. Loveless https://orcid.org/0009-0006-8099-3824
Utah State University
Naomi R. Aguiar https://orcid.org/0000-0002-3139-9654
Oregon State University
Yu-Chun Kuo https://orcid.org/0000-0002-7147-9798
Rowan University
Raechel N. Soicher https://orcid.org/0000-0002-2142-625X
Massachusetts Institute of Technology
Jessica Resig https://orcid.org/0000-0003-2706-8446
Applied Research Institute

Abstract

Prior meta-analytic work suggests the relationship between student motivation and academic performance among in-person learners is strong; however, evidence is weaker for college students learning in online, digital, and distance education (ODDE). This could be due to a variety of factors, including the rigor and quality of existing research, which has not been systematically accounted for in (ODDE) motivation reviews since 2008. The primary objective of this expanded meta-analytic review was to examine measurement quality and methodological quality first as moderators and then as adjustments in exploring the relationships between motivational constructs or synchronicity and academic performance in ODDE courses. Informed by DISCAR and adapted from PRISMA, we searched nine databases of studies published between 1990 and 2025 and conducted a secondary search of references, ultimately analyzing 64 studies and 117 outcomes that met inclusion and exclusion criteria. Both measurement quality and methodological quality were moderators of the relationship between student motivation and academic performance in online and blended courses. Adjusting for significant quality differences, overall, modest correlation values ($r = 0.14$) overestimated adjusted relationships ($r_{adj} = 0.10$). Despite challenges with some instances of missing data and low interrater reliability, these results indicate that motivational factors related to student success may be overstated when research quality is taken into account. Variations were found in both motivational constructs as well as synchronicity, suggesting that what is measured and how learners interact make a difference. These results have implications for future ODDE research, pedagogical designers, and policymakers.

Keywords: meta-analysis, online learning, blended learning, motivation, research quality

Corresponding author: Dr. Andrew Walker, andy.walker@usu.edu
This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 The Author(s).

Introduction

In higher education, online and blended course modalities are of increasing importance to institutional success. Online and blended course offerings provide the opportunity to expand access to higher education beyond the "traditional" in-person learner. This expanded access to higher education not only benefits diverse learners everywhere; it benefits academic institutions, many of which are facing emergent enrollment challenges (National Center for Educational Statistics, 2024). Although these and other benefits of online and hybrid modalities are well documented (Lucas & Vicente, 2023), so too are its challenges (Kebritchi et al., 2017; Lucas & Vicente, 2023) and complexities, including variations in synchronous and asynchronous instruction (Johnson et al., 2022). One particular concern is student academic performance, as it is associated with retention and graduation rates (Gardner et al., 2024), which tend to be lower in online programs (Bawa, 2016).

Student motivation is currently of high interest in online and hybrid courses. In addition to a large volume of research, it plays a prominent role in conceptual frameworks like Community of Inquiry (CoI) for online settings (Shea et al. 2022). This focus coincides with investigations of in-person learners, where the relationship between student motivation and academic performance is strong (Honicke et al., 2020; Richardson et al., 2012). These findings with in-person learners have contributed to the widely held belief that motivation is one of the most important predictors of college students' academic performance (Honicke et al., 2020). However, our recent meta-analytic review indicates that the relationship between student motivation and academic performance is more modest for online distance and digital education (ODDE), and accounts for little of the variance in academic performance (Walker, Aguiar, Soicher, Kuo, & Resig, 2024). This appears true across a broad range of motivational constructs, including self-efficacy, attributional style, achievement goal orientation, self-determination and task value.

It is possible that these modest relationships are partially explained by a range of moderating factors. Methodological quality clearly matters, both to the academic community and as context for findings. Early reviews of motivation and learning express concerns regarding study quality (Bernard et al., 2004; Multon et al., 1991; Perkowski, 2013; Storrings, 2005). Reductions in study quality are associated with reductions in effect sizes for in-person learning (Macnamara et al., 2023) and for ODDE learners (Bernard et al., 2004). Measurement quality also matters in reviews of motivation. Some reviews code for and include reliability of measurement in their analyses (Richardson et al., 2012) but not validity. Others, make strong contributions to validity in specific constructs like achievement goals (Huang, 2012) but do not address reliability. Investigations of ODDE literature that address study quality only cover literature to 2006 (Bernard et al., 2009) and measurement quality only covers to 2008 (Huang, 2012). Of the approaches to addressing quality (Fromm et al., 2025), existing reviews have employed selection criteria (Storrings, 2005), and novel coding (Multon et al., 1991). None have used an establish quality rubric (Katrak et al., 2004; Rosella et al., 2016a; Sterne et al., 2019; Whiting et al., 2016) or comprehensively coded for methodological quality alongside measurement reliability and validity. Finally, asynchronous and synchronous interactions clearly matter in this literature and are associated with variations in effect sizes (Bernard et al., 2004) and persist when coding for a mixture of the two (Bernard et al., 2009). Like study and

measurement quality, reviews of the ODDE literature have not coded for synchronicity for almost two decades of primary research. A recent critical review (Krebs-Lazendic et al., 2026) indicated that synchronicity within online courses may be an important factor in the focused area of instructor presence and academic performance, warranting replication newer results in broader ODDE literature.

The goal of the current meta-analytic review was to expand the previous work of Bernard et al. (2009) and Richardson et al. (2012) by comprehensively examining methodological quality and measurement quality as moderators of the relationship between student motivation and academic performance in online and blended college-level courses. Importantly, not only was it crucial to assess the impact of methodological and measurement quality on the relationship between motivation and academic performance in ODDE research, it was important to replicate and extend adjustments for measurement reliability (Richardson et al., 2012) to include adjustments for measurement validity and methodological quality in the examination of moderators for motivational constructs and synchronicity.

Existing Reviews

In an effort to be comprehensive, we examined 19 historical, scoping, systematic or meta-analytic reviews overlapping with post-secondary education. About half drew from in-person research (Howard et al., 2021; Hulleman et al., 2010; Lazowski & Hulleman, 2016; Macnamara et al., 2023; Multon et al., 1991; Richardson et al., 2012; Senko & Dawson, 2017; Theobald, 2021) and the rest summarized the ODDE literature (Bernard et al., 2004, 2009; Huang, 2012; Lee, 2017; Mahmud et al., 2022; Martin et al, 2022; Perkowski, 2013; Storrings, 2005; Walker et al., 2024; L. Wang, 2022; Zhao et al., 2005). Variation between in-person and ODDE learners across reviews is stark. The relationship between variations of self-efficacy and academic performance in Richardson et al. (r = 0.31 to 0.59) are clearly stronger than the aggregate self-efficacy for ODDE contexts, Walker et al. (2024) (r = 0.14). Further exploration of moderators is clearly needed.

Not only were meta-analyses addressing primary research quality sporadic (Bernard et al., 2004, 2009; Macnamara et al., 2023; Multon et al., 1991; Storrings, 2005), but none of them used an establish quality coding rubric. Reviews addressing methodological quality ranged from simple reporting (Multon et al., 1991) to using publication in peer reviewed journals as a form of quality control (Richardson et al., 2012), to analyses indicating that variations in quality do impact effect sizes (Bernard et al., 2004; Macnamara et al., 2023). For example, Macnamara et al. (2023) found that significant effect sizes became non-significant when focusing on acceptable or best available study quality. Existing reviews note challenges of insufficient data reporting (Multon et al., 1991) including methodological quality, making recommendations to stakeholders challenging (Bernard et al., 2004). Other quality concerns range from methodological rigor (Storrings, 2005) to convenience samples and treatment variations (Perkowski, 2013). Despite clear concerns, the most recent review incorporating quality in the analysis of motivation for ODDE learners (Bernard et al., 2009) only covers studies as recent as 2006. All of these reviews

use novel coding for methodological features, bypassing modern quality rubrics (Rosella et al. 2016).

Measurement quality is not routinely addressed within motivation reviews. Some use measurement rigor as a selection criterion (Howard et al., 2021; Storrings, 2005). Others, working in the niche area of Achievement Goals report reliabilities across scales (Hulleman et al., 2010). Rare reviews (Richardson et al., 2012) report and adjust for internal construct reliability of both predictors and reliability of learning outcomes. Even studies that address reliability in some fashion fail to address validity (Hulleman et al., 2010; Richardson et al., 2012). Among ODDE reviews, the only analysis of measurement focused on discriminant and criterion validity of achievement goals as moderators (Huang et al. 2012) and only covered studies through 2008. For discriminant validity, Huang et al. (2012) found four factor models of achievement goals had the highest correlation values with learning. A meta-regression for criterion validity determined that 6% of the variance in academic achievement could be accounted for when using all four sub-constructs. Clearly, there is a need to examine literature at the intersection of motivation and ODDE using an established methodological quality rubric and comprehensively addressing measurement reliability and validity for a wide range of motivational constructs, covering two decades of recent literature.

Online and Blended Synchronicity

Much like selecting a quality coding rubric for educational research, definitions of synchronous, asynchronous and mixed interactions can be challenging. Some definitions center on interaction in combination with group size. At one extreme, online classes are synchronous when they scale the in-person classroom to multiple sites. Asynchronous interactions are either individual or small group based and utilizing technologies like email that allow for separation of time (Bernard et al., 2004). Refined definitions acknowledge directionality as a difference—reading an article is a one-way transfer and fundamentally asynchronous, contrasted with synchronous exchanges like the bi-directional transfer of a chat or Zoom meeting (Bernard et al., 2009). Our definition borrows from the consensus-built work of Johnson et al. (2022), where synchronous learning requires fixed, real-time exchanges, whether in-person or virtually, and asynchronous learning provides students with flexibility to access learning experiences when and where it works for them.

Based on both past and recent reviews (e.g. Bernard et al. 2004; Krebs-Lazendic et al. 2026), course synchronicity may be as important as the nature of the motivational construct. Bernard et al. (2004) examined differences between distance education (DE) and classroom instruction (CI) across affective, performance, and completion outcomes. Splitting results out by asynchronous and synchronous contexts, authors found variations in effect size for academic performance, as well as affective and course completion outcomes. Expanding their prior coding to include mixed synchronicity, Bernard et al. (2009) explored variations in online learning coding for type of interaction (e.g., student-student) and mean difference based on strength of interaction (minimal, moderate, high). Although they did not find significant differences in effect sizes between synchronous and asynchronous courses, they did find that the strength of

interactions in asynchronous courses predicted academic performance; this was not the case for mixed or synchronous courses, suggesting that the strength of interactions is of particular importance when students engage asynchronously with course content. Note that both of these meta-analyses explore mean differences—either between online and in-person settings (Bernard et al., 2004) or between interaction strength variations in online classes (Bernard et al., 2009). Given that other research examining the effects of synchronicity on academic performance remains limited (Krebs-Lazendic et al., 2026), extending early mean difference reviews to explore the moderating effect of synchronicity in motivation relationships is needed.

Motivational Constructs

Building from social cognitive theory to the constructs and subconstructs of student motivation is another review challenge. Motivation is a multidimensional latent variable that consists of several constructs including but not limited to self-efficacy, achievement goal orientation, self-determination, task value and attributional style (see Table 1 for well-established definitions).

Table 1

Select motivational constructs and their definitions

Construct	Definition
Self-efficacy	A person's belief in their ability to organize and execute the tasks required to achieve a desired level of performance, derived from Bandura's Social Cognitive Theory (1986, 2006) and thought to be domain specific.
Achievement Goal Orientation	"A future focused cognitive representation" (Hulleman et al. 2010, p. 423) that guides behavior either towards or away from a specific outcome; measured across dimensions of mastery vs. performance goals and approach vs. avoidance goals (Bardach et al., 2020; Elliot & McGregor, 2001).
Self-Determination	Intrinsic motivation involves pursuing a goal because of its perceived internal value (e.g., for interest or fun); extrinsic motivation involves pursuing a goal that has outcomes separate from the activity (e.g., an external reward, Ryan & Deci, 2000, 2020)
Task Value	Operationalized as attainment value (importance of doing well on a task), intrinsic value of the task, utility value (how a task relates to current and future goals), and cost (Eccles & Wigfield, 2020).
Attributional Style	A perception of successes and failures, made up of three components: (1) locus of causality (internal or external causes), (2) stability (constant or variable), and (3) controllability (Weiner, 2010).

Theoretical Framework for Motivation

There are several intersecting theories in ODDE and motivation literature, such as the Community of Inquiry Framework (Garrison, 1999, 2007), which incorporates self-regulated learning and self-efficacy in learner presence (Chang et al., 2026). The primary interest for this

review is motivation. We adopted the review stance of Lazowski and Hulleman (2016) that motivation is complex, multi-dimensional, and in the context of educational research shares common roots in social cognitive theory (Bandura, 1986, 2006). Some motivational constructs, such as self-efficacy are featured in social cognitive theory. Others, such as achievement goal theory lean heavily on student perceptions and thus a social-cognitive orientation to illuminate goal structures (Ames & Archer, 1988).

Motivation, Measurement, and Synchronicity

In the extant literature specific to ODDE, much of the research has focused on self-efficacy as it relates to academic performance. For example, in our recent meta-analysis, nearly half of the outcomes included in the analyses examined the relationship between self-efficacy and academic performance (Walker et al., 2024). No other single motivational construct (e.g., task value, attributional style) was represented as frequently in empirical research. That could be a function of reviews with in-person learners (Richardson et al., 2012) where the strongest relationships are between self-efficacy scales and academic performance. Note that at least with Richardson et al. (2012), subconstructs of goal orientation accounted for more than half (n = 164) of the total outcomes.

It is possible that the focus on self-efficacy partially stems from the robust empirical testing of instruments that are available to measure it. The Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, 1991) is one such measure that has been rigorously vetted in the literature (de Araujo et al., 2023; Hilpert et al., 2013; Pintrich et al., 1993) and is frequently used. Although this instrument was designed to measure a range of motivational constructs, the Self-Efficacy for Learning and Performance subscale of the MSLQ is often used to measure motivation among online and blended learners (e.g., Alkış & Temizel, 2018; Goulão, 2014; Joo et al. 2013). However, the value of this tool in assessing motivation in online courses has been called into question in past research. Cho and Summers (2012) found that the original factor structure of the MSLQ was not supported in students taking asynchronous online courses. This could partially account for the modest overall relationship we found between self-efficacy and academic performance in our recent meta-analytic review (Walker et al., 2024). However, we also found some variation in the overall effect size, with stronger correlations among self-efficacy for learning in general, and weaker relationships with one of several technology contexts (e.g., online, computer, or LMS self-efficacy). This variation could also be due to the range of approaches in measuring self-efficacy in the literature. While some authors used the original self-efficacy subscale in the MSLQ to measure motivation (e.g., Alkış & Temizel, 2018; McPhaul-Moore, 2013; Savoji, 2013), others modified the original subscale (Hodges, 2008; Liu et al., 2019), and some authors developed their own measurement instruments (Kim, 2012; Martin et al., 2010; Warren et al., 2020). The example of self-efficacy alone highlights the potential for measurement decisions, including the synchronicity of classroom experiences to impact the correlation between motivation and academic performance in online and blended settings. Clearly, an updated review that not only explores the measurement space, but contextualizes findings related to motivational constructs and synchronicity after accounting for measurement quality and methodological quality is needed.

Assessing Study Quality

In meta-analytic reviews, measurement reporting quality is one of several variables best paired with an overall assessment of study quality. Appraisal of study quality, which often looks for reliability and validity of study methods, sampling, results reporting, replicability, and generalizability, is critical for evaluating the relevance and trustworthiness of a study (Rosella et al. 2016). Beyond ensuring a study and its results are trustworthy and relevant, prior meta-analytic work has shown that methodology quality accounts for considerable amounts of variation in effect size (Bernard et al., 2004). Unfortunately, extant reviews cite issues such as high attrition, large amounts of missing data, and differences in course content, duration, and delivery between treatment and control groups as contributors to low quality studies (e.g., Bernard et al. 2004; Huang, 2012; Perkowski, 2013). In a three-round meta-analysis study evaluating methodological quality control in intervention designs (no control, minimum standards, and best available standards), Macnamara et al. (2023) found that failure to account for study quality leads to inflated effect sizes. Additionally, of the 63 studies and 96 effect sizes included in their first-round analysis, only 6 studies and 10 effect sizes met more than 60% of the 10 relevant best practices criteria for third-round analysis. The systematic association between study quality and effect sizes, coupled with the scarcity of methodological quality in the literature, reinforces the importance of quality-controls in analyses meant to advise practitioners and policymakers on best practices.

Though study quality has been shown to impact effect sizes, few meta-analyses of academic achievement in ODDE have accounted for study quality in their analysis or discussion of limitations. Further, when quality is evaluated, authors utilize varying levels and types of criteria to support their assessment. For example, Bernard et al. (2004, 2009) coded for the type of publication, selection bias, treatment-control group equivalence, and attrition while Macnamara et al. (2023) coded for study group similarities, random assignment, a priori power analysis, clear design, transparent and complete results, and reporting of conflicts of interest. Another shortcoming of extant reviews is the exclusion of observational research from analysis. To the best of our knowledge, extant reviews that account for study quality focus primarily on experimental or interventional studies. The shortage of common frameworks or tools to conduct quality appraisals adapted to the wide range of educational research (Fromm et al., 2025) may be a contributor to the shortage of reviews that acknowledge study quality.

While less prominent in educational research, quality appraisal tools have a foothold in evidence-based medicine (EBM), where many tools were developed with a focus on risk of bias (Katrak et al., 2004). Such tools include AMSTAR, a measurement tool to assess systematic reviews (Shea et al. 2007); RoB 2, a revised Cochrane risk-of-bias tool for randomized trials (Sterne et al., 2019); and ROBIS, a tool to assess risk of bias in systematic reviews (Whiting et al., 2016), among others. Quality appraisal tools share at least one limiting factor in their potential usefulness: the inability to adapt to different settings, contexts, and study designs. Rosella et al. (2016) proposed the development of a meta-tool that extends to a wide range of applicability and flexibility, such as to observational studies, as well as maintain a rigorous and transparent process to carry out a detailed assessment. This tool, the Meta Quality Appraisal Tool

(MetaQAT), was developed and validated for the evaluation of a wide range of research applications and practices.

The MetaQAT framework encompasses four domains of quality, including relevancy, reliability, validity, and applicability. The assessment of relevancy focuses on evaluating the degree to which a topic is relevant to the issue that is being investigated. The assessment of reliability focuses on a thorough examination of the quality and transparency of the research, including a clear presentation of a study, the description of research methodology, results, and the inclusion of ethics procedures, etc. In terms of the assessment of validity, it examines both the internal validity and external validity to address the likelihood and magnitude of error or bias in the study; for example, the appropriateness of the study methodology for the scope of research, explicit conclusions, etc. The assessment of applicability evaluates whether the generated results can be applied within the scope of the field or discipline being studied. In the absence of a flexible quality appraisal tool in educational research, the current review adopted a modified MetaQAT to narrow the literature gap examining the influence of study quality on the magnitude of study effects and results in experimental, quasi-experimental, and observational research. In our study, we coded for methodological reliability, validity, and applicability. Relevancy, as outlined in MetaQAT, was addressed through our inclusion and exclusion criteria.

Research Questions

Based on existing reviews, it is clear the quality of primary ODDE research matters. Primary research suggests in-person and ODDE settings are not the same and can impact measurement models (Cho & Summers, 2012). Early review efforts to account for measurement reliability (Richardson et al. 2012), make contributions to measurement validity (Huang, 2012), and explore study quality (Bernard et al., 2004), but have not been replicated in modern ODDE reviews and have never been tackled comprehensively regardless of setting. It is also clear there are differences in the relationship between motivation and academic performance between in-person (Richardson et al., 2012) and ODDE learners (Walker et al., 2024). Regardless of setting, motivational construct is an important moderator in these relationships. In addition, there is a clear need to explore synchronicity of ODDE class interactions based on prior mean difference reviews (Bernard et al, 2004 & 2009) and a recent critical review of the literature (Krebs-Lazendic et al. 2026). Combining these approaches, a replication of Richardson et al. (2012) is needed that includes a comprehensive examination of and adjustments for study quality, an exploration of a broad range of motivational constructs, and an examination of synchronicity levels in online courses.

The current meta-analytic review addresses the following objectives and research questions in ODDE college-level courses:

Objective$_1$: Explore the impact of measurement and methodological quality on studies of motivation and learning performance.

RQ$_1$: To what extent does measurement reporting quality (validity and reliability) moderate the relationships between motivation and learning performance?

RQ_2: To what extent does methodological quality (via MetaQAT and attrition) of primary research moderate the relationships between motivation and learning performance?

$Objective_2$: Explore synchronicity and motivational constructs as moderators, using research quality adjusted effect sizes.

RQ_3: Adjusting for measurement reporting and methodological quality, to what extent does class synchronicity (*asynchronous*, *synchronous*, *mixed*) moderate the relationships between motivation and academic performance?

RQ_4: Adjusting for measurement reporting and methodological quality, to what extent does motivational construct (*self-efficacy, attributional style, achievement goal orientation, and self-determination and task value*) moderate the relationships between motivation and academic performance?

Methods

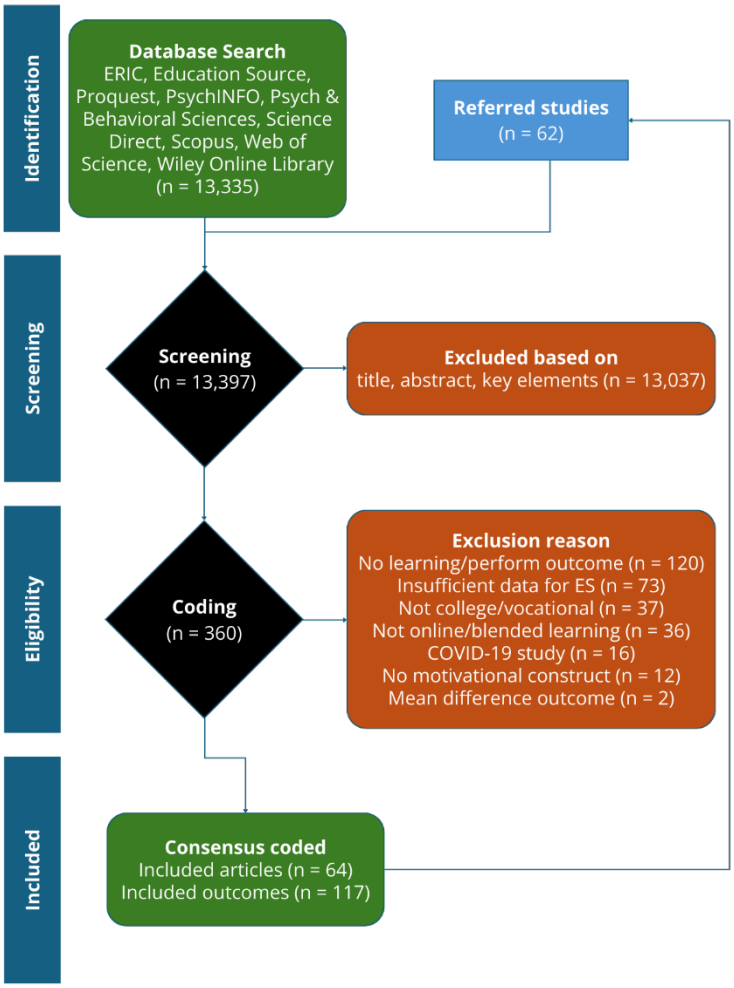
To address our research questions, we employed a systematic review and meta-analysis of ODDE literature examining student motivation and academic performance. Our study protocol was informed by the DISCAR (designing, including/excluding, screening, coding, analyzing, reporting) process (West & Martin, 2024), which identifies best practices for rigorous reviews. Our methods focus on all sections except for reporting, which is addressed in the results section.

Design

Our study and question design was driven by a desire to replicate and extend an exploration of motivational constructs (Richardson et al., 2012) and synchronicity (Bernard et al., 2004, 2009) after exploring and accounting for a broad range of motivational constructs (Walker et al., 2024) and a broad range of research quality in ODDE settings. We began with a scoping review to determine meaningful questions and feasibility of the study, assessing 19 reviews of post-secondary education. Through this scoping review, we identified three major gaps in post-secondary and ODDE literature: 1) inconsistencies in measurement reporting; 2) variation in methodological rigor assessment; and 3) an updated examination of synchronicity. In accordance with existing guidelines, an adapted PRISMA (Page et al., 2021) flow chart (Figure 1) was used for search transparency.

Figure 1

Literature search, referrals, and screening



Inclusion and Exclusion criteria

An early and crucial intersection of quality is selection of the range of methodologies included in the review (Lipsey & Wilson, 2001). Because of our interest in determining the impact of study quality, we cast a wide net for inclusion criteria: 1) Peer reviewed and gray literature; 2) Any study design (descriptive, pre-experimental, quasi-experimental, experimental) as long as it was empirical and quantitative; 3) Measured-attributional style, self-determination, self-efficacy, achievement goals, or task value; 4) Reported a relationship (*r*) between motivation and academic performance; 5) Published from 1990-2025; and 6) Used the English language. Studies were excluded if they: 1) Did not report a learning or performance outcome (most often due to it being a self-report such as anticipated grade); 2) Lacked enough information to calculate a correlation effect size; 3) Had either a mix of learners or education levels that did not align with the college/vocational sampling frame; 4) Were not online or blended learning

experiences; 5) Collected data during COVID-19, defined as any time during 2020; 6) Discussed but failed to report a motivational construct; or 7) Reported results rooted in mean differences instead of relationships.

Replicating the search from our most recent review (Walker et al., 2024), nine academic databases (Education Source, ERIC, PsychInfo, Psychological and Behavioral Sciences, ProQuest, Science Direct, Scopus, Web of Science, and Wiley Online Library) were searched for literature at the intersection of motivational constructs and variations of online teaching modes. Specifically, our search terms were:

(internality OR internal locus of control OR external locus of control OR globality OR stability OR attributional style OR academic locus of control OR optimism ORself-efficacy OR intrinsic motivation OR extrinsic motivation OR goal orientation OR achievement goal) AND (online learning OR distance learning OR web-based learning OR blended learning).

When available, limiters were added for English language, research manuscripts, post-secondary, applied related words, for all available fields encompassing 1990 through 2025. Our search process was inclusive of both peer reviewed journal articles and gray literature including conference papers, research reports and dissertations. Database results were supplemented by referrals from included studies, as well as systematic and narrative literature reviews.

Screening

After the data were extracted from databases, the citations were shared using Zotero and all search results were aggregated in Rayyan.ai (Ouzzani et al., 2016). Rayyan.ai was used to deduplicate results and track pre-pass screening assignments between a team of six screeners. Each study went through a screening of the title, abstract, and in cases of ambiguity about measuring motivation or reporting academic performance, a brief check of the full text. Screeners evaluated exclusion decisions based on key words related to the exclusion criteria, such as COVID; primary, secondary or graduate education; qualitative studies; self-reported performance; and MOOC studies. Generative AI tools were not used in the search, screening, or coding process, nor in the analyses and reporting.

Coding

Following the screening phase, a three-round coding sequence of first-pass coding, second-pass coding, and consensus coding was followed by a team of six researchers. Coding decisions were made according to a live code book which was developed and maintained by the research team. The code book began with deductive, a priori codes based on existing theory or frameworks. Occasionally, inductive coding was necessary when borderline or novel cases emerged in the literature. These borderline or novel cases were documented in the code book with consensus decisions to help ensure future decisions were made in a consistent manner.

Coding included (1) article citation information; (2) class interactions (*synchronous*, *asynchronous*, *mixed*, or *unclear*); (3) motivational construct; (4) sample size; (5) type of correlation; (6) correlation; (7) study attrition based on percentage of sampling frame; (8) quality

of measurement validity reports (*strong*, *attempt*, *none*); (9) quality of measurement reliability reports (*strong*, *attempt*, *none*); and (10) study quality (*good* or *bad/unclear*) utilizing the MetaQAT tool (Rosella et al. 2016).

Promising studies were first-pass coded and rejected if they failed to meet inclusion criteria. Studies that were accepted after a first-pass code were second-pass coded by another researcher. Independent coding was followed by a consensus round where disagreements were resolved including final decisions about inclusion, exclusion, and coding. The resolution of disagreements among coders was centered around finding different information to justify coding choices rather than differences in applying the coding guide. Data from the consensus coding was used in the final analysis.

Analysis

Analysis and reporting were supported through STATA 18.0. Base STATA functions for meta-analysis were supplemented by community contributed functions in MAIS and ROBMETA.

Preliminary Analyses

The following analyses are provided to promote transparency or because they were supportive of analyses that directly addressed research questions. Data extraction for all included studies is available in Appendix A.

Inter-coder reliability

Reliability scores (see Table 2) were not based on a sample, but a comprehensive analysis of all independent first- and second-pass coding. We used Krippendorff's Alpha (2004) for interval, nominal, and ordinal data and Cohen's Kappa (Cicchetti & Feinstein, 1990) for boolean variables (*good* or *bad/unclear*).

Table 2

Inter-rater reliability of all first- and second-pass coding

Variable	reliability	method	data type
Effect size (r)	0.96	alpha	interval
Measurement Reporting Quality			
Validity (*strong*, *attempt*, *none*)	0.90	alpha	ordinal
Reliability (*strong*, *attempt*, none)	0.84	alpha	ordinal
Study Quality (good, bad/unclear unless otherwise noted)			
Reliability: Clear presentation	0.70	kappa	boolean
Reliability: Clear methods and results	0.57	kappa	boolean
Reliability: Ethics procedures described	0.76	kappa	boolean

Variable	reliability	method	data type
Validity: Methods align with scope of research	0.67	kappa	boolean
Validity: Free from bias	0.61	kappa	boolean
Validity: Explicit and transparent conclusions	0.67	kappa	boolean
Applicable to setting (higher education)	0.47	kappa	boolean
Confident in findings	0.79	kappa	boolean
Attrition for sample (0 - 1.00; *missing*)	0.86	alpha	interval
Moderators			
Synchronicity (*asynch*, *synch*, *mixed*, *unclear*)	0.75	alpha	nominal
Motivational construct (thirteen options)	1.00	alpha	Nominal

Most of the Krippendorff's alpha values were at or near acceptable reliability of > .80 (Marzi et al., 2024). For Cohen's Kappa, inter-rater reliability was weak at < 0.59 to moderate at < 0.79 (McHugh, 2012).

Outlier detection

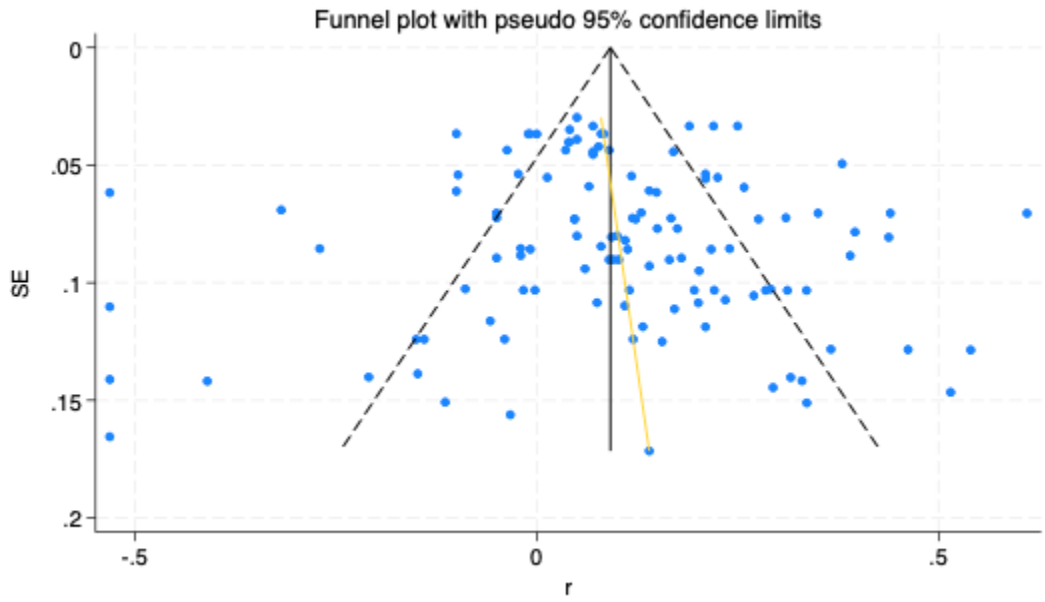
Three negative outliers were detected (Liu et al., 2019; Stephen et al., 2020; Tladi, 2017), prior to any adjustments. Rather than drop meaningful data, the values for each were trimmed to -.411, which is the lowest effect size that is not an outlier. Note these same three outliers were found before and after corrections for quality.

Publication bias

Publication bias using Egger's test was not significant $t(116) = 1.38$, $p = 0.17$ before trimming outliers. After trimming outliers as well as after adjustments for study quality, there was still no support for publication bias $t(116) = 0.79$, $p = 0.43$. A funnel plot of trimmed outlier and quality adjusted effect sizes by standard error (see Figure 2) also suggests there was no publication bias.

Figure 2

Visualization of standard error and adjusted effect size estimates



Multiple outcomes

With a total of 117 outcomes across 64 studies, there were clearly nested data. On average, there were 1.8 outcomes per study, with a range of a single outcome to 8 outcomes. Keeping these data was important to observe the full range of quality and also to keep information about moderators. For example, Bahçekapili and Karaman, (2020) examined the relationships between self-efficacy ($r = 0.14$), external ($r = 0.16$) and internal ($r = 0.03$) locus of control to academic performance. These not only represent different motivational constructs; they represent effect sizes that are both above and indistinguishable from 0. Rather than drop outcomes, we used Robust Variance Estimation (Hedges et al., 2010) to determine the degree to which the study of origin impacts effect size estimates of individual outcomes. Using adjusted effect sizes, the difference between assuming that no variation (*rho* = 0.00) was from the study of origin and all variation (*rho* = 0.99) was from the study of origin was minimal. Specifically, there was a 0.01 change in the overall effect size for the differences between no variation and all variation attributed to the study of origin; there was no change for quality adjusted overall effect size. As a result, outcomes were treated as if they were independent.

Academic Performance

Given the self-reported nature of most measures of motivation, our interest in academic performance was in scored performances including tests, quizzes, or class grades. We also included class completion and persistence (e.g., returning for additional coursework; Stephen, 2020). We excluded self-report of performances like grades (e.g., Radovan, 2011). We also

excluded cases where scored performances, such as exam scores were combined with self-assessment of learning (e.g., Zhang, 2020).

Rigor and Quality

The deep examination of study features forced by coding for quality did result in some adjustments to the prior data. Three outcomes, two technology and one online learning self-efficacy, from Torun (2020) were collapsed into a single technology self-efficacy outcome using a mini meta-analysis. One external locus of control outcome in Wang and Newlin (2000) was changed to internal locus of control after closer examination of the measure. These adjustments underscore the need for a comprehensive review of primary research quality.

Results

Where relevant, a random-effects (Borenstein et al., 2021) model was used, reflecting a focused (college and vocational) education level coupled with wide variations in study quality, motivational constructs, outcomes, and subject areas. Significance testing was set at the .05 level for all analyses. Finally, a Hunter and Schmidt (1990, 2007) correction was used to adjust for sample size for all analyses and to facilitate corrections for quality as detailed below. To compare moderators, ANOVA style Z-tests were used (Borenstein et al., 2021). Much like a primary study ANOVA, these offer direct comparisons between two groups using a ratio of the difference (r) and square root of the summed variance for any two groups. Similarly, Z-tests were used to determine if effect sizes were significantly different than $r = 0$ or no relationship between motivation and academic performance.

Objective 1: Measurement reporting and study quality

For the first two research questions, effect sizes were adjusted only for sample size. The primary purpose of these questions was to determine which, if any, variations in measurement reporting and methodological quality impact the correlation between motivation as a whole and academic performance.

RQ_1 Measurement reporting quality as a moderator

Most levels (*good*, *attempt*, *none*) in both measurement reliability and measurement validity (see Table 3) were associated with correlations significantly larger than zero. The lone exception was reliability (*attempt* $r = 0.02$). The expectation is that stronger reporting (*good*) should result in larger effect size relationships. For reliability, *good* reporting is associated with reduced measurement error and a closer approximation of the true (larger) effect size. For validity, stronger reporting should more precisely measure the motivational construct and be a closer approximation of the true (larger) effect size.

Table 3

Measurement reporting moderators

Measurement Reporting	*n*	*r*	*CI*[95%]	I^2	*Q*	*p*
Reliability						
good	89	0.14	[0.12, 0.16]	84.2%	557.43	0.01
attempt**	10	0.02	[-0.04, 0.08]	87.2%	70.36	0.53
none*	18	0.19	[0.14, 0.16]	81.1%	89.95	0.01
Validity						
good	34	0.14	[0.11, 0.17]	89.5%	315.29	0.01
attempt**	28	0.10	[0.07, 0.13]	84.5%	174.67	0.01
none	55	0.16	[0.13, 0.18]	77.1%	235.51	0.01
Overall	**117**	**0.14**	**[0.12, 0.15]**	**83.7%**	**711.46**	**0.01**

**Indicates a significant and stronger effect size (r) difference in need of reduction for quality*

***Indicates a significant and weaker effect size (r) difference in need of an increase for quality*

Pairwise comparisons revealed that measurement reliability reported as *none* was significantly higher than *attempt* ($r = 0.19$; $r_{diff} = 0.17$; $Z = 4.06$, $p < 0.01$) suggesting outcomes with no attempt to report measurement reliability may overestimate the effect size (*r*). Two other significant relationships were in an expected direction. Measurement reliability of *attempt* was lower than *good* ($r = 0.02$; $r_{diff} = 0.12$; $Z = 3.72$, $p < 0.01$). For measurement validity, *attempt* was significantly lower than *good* ($r = 0.10$; $r_{diff} = 0.04$; $Z = 2.17$, $p = 0.03$). Both reliability and validity reporting attempts indicate underestimates of the effect size (*r*). Results suggest that variations in measurement reliability and validity reporting do significantly moderate effect sizes (*r*) between motivation and academic performance. Under the assumption that high quality studies are transparent about reporting measurement reliability, outcomes with no reporting or *none* ($n = 18$) were adjusted down for later analyses. Similarly, reliability ($n = 10$) and validity ($n = 22$) reporting *attempts* were adjusted up for later analyses. Overall, these results indicate that measurement reporting quality plays a significant role in moderating the effect size relationships between motivation and academic performance, both over and under-estimating the relationship.

RQ_2 Methodological quality as a moderator

As with measurement reporting, effect sizes were adjusted only for sample size for analysis. These ratings were based on judgements about the methodological quality of each outcome. Out of nine coded MetaQAT plus attrition variables, three indicators of methodological quality acted as significant moderators (see Table 4). All of the significant differences paired *bad/unclear* methods, *bad/unclear* freedom from bias, and *missing* enough data to calculate attrition with stronger relationships.

Table 4

Methodological quality moderators

Methodological Quality	*n*	*r*	*CI*[95%]	I^2	*Q*	*p*
Reliable						
Clear study presentation	111	0.14	[0.12, 0.15]	83.3%	658.86	0.01
bad/unclear presentation	6	0.13	[0.02, 0.24]	90.6%	53.05	0.01
Clear methods & results	92	0.11	[0.10, 0.13]	87.6%	515.12	0.01
bad/unclear methods*	25	0.17	[0.14, 0.19]	69.6%	167.90	0.01
Clear ethics	32	0.12	[0.08, 0.15]	76.7%	133.28	0.01
bad/unclear ethics	85	0.15	[0.13, 0.17]	85.5%	578.75	0.01
Valid						
Methods match study scope	92	0.13	[0.11, 0.15]	84.7%	596.08	0.01
bad/unclear match	25	0.17	[0.13, 0.21]	73.7%	91.29	0.01
Freedom from bias	101	0.13	[0.11, 0.14]	84.4%	640.82	0.01
bad/unclear bias*	16	0.21	[0.17, 0.25]	81.4%	80.83	0.01
Explicit conclusions	24	0.13	[0.11, 0.15]	82.1%	425.55	0.01
bad/unclear conclusions	77	0.14	[0.11, 0.18]	84.3%	146.55	0.01
Generalizable to higher ed	97	0.15	[0.13, 0.16]	85.4%	658.32	0.01
bad/unclear generalizability	20	0.10	[0.05, 0.15]	71.9%	67.71	0.01
Confident in findings	68	0.13	[0.11, 0.15]	85.6%	466.44	0.01
bad/unclear confidence	49	0.15	[0.12, 0.18]	79.6%	235.60	0.01
Attrition reported	61	0.11	[0.09, 0.13]	87.1%	464.22	0.01
Attrition missing*	56	0.17	[0.14, 0.20]	74.7%	217.00	0.01
Overall	**117**	**0.14**	**[0.12, 0.15]**	**83.7%**	**711.46**	**0.01**

**Indicates a significant and stronger effect size (r) difference in need of correction*

For these data, methodological quality clearly mattered. As noted in the preliminary analysis, specific attrition values were calculated when available ($n = 61$). A meta-regression was run to determine if specific attrition rates can predict effect size relationships (r). Results suggested there was no predictive relationship between specific attrition values and academic performance (*beta* = -0.12*)*, $t(55) = 1.15$, $p = 0.25$. There was, however, a marked increase after binary coding attrition and using all available data. When attrition was missing, the effect size went up *(r = 0.17;* $r_{diff} = 0.06$*;* $Z = 3.72$*,* $p < 0.01$*)*. Of the four significant moderators, all of them associated poor methodological quality with increases in the relationship (r) between motivation and academic performance.

Based on findings from the first objective and first two research questions, motivation may be underestimated (reliability and validity reporting *attempts*). More frequently, motivation relationships with academic performance may be over-estimated (attrition, unclear bias, unclear

methods, *no reporting* of reliability). Adjustments are needed to compensate for these significant variations in measurement and methodological quality.

Objective 2: Synchronicity and motivational construct moderators with quality adjusted effect sizes

Building on findings from the first objective, effect sizes were adjusted for sample size as well as methodological quality and measurement reporting quality. For the remaining research questions, overestimates of relationships (*bad/unclear* presentation, ethics, *missing* attrition, *no reporting* of reliability) were adjusted down and underestimates (reliability and validity reporting *attempts*) were adjusted up. The quality adjusted relationship between motivation and academic performance dropped (r_{adj} = 0.10) by almost one third (see Table 5). Note that the overall correlation is not a meaningful value because of the wide range of study characteristics. It is a useful reference point for subgroups and to capture how profound the adjustments for measurement and methodological quality are.

RQ_3 Class synchronicity as moderators, adjusting for quality

Class synchronicity acted as a stark moderator of the relationship (*r*) between motivation and academic performance. Asynchronous class interactions (n = 17, r_{adj} = 0.04) were not significantly different from 0 (Z = 1.73, p = 0.08. In other words, in asynchronous courses, there was no relationship between motivational constructs and academic performance.

Table 5

Class synchronicity moderators

Class Synchronicity	***n***	r_{adj}	***CI*[95%]**	I^2	***Q***
asynchronous***	17	0.04	[-0.01, 0.09]	86.6%	119.51
mixed	48	0.12	[0.09, 0.15]	77.4%	208.24
synchronous	15	0.11	[0.07, 0.15]	94.5%	253.52
unclear	37	0.10	[0.07, 0.12]	80.6%	185.80
Overall (adj)	**117**	**0.10**	**[0.08, 0.12]**	**83.8%**	**717.78**

***Asynchronous classes were not significantly different than r_{adj} = 0.

Asynchronous interactions were also significantly weaker relationships than both *synchronous* (r_{adj} *= 0.04;* r_{diff} = 0.07*; Z = 2.14, p* = 0.03) and *mixed* (r_{diff} = 0.08*; Z = 2.69, p* = 0.01) class interactions. Finally, *unclear* interactions which could have *asynchronous*, *mixed*, and *synchronous* ODDE elements had similar relationships to all forms of synchronicity. Almost a third of the outcomes (n = 37) were *unclear*, lacking enough of a study description to characterize. From these results, it is clear that class synchronicity acted as a significant moderator in the relationship between motivation and academic performance. Truly *asynchronous* classes are rare in this sample (n = 17), and also show no relationship between motivation and academic performance when adjusting for study quality.

RQ$_4$ Motivational constructs as moderators, adjusting for quality

Making corrections for sample size, methodological quality, and measurement reporting quality clearly impacted the strength of relationships between both motivational categories and specific constructs with academic performance. In particular, one entire motivational category, *attributional style* (r_{adj} = 0.04*; Z* = 1.65*, p* = 0.10), and seven specific constructs were not significantly different from r_{adj}= 0 (see Table 6). The results suggest that *internal* and *external locus of control*, r*elative autonomy index*, *extrinsic motivation*, *technology self-efficacy*, *performance avoidance*, and *mastery avoidance* have no relationship with academic performance in ODDE courses.

Table 6

Motivational construct moderators

Motivational Construct	*n*	r_{adj}	*CI*[95%]	I^2	*Q*
Attributional Style* **	14	0.04	[-0.01, 0.09]	71.4%	45.43
internal locus of control***	6	-0.02	[-0.10, 0.06]	87.9%	41.29
external locus of control (rev)***	4	0.10	[-0.02, 0.21]	38.4%	4.87
optimism	4	0.08	[0.04, 0.12]	51.3%	6.16
Self-determination Motivation	28	0.10	[0.06, 0.13]	83.0%	159.20
relative autonomy index***	1	-0.10	[-0.22, 0.02]	-	-
extrinsic motivation***	7	0.04	[-0.04, 0.11]	74.2%	23.22
general motivation	7	0.12	[0.04, 0.21]	88.4%	51.85
intrinsic motivation	13	0.13	[0.08, 0.18]	77.4%	53.05
Self-efficacy	56	0.11	[0.08, 0.13]	83.5%	333.75
technology self-efficacy***	20	0.04	[-0.01, 0.08]	89.8%	186.00
learning self-efficacy	36	0.15	[0.12, 0.18]	76.1%	146.41
Achievement Goals	11	0.10	[0.07, 0.14]	93.0%	143.77
performance avoidance***	2	0.02	[-0.06, 0.09]	76.2%	4.21
mastery avoidance***	3	0.05	[-0.02, 0.13]	95.8%	47.48
mastery approach	4	0.15	[0.09, 0.21]	95.9%	74.05
performance approach	2	0.17	[0.09, 0.25]	96.7%	30.31
Task Value	8	0.16	[0.11, 0.21]	64.5%	19.74
Overall (adj)	**117**	**0.10**	**[0.08, 0.12]**	**83.8%**	**717.78**

****Motivational construct or category was not significantly different than $r_{adj} = 0$*

Note that the relative autonomy index, as a single outcome, does not have heterogeneity (I^2 or Q) statistics. It represents a single extreme outcome. Clearly, along with several other constructs with emerging sets of evidence, additional research is needed. Half of the motivational constructs had four or fewer outcomes. Authors of included studies clearly responded to theoretical calls to vary measurement of self-efficacy by context. Adaptations included scales for online, computer, and even LMS self-efficacy. Aggregated under technology self-efficacy in our

coding, technology self-efficacy was lower than learning self-efficacy *(r_{adj} = 0.04; r_{diff}* = 0.11 *Z = 3.97, p* < 0.01*)*. Neither approached the self-efficacy relationships found among in-person courses (*r* = 0.31 to 0.59) in the meta-analysis of in-person classes replicated here (Richardson et al. 2012). In summary, motivational construct does moderate the relationship between motivation and academic performance after adjusting for study quality. Many constructs are in need of additional research. From available evidence, half of the constructs, representing about one third of all outcomes show no relationship between motivation and academic performance after adjusting for study quality.

Discussion

ODDE scholars displayed a clear interest in the relationship between motivation and academic performance. Ignoring COVID-19 and MOOC studies and focusing on intentionally designed college and vocational courses. A large volume of outcomes (*n* = 117) was represented in this meta-analytic review. That interest increased rapidly in the form of empirical literature with 23 of those outcomes added just over the past year and a half.

On balance, quality analyses suggested that effect sizes (*r*) for online and blended learning are over-estimated by as much as a third. In some cases, there were clear indications of scholarly rigor for measurement reliability reporting. The only two cases of likely underestimates of effect size were both in the area of measurement reporting quality. For *attempts* to report measurement reliability and measurement validity, the effect size was lower than strong reports as expected, and also lower than no reports. This suggests a level of attenuation in need of positive correction. Only 15% (*n* = 18) of outcomes failed to report any measurement reliability and for the vast majority of cases (*n* = 89) reliability specific to that sample was reported. Given the measurement challenges and range of effect sizes for various measures and study contexts already in the literature (e.g., Cho & Summers, 2012; Huang, 2012; Hulleman et al. 2016), this is to be commended and should continue. Measurement validity reporting was far less clear with almost half (*n* = 55) of the outcomes lacking any kind of report. In many ways, validity is even more crucial given reported relationships between factors and effect size (Huang, 2012). Clearly ODDE scholars made attempts at following theoretical recommendations to assess self-efficacy in domain-specific ways (Bandura, 1986, 2006). However, it is also clear that variations on technology self-efficacy may not align with Bandura's thinking about domain. In a broad sense, through the lens of motivational constructs at least, it is clear that ODDE is a different measurement context for motivation than in-person settings.

Under-reporting was also present in methodological quality. Attrition had so many missing values (*n* = 56) that it lacked sensitivity to use specific attrition rates despite a wide range (see Appendix A) and had to be binary coded to adjust for quality. Methodological quality was not a significant factor most of the time, but when it was significant, lower quality (bad/unclear methods, bad/unclear bias, and no reports of attrition) resulted in significantly stronger correlation values that required correction. Inflated effect sizes for less rigorous methods matched prior reviews of in-person and digital learning (Macnamara et al., 2023). Contextualizing quality in terms of MetaQAT, casting a wide net for relevancy was rewarded

with quality challenges to methodological reliability and methodological validity. The ODDE literature for relationships between motivation and academic performance did appear to have questionable applicability. Note that strong MetaQAT indicators should be more sensitive to real differences in outcomes and more likely to show higher correlations, which was never the case.

Across the combination of measurement reporting and methodological quality, significant differences existed. Quantifying and adjusting for those differences reduced the overall effect size from $r = 0.14$ to $r_{adj} = 0.10$. Quality adjustments similarly resulted in significant changes when exploring mean difference (*g*) variations in interaction strength for online and blended learners (Bernard et al., 2009). The work of comprehensively quantifying and adjusting for quality differences is a direct benefit of this review, and consistent with prior reviews (Macnamara et al., 2023) of in-person learners. Based on the number of quality moderators and the magnitude of effect size reduction from quality adjustments, researchers should strive for more rigorous measurement validity, keep up efforts on measurement reliability, and recognize that reductions in methodological quality may overestimate relationships between motivation and academic performance. An indirect benefit of intense quality coding was the important adjustments to motivational construct assignments for two different studies. Thus, for reviewers, both coding and adjusting for quality were clearly important for properly characterizing and contextualizing the ODDE literature as it intersects with motivation. In a dramatic shift from in-person meta-analyses, several moderators such as asynchronous class interactions and half of the motivational constructs were not just weak effect sizes, they were so weak that they supported the null hypothesis of no relationship between motivation and academic performance in ODDE. Bernard et al. (2009) did not find overall mean differences (*g*) between asynchronous, synchronous or a mix of synchronicity when examined in isolation, but they did detect differences when simultaneously accounting for low, moderate, and high interaction. Our results showed that synchronicity did matter across all observations, specifically that asynchronous settings showed no relationship ($r_{adj} = 0.04$) between motivation and academic performance. As a key finding, our results were a sharp contrast from the widespread agreement about motivation (Richardson et al., 2012), including specific recommendations to enhance academic self-efficacy in order to improve academic outcomes in classroom settings (Honicke et al., 2020). Specific to the ODDE literature, intentional choices to select for (Krebs-Lazendic et al., 2026) or code for synchronous experiences in reviews is warranted—at least in the context of motivation.

Limitations and Future Work

Coding for quality relied on the transparency of primary research. There was an ordinal relationship between good validity (actual analysis as part of a given study) an attempt (citing the validity efforts of another study) or no reporting. Caution is warranted when generalizing findings, especially for validity given the large number ($n = 55$) of studies with no reporting. Additionally, both validity and reliability at times reported specific scores or analyses. In the case of measurement reliability, the vast majority of outcomes reported coefficients for their specific sample ($n = 89$). In short—there was hidden precision in measurement validity and reliability that should be explored. These actual values could be used to model unattenuated correlations (Hunter & Schmidt, 2007) by means of a psychometric meta-analysis, a logical next

step for this research. That work would not be possible without the existing quality efforts of authors in this area of literature—particularly for measurement reliability.

One moderator, applicability to higher education showed weak inter-rater reliability. While it was not a significant moderator of effect size (*r*), measurement error could have masked meaningful differences between good and bad or unclear applicability. Refinement of MetaQAT may be warranted, particularly with uses like this meta-analysis that represent literatures outside of the public health context it was created for (Rosella et al. 2016). Note that this limitation was ameliorated given the consensus coding context for the reported results.

The evidence base for this literature clearly has missing data, such as measurement validity that held true for prior examinations of the ODDE literature. Bernard et al. (2004) reported as much as 60% of study features were missing. Missing data in this study have clearly gone down, suggesting the ODDE literature is improving. The rate does still warrant caution when generalizing to the literature as a whole, especially because both bad/unclear quality elements at times and not reporting attrition were significant moderators of effect size (r). Relevance to select, keep, and code for the full range of study and measurement quality available was an intentional choice for this analysis, in part to quantify the impact and to promote better reporting of future ODDE research.

While all of the motivational constructs do have ties to social cognitive theory, they also have rich independent underlying motivational theories. This was not the first attempt to synthesize across an eclectic mix of motivational constructs and their underlying theories (Lazowski & Hulleman, 2016). Future reviewers may want to specifically code for motivational theory and measurement alignment, or incorporate additional orientations such as student interaction (Moore, 1989) extending the work of Bernard et al., 2009, community of inquiry (Shea et al. 2022), or construct validity, extending the work of Hulleman et al., 2010 for ODDE learners and in a more comprehensive way than their focus on achievement goal theory.

Conclusion

This review combines prior efforts to examine measurement reliability (Richardson et al., 2012), validity (Huang et al. 2012), or methodological quality (Macnamara et al., 2023). Unique to this review space, we use an established quality tool (Rosella et al. 2016). Identifying and adjusting for significant quality moderators we contribute a much needed update of synchronicity bridging more than two decades of research (Bernard et al., 2009) as well as providing a quality adjusted relationships for a range of motivational constructs (Richardson et al., 2012; Walker et al., 2024).

For relationships between motivation and academic performance it was clear that quality did indeed matter. Initially, the primary importance looks academic. ODDE authors are to be commended on their attention and reporting of measurement reliability. For measurement quality reporting as a whole, effect size relationships were largely expected or indicated possible under-estimates of motivation and academic performance. There is a clear need to expand reporting of validity, especially because of the challenges in the online measurement space and how validity

clearly differed from in-person contexts. It is possible that ODDE motivation may need to be measured as part of a larger set of constructs, like the representations of self-efficacy alongside self-regulated learning in the Community of Inquiry framework (Shea et al. 2022) rather than in isolation. The scholarly community may also want to stop exploring variations of technology self-efficacy as an attempt to address calls for domain specificity (Bandura, 2006). Reviewers should clearly address quality in a comprehensive way, not only because of these findings but because of past reviews (Bernard et al., 2009; L.-F. Huang et al., 2012; Macnamara et al., 2023).

Modest correlations for motivation and academic performance (Walker et al., 2024) are even lower when adjusted for measurement reporting quality and methodological quality, to the point that whole categories and half of the constructs show no relationship at all. These findings underscore the need to replicate prior efforts to provide context through exploring study quality (Bernard et al., 2004) or make adjustments for measurement (Richardson et al., 2012). Policy makers should be cautious about extending in-person relationships and recommendations (Honicke et al. 2020) to online learning and should also pay close attention to the impact of truly asynchronous class experiences. These results add to the extant findings that ODDE is a different setting (Cho & Summers, 2012) in regards to motivation. Research quality, review implications, and main findings for synchronicity and motivation should be of interest to policymakers. Policy makers and practitioners may want to explore alternatives or supplements to asynchronous classes in this context. This is an applied field, and there is much to be learned. If motivation does not carry the kind of relationships in ODDE settings that it does face to face—we need to explore why. It could be due to the measurement context; it could be a function of factors for non-traditional students that trump motivation. Quality research is best done with appropriate funding, time, and as a partnership between designers, instructors and researchers—as well as the many academics the engage in multiple roles simultaneously. Given recent increases in online and blended learning experiences (Bay, 2026), the interest in student success models (Jacobsen, 2025), and the high stakes of a declining higher education student population (National Center for Educational Statistics, 2024), practitioners and policy makers stand to benefit from understanding why motivation has such a different relationship for ODDE learners.

References

Alkış, N., & Temizel, T. T. (2018). The impact of motivation and personality on academic performance in online and blended learning environments. *Journal of Educational Technology & Society*, *21*(3), 35–47. https://doi.org/10.30191/ETS.201807_21(3).0004

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*(3), 260–267. https://doi.org/10.1037/0022-0663.80.3.260

Bahçekapili, E., & Karaman, S. (2020). A path analysis of five-factor personality traits, self-efficacy, academic locus of control and academic achievement among online students. *Knowledge Management & E-Learning*, *12*(2), 191–208. https://doi.org/10.34105/j.kmel.2020.12.010

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Bandura, A. (2006). Guide for constructing self-efficacy scales. in f. pajares & t. urdan (eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Information Age Publishing.

Bardach, L., Oczlon, S., Pietschnig, J., & Lüftenegger, M. (2020). Has achievement goal theory been right? a meta-analysis of the relation between goal structures and personal achievement goals. *Journal of Educational Psychology*, *112*(6), 1197–1220. https://doi.org/10.1037/edu0000419

Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—a literature review. *SAGE Open*, *6*(1). https://doi.org/10.1177/2158244015621777

Bay, J. (2026, January 15). *Distance Education Enrollment Surges Nationwide*. Inside Higher Education. https://www.insidehighered.com/news/student-success/academic-life/2026/01/15/distance-education-enrollment-surges-nationwide

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, *79*(3), 1243–1289. https://doi.org/10.3102/0034654309333844

Bernard, R. M., Abrami, P. C., Wade, A., Borokhovski, E., & Lou, Y. (2004, October 19–23). The effects of synchronous and asynchronous distance education: A meta-analytical assessment of Simonson's "equivalency theory" [Conference paper]. 27th Annual Convention of the Association for Educational Communications and Technology, Chicago, IL, United States. https://eric.ed.gov/?id=ED485078

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to Meta-Analysis* (2nd ed.). Wiley.

Chang, C.-C., Yen, W.-H., Liu, S.-C., & Kuo, C.-T. (2026). Learning presence in the revised community of inquiry framework: A systematic review of empirical research (2010–2024). *The Internet and Higher Education*, *68*, 101058. https://doi.org/10.1016/j.iheduc.2025.101058

Cho, M. H., & Summers, J. (2012). Factor validity of the motivated strategies for learning questionnaire (MSLQ) in asynchronous online learning environments. *Journal of Interactive Learning Research*, *23*(1), 5-28.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: Ii. resolving the paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 551–558. https://doi.org/10.1016/0895-4356(90)90159-m

de Araujo, J., Gomes, C. M. A., & Jelihovschi, E. G. (2023). The factor structure of the motivated strategies for learning questionnaire (MSLQ): New methodological approaches and evidence. *Psicologia, Reflexao e Critica: Revista Semestral Do Departamento de Psicologia Da UFRGS*, *36*(1), 38. https://doi.org/10.1186/s41155-023-00280-0

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501

Fromm, Y. M., Martin, F., Gezer, T., & Ifenthaler, D. (2025). Best practices for conducting systematic reviews: Perspectives of experienced systematic review researchers in educational sciences. *Technology, Knowledge and Learning*, *30*(1), 1–28. https://doi.org/10.1007/s10758-025-09819-9

Gardner, A., Ibrahim, M., Karamarkovich, S., Holsapple, M., & Shapiro, D. (2024). PDP insights: Early indicators of timely completers. in *National Student Clearinghouse*. National Student Clearinghouse Research Center. https://eric.ed.gov/?id=ED673395

Goulão, M. (2014). The relationship between self-efficacy and academic achievement in adults' learners. *Athens Journal of Education*, *1*, 237–246. https://doi.org/10.30958/aje.1-3-4

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Hilpert, J. C., Stempien, J., van der Hoeven Kraft, K. J., & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: A new conceptualization of an established questionnaire. *SAGE Open*, *3*(4), 2158244013510305. https://doi.org/10.1177/2158244013510305

Hodges, C. B. (2008). Self-efficacy, motivational email, and achievement in an asynchronous math course. *The Journal of Computers in Mathematics and Science Teaching*, *27*(3), 265–285.

Honicke, T., Broadbent, J., & Fuller-Tyszkiewicz, M. (2020). Learner self-efficacy, goal orientation, and academic achievement: Exploring mediating and moderating relationships. *Higher Education Research & Development*, *39*(4), 689–703. https://doi.org/10.1080/07294360.2019.1685941

Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(1), 48–73. https://doi.org/10.1037/a0026223

Huang, L.-F., Wu, T.-Y., Hou, S.-H., & Horng, R.-Y. (2012). Effects of feedback type, problem presentation, and practice on problem solving. *Bulletin of Educational Psychology*, *43*(4), 855–874.

Hulleman, C.S., Barron, K.E., Kosovich, J.J., Lazowski, R.A. (2016). Student Motivation: Current Theories, Constructs, and Interventions Within an Expectancy-Value Framework. In: Lipnevich, A., Preckel, F., Roberts, R. (eds) Psychosocial Skills and School Systems in the 21st Century. The Springer Series on Human Exceptionality. Springer, Cham. https://doi.org/10.1007/978-3-319-28606-8_10

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A Meta-Analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, *136*(3), 422–449. https://doi.org/10.1037/a0018947

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting Error and Bias in Research Findings*. Sage Publications.

Hunter, J. E., & Schmidt, F. L. (2007). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. SAGE Publications, Inc.

Jacobsen, L. J. (2025). *ACCESS: A Framework for Student Success*. Inside Higher Ed. https://www.insidehighered.com/opinion/views/2025/08/06/framework-organizing-student-success-efforts-opinion

Johnson, N., Seaman, J., & Poulin, R. (2022). Defining different modes of learning: Resolving confusion and contention through consensus. *Online Learning*, *26*(3). https://doi.org/10.24059/olj.v26i3.3565

Joo, Y. J., Lim, K. Y., & Kim, J. (2013). Locus of control, self-efficacy, and task value as predictors of learning outcome in an online university context. *Computers & Education*, *62*, 149–158. https://doi.org/10.1016/j.compedu.2012.10.027

Katrak, P., Bialocerkowski, A. E., Massy-Westropp, N., Kumar, S., & Grimmer, K. A. (2004). A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology*, *4*, 22. https://doi.org/10.1186/1471-2288-4-22

Kebritchi, M., Lipschuetz, A., & Santiague, L. (2017). Issues and challenges for teaching successful online courses in higher education: A literature review. *Journal of Educational Technology Systems*, *46*(1), 4–29. https://doi.org/10.1177/0047239516661713

Kim, J.-Y. (2012). A study on learners' perceptional typology and relationships among the learner's types, characteristics, and academic achievement in a blended E-Education environment. *Computers & Education*, *59*(2), 304–315.

Krebs-Lazendic, L., Dello Stritto, M. E., & Park, E. (2026). A critical review of research on synchronous instructor presence in online learning: Elusive evidence and mixed results. *Quarterly Review of Distance Education*, 27(1)*,* 1–19. https://doi.org/10.1108/QRDE-09-2025-0016

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A Meta-Analytic review. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Lee, K. (2017). Rethinking the ACCESSibility of online higher education: A historical review. *The Internet and Higher Education*, *33*, 15–23. https://doi.org/10.1016/j.iheduc.2017.01.001

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Sage.

Liu, C., Sands-Meyer, S., & Audran, J. (2019). The effectiveness of the student response system (SRS) in english grammar learning in a flipped english as a foreign language (EFL) class. *Interactive Learning Environments*, *27*(8), 1178–1191.

Lucas, M., & Vicente, P. N. (2023). A double-edged sword: Teachers' perceptions of the benefits and challenges of online teaching and learning in higher education. *Education and Information Technologies*, *28*(5), 5083–5103. https://doi.org/10.1007/s10639-022-11363-3

Macnamara, B. N., & Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? a systematic review and meta-analysis with recommendations for best practices. *Psychological Bulletin*, *149*(3–4), 133–173. https://doi.org/10.1037/bul0000352

Mahmud, M., Freeman, B., & Abu Bakar, M. (2022). Technology in education: Efficacies and outcomes of different delivery methods. *Interactive Technology and Smart Education*, *19*(1), 20–38. https://doi.org/10.1108/ITSE-01-2021-0021

Martin, F., Wu, T., Wan, L., & Xie, K. (2022). A Meta-Analysis on the Community of Inquiry Presences and Learning Outcomes in Online and Blended Learning Environments. *Online Learning (Newburyport, Mass.)*, *26*(1), 325. https://doi.org/10.24059/olj.v26i1.2604

Martin, F., Tutty, J. I., & Su, Y. (2010). Influence of learning management systems self-efficacy on e-learning performance. *Journal on School Educational Technology*, *5*(3), 26–35.

Marzi, G., Balzano, M., & Marchiori, D. (2024). K-Alpha calculator–Krippendorff's alpha calculator: A user-friendly tool for computing Krippendorff's alpha inter-rater reliability coefficient. *MethodsX*, *12*, 102545. https://doi.org/10.1016/j.mex.2023.102545

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

McPhaul-Moore, E. (2013). Predictors of online learning success in a rural community college [doctoral dissertation, northcentral university]. in *ProQuest LLC*. https://www.proquest.com/docview/1444627614

Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A Meta-Analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

National Center for Education Statistics. (2024). *College Enrollment Rates* (Condition of Education). Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cpb

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinnrdd, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *71*. https://doi.org/10.1136/bmj.n71

Perkowski, J. (2013). The role of gender in distance learning: A Meta-Analytic review of gender differences in academic performance and self-efficacy in distance learning. *Journal of Educational Technology Systems*, *41*(3), 267–278. https://doi.org/10.2190/ET.41.3.e

Pintrich, P. R. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. https://eric.ed.gov/?id=ED338122

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rosella, L., Bowman, C., Pach, B., Morgan, S., Fitzpatrick, T., & Goel, V. (2016). The development and validation of a meta-tool for quality appraisal of public health evidence: Meta quality appraisal tool (MetaQAT). *Public Health*, *136*, 57–65. https://doi.org/10.1016/j.puhe.2015.10.027

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Savoji, A. P. (2013). Motivational strategies and academic achievement in traditional and virtual university students. *Procedia - Social and Behavioral Sciences*, *84*, 1015–1020. https://doi.org/10.1016/j.sbspro.2013.06.691

Senko, C., & Dawson, B. (2017). Performance-approach goal effects depend on how they are defined: Meta-Analytic evidence from multiple educational outcomes. *Journal of Educational Psychology*, *109*(4), 574–598. https://doi.org/10.1037/edu0000160

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, *7*, 10. https://doi.org/10.1186/1471-2288-7-10

Shea, P., Richardson, J., & Swan, K. (2022). Building bridges to advance the community of inquiry framework for online learning. *Educational Psychologist*, *57*(3), 148–161. https://doi.org/10.1080/00461520.2022.2089989

Stephen, J. S., Rockinson-Szapkiw, A. J., & Dubay, C. (2020). Persistence model of non-traditional online learners: Self-efficacy, self-regulation, and self-direction. *American Journal of Distance Education*, *34*(4), 306–321. https://doi.org/10.1080/08923647.2020.1745619

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Storrings, D. (2005). Attrition in distance education: A meta-analysis. *Instructional Design, Development and Evaluation - Dissertations*. https://surface.syr.edu/idde_etd/8

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, *66*, 101976. https://doi.org/10.1016/j.cedpsych.2021.101976

Tladi, L. S. (2017). Perceived ability and success: Which self-efficacy measures matter? a distance learning perspective. *Open Learning: The Journal of Open, Distance and e-Learning*, *32*(3), 243–261. https://doi.org/10.1080/02680513.2017.1356711

Torun, E. D. (2020). Online distance learning in higher education: E-learning readiness as a predictor of academic achievement. *Open Praxis*, *12*(2), 191–208. https://search.informit.org/doi/10.3316/informit.352722164544178

Walker, A., Aguiar, N. R., Soicher, R., Kuo, Y.-C., & Resig, J. (2024). Exploring the relationships between motivation and academic performance among online and blended learners: A Meta-Analytic review. *Online Learning*, *28*(4). https://doi.org/10.24059/olj.v28i4.4602

Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology*, *92*(1), 137–143. https://doi.org/10.1037/0022-0663.92.1.137

Wang, L. (2022). Student Intrinsic Motivation for Online Creative Idea Generation: Mediating Effects of Student Online Learning Engagement and Moderating Effects of Teacher Emotional Support. *Frontiers in Psychology*, *13*, 954216. https://doi.org/10.3389/fpsyg.2022.954216

Warren, L., Reilly, D., Herdan, A., & Lin, Y. (2020). Self-efficacy, performance and the role of blended learning. *Journal of Applied Research in Higher Education*, *13*(1), 98–111. https://doi.org/10.1108/JARHE-08-2019-0210

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist*, *45*(1), 28–36. https://doi.org/10.1080/00461520903433596

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*(5), 2443–2476. https://doi.org/10.1007/s11423-023-10233-0

Whiting, P., Savović, J., Higgins, J. P. T., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., & Churchill, R. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, *69*, 225–234. https://doi.org/10.1016/j.jclinepi.2015.06.005

Zhao, Y., Lei, J., Lai, B. Y. C., & Tan, H. S. (2005). What makes the difference? a practical analysis of research on the effectiveness of distance education. *Teachers College Record*, *107*(8), 1836–1884. https://doi.org/10.1111/j.1467-9620.2005.00544.x

Acknowledgements

This work was supported by a three-year seminar series on meta-analysis by Oregon State University. Authors received travel support and Andrew Walker received a small stipend as co-leader of the seminar.

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5822
https://doi.org/10.24059/olj.5822

Appendix A -- Studies, Outcomes, and Coding

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Al Fadda (2019)														
learning SE	asynch	attempt	good	--	--	--	--	--	--	--	--	80%	0.29	0.21
Alkis & Temizel (2018)														
intrinsic motiv	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.17	0.12
learning SE	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.22	0.17
intrinsic motiv	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.44	0.39
learning SE	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.03	-0.02
Amin (2019)														
intrinsic motiv	mixed	none	none	good	--	--	--	good	good	good	--	--	0.65	0.54
extrinsic motiv	mixed	none	none	good	--	--	--	good	good	good	--	--	0.48	0.37
learning SE	mixed	none	none	good	--	--	--	good	good	good	--	--	0.57	0.46
Bahçekapili & Karaman (2020)														
internal LoC	synch	none	good	good	good	--	good	--	--	good	good	18%	0.03	-0.04
external LoC	synch	none	good	good	good	--	good	--	--	good	good	18%	0.16	0.09
learning SE	synch	none	good	good	--	--	good	--	--	good	good	18%	0.14	0.04
Baturay & Yukselturk (2015)														
technology SE	mixed	none	good	good	good	--	good	good	--	--	--	67%	0.11	0.11
Broadbent & Howe (2023)														
technology SE	mixed	none	good	good	--	good	good	good	good	good	--	--	0.27	0.21
Cannon et al. (2021)														
learning SE	mixed	good	good	good	good	good	good	good	good	good	good	25%	0.05	0.05
Chang & Ho (2009)														
internal LoC	asynch	none	good	good	good	--	--	good	good	good	--	--	0.17	0.14
Chang et al. (2014)														
technology SE	unclear	none	good	good	good	--	good	--	--	good	good	8%	0.24	0.17

Corresponding author: Dr. Andrew Walker, andy.walker@usu.edu
This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 The Author(s).

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[i]**	**8[j]**	**9[k]**	***r***	r_{adj}
Chen & Jang (2010)														
RAI	mixed	good	good	good	--	good	good	good	good	good	good	--	-0.04	-0.10
Cheng et al. (2023)														
learning SE	unclear	attempt	attempt	good	--	good	good	good	good	good	good	--	0.10	0.18
task value	unclear	attempt	attempt	good	--	good	good	good	good	good	good	--	0.07	0.15
Cho & Shen (2013)														
intrinsic motiv	asynch	none	good	good	--	good	good	good	good	good	--	--	-0.09	-0.15
extrinsic motiv	asynch	none	good	good	--	good	good	good	good	good	--	--	-0.08	-0.14
learning SE	asynch	none	good	good	--	good	good	good	good	good	--	--	0.18	0.12
Cigdem & Ozturk (2016)														
technology SE	mixed	attempt	good	good	--	--	good	--	--	--	--	--	0.21	0.10
general motiv	mixed	attempt	good	good	--	--	good	--	--	--	--	--	0.16	0.05
de Fátima Goulão (2014)														
learning SE	unclear	none	good	good	--	--	good	--	good	good	good	--	0.29	0.16
DeTure (2004)														
technology SE	mixed	good	good	good	--	--	good	good	good	good	good	55%	-0.03	-0.06
Ergul (2004)														
learning SE	unclear	none	good	good	--	--	good	good	good	good	good	--	0.24	0.18
mastery approach	unclear	none	good	good	--	--	good	good	good	good	good	--	0.01	-0.05
Gerlich et al. (2009)														
external LoC	asynch	none	none	good	--	--	good	good	good	--	--	--	0.08	-0.03
Goad et al. (2021)														
mastery approach	asynch	attempt	none	good	--	good	--	--	good	good	--	5%	0.17	0.04
Gravelle et al. (2023)														

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
learning SE	mixed	none	good	good	good	good	good	good	good	good	good	12%	0.08	0.08
learning SE	mixed	none	good	good	--	good	good	good	good	good	good	24%	0.10	0.07

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Gravelle et al. (2023)														
learning SE	mixed	none	good	good	good	good	good	good	good	good	good	12%	0.08	0.08
learning SE	mixed	none	good	good	--	good	good	good	good	good	good	24%	0.10	0.07
Hamm et al. (2019)														
optimism	unclear	good	good	good	good	--	good	good	good	good	good	19%	0.07	0.07
optimism	unclear	good	good	good	good	--	good	good	good	good	good	2%	0.04	0.04
optimism	unclear	good	good	good	good	--	good	good	good	good	good	19%	0.17	0.17
optimism	unclear	good	good	good	good	--	good	good	good	good	good	2%	0.04	0.04
Hefter & Nitsch (2024)														
mastery avoidance	synch	none	good	good	--	good	good	good	good	good	--	--	-0.21	-0.27
learning SE	synch	none	good	good	--	good	good	good	good	good	--	--	0.04	-0.02
learning SE	synch	none	good	good	--	good	good	good	good	good	--	--	0.30	0.24
Hobson & Puruhito (2018)														
learning SE	unclear	attempt	attempt	good	good	--	good	--	good	good	good	28%	0.31	0.38
Hodges (2008)														
learning SE	mixed	attempt	good	--	good	good	good	--	good	good	--	56%	0.28	0.23
Jadric, Bubas & Hutinski (2010)														
technology SE	mixed	attempt	good	good	--	--	good	good	good	--	good	38%	0.15	0.14
Jiang et al. (2024)														
general motiv	mixed	good	good	good	--	good	--	good	good	good	good	--	0.35	0.29
Johnson et al. (2008)														
technology SE	asynch	good	good	good	--	--	good	good	good	good	good	--	0.27	0.21
Joo, Lim, & Kim (2013)														

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
internal LoC	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.07	0.07
learning SE	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.22	0.22
task value	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.25	0.25
task value	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.19	0.19
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[i]**	**8[j]**	**9[k]**	***r***	r_{adj}
Kim (2012)														
learning SE	mixed	none	none	good	good	--	--	good	good	good	--	--	0.40	0.32
learning SE	mixed	none	none	good	good	--	--	good	good	good	--	--	0.40	0.32
Kong et al. (2025)														
intrinsic motiv	asynch	none	good	good	--	--	good	good	good	--	--	--	0.16	0.10
extrinsic motiv	asynch	none	good	good	--	--	good	good	good	--	--	--	0.16	0.10
Lee & Choi (2013)														
internal LoC	unclear	good	good	good	good	--	good	good	good	good	good	--	0.29	0.26
Lin et al. (2024)														
intrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.34	0.28
extrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.06	0.00
task value	mixed	none	good	good	--	--	good	good	good	good	good	--	0.18	0.12
learning SE	mixed	none	good	good	--	--	good	good	good	good	good	--	0.28	0.22
intrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.37	0.31
extrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.04	-0.02
task value	mixed	none	good	good	--	--	good	good	good	good	good	--	0.40	0.34
learning SE	mixed	none	good	good	--	--	good	good	good	good	good	--	0.26	0.20
Liu et al. (2019)														
learning SE	mixed	none	none	good	good	--	good	--	good	--	--	--	-0.06	-0.21
general motiv	mixed	none	none	good	good	--	good	--	good	--	--	--	-0.58	-0.53
Liu, Ma, & Chen (2024)														
extrinsic motiv	unclear	good	good	good	good	--	good	good	good	good	good	8%	-0.10	-0.10
intrinsic motiv	unclear	good	good	good	good	--	good	good	good	good	good	8%	0.08	0.08

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Lynch & Dembo (2004)														
learning SE	mixed	good	good	good	good	--	good	good	good	good	good	73%	0.29	0.29
technology SE	mixed	good	good	good	good	--	good	good	good	good	good	73%	-0.09	-0.09
Martin et al. (2010)														
technology SE	mixed	good	good	good	--	--	good	good	good	good	good	--	0.20	0.14
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[i]**	**8[j]**	**9[k]**	***r***	r_{adj}
McPhaul-Moore (2013)														
learning SE	unclear	attempt	none	good	good	good	good	good	good	good	good	24%	0.23	0.20
intrinsic motiv	unclear	attempt	none	good	good	good	--	good	good	good	--	29%	0.09	0.06
Morris (2005)														
internal LoC	unclear	attempt	attempt	good	--	good	--	good	good	good	--	87%	-0.26	-0.15
Neroni et al. (2022)														
learning SE	asynch	none	good	good	good	good	good	good	--	good	--	77%	0.05	0.05
Noteborn et al. (2012)														
task value	unclear	none	good	good	--	--	good	good	good	good	good	--	0.14	0.08
Ongor & Uslusoy (2023)														
general motiv	synch	good	good	good	good	good	good	--	good	good	good	--	0.23	0.13
Parker et al. (2021)														
intrinsic motiv	unclear	none	good	good	--	--	good	good	good	good	good	--	0.29	0.23
task value	unclear	none	good	good	--	--	good	good	good	good	good	--	0.07	0.01
Redstone (2023)														
technology SE	asynch	attempt	good	good	--	good	good	good	good	good	good	55%	-0.10	-0.11
Remedios & Richardson (2012)														
mastery approach	unclear	good	good	good	good	--	good	good	good	good	good	35%	0.00	0.00
mastery avoidance	unclear	good	good	good	good	--	good	good	good	good	good	35%	-0.01	-0.01
performance approach	unclear	good	good	good	good	--	good	good	good	good	good	35%	-0.01	-0.01

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
performance avoidance	unclear	good	good	good	good	--	good	good	good	good	good	35%	0.08	0.08
Runyon (2013)														
technology SE	unclear	good	attempt	good	good	good	good	good	good	good	good	58%	0.28	0.40
Sankaran & Bui (2001)														
general motiv	unclear	none	good	--	--	--	--	good	good	good	--	--	0.58	0.52
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[j]**	**8[k]**	**9[l]**	***r***	r_{adj}
Savoji (2013)														
intrinsic motiv	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.06	0.05
extrinsic motiv	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.06	0.05
task value	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.13	0.12
learning SE	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.29	0.28
Simmering, Posey & Piccoli (2009)														
technology SE	mixed	attempt	good	good	good	--	good	good	good	good	good	12%	0.29	0.31
general motiv	mixed	attempt	good	good	good	--	good	good	good	good	good	12%	-0.07	-0.05
Stephen et al. (2020)														
technology SE	unclear	attempt	attempt	--	good	--	good	good	good	good	good	54%	-0.78	-0.53
learning SE	unclear	attempt	attempt	good	good	--	good	good	good	good	good	54%	-0.03	0.11
Tai (2016)														
learning SE	mixed	good	good	good	good	good	good	good	good	good	good	--	-0.29	-0.32
Talsma et al. (2023)														
learning SE	mixed	none	none	good	--	good	--	good	good	good	--	--	0.31	0.20
learning SE	mixed	none	good	good	--	good	good	good	good	good	--	--	0.14	0.08
Tladi (2017)														
learning SE	synch	attempt	good	good	good	--	good	good	good	good	good	98%	0.13	0.15
technology SE	synch	attempt	good	good	good	--	good	good	good	good	good	98%	-0.74	-0.53
Torun (2020)														
general motiv	synch	good	good	good	good	--	--	--	--	good	--	1%	0.51	0.44

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
technology SE	synch	good	good	good	good	--	good	--	--	good	--	1%	0.16	0.09
Tseng, Chen, & Lin (2023)														
learning SE	mixed	attempt	good	good	--	good	good	--	good	good	--	--	0.18	0.07
Wadsworth et al. (2007)														
learning SE	mixed	none	none	good	good	--	good	good	good	--	--	83%	0.32	0.27
Wang & Newlin (2000)														
external LoC	mixed	none	none	good	--	--	--	good	good	--	--	4%	0.41	0.33
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[j]**	**8[k]**	**9[l]**	***r***	r_{adj}
internal LoC	mixed	none	none	good	--	--	--	good	good	--	--	4%	-0.33	-0.41
Wang & Newlin (2002)														
learning SE	mixed	none	none	good	good	--	--	--	good	--	--	--	0.32	0.17
technology SE	mixed	none	none	good	good	--	--	--	good	--	--	--	0.24	0.09
Wang et al. (2008)														
intrinsic motiv	unclear	none	attempt	good	--	--	good	good	good	--	--	--	0.16	0.22
technology SE	unclear	none	attempt	good	--	--	--	good	good	--	--	--	0.05	0.11
external LoC	unclear	none	attempt	good	--	--	--	good	good	--	--	--	-0.07	-0.01
Warren et al. (2020)														
learning SE	mixed	none	none	--	good	good	--	--	--	--	--	74%	0.46	0.34
Wipawayangkool et al. (2022)														
technology SE	unclear	good	good	good	good	--	good	good	good	good	good	--	-0.01	-0.04
Wuellner (2015)														
intrinsic motiv	asynch	none	none	--	good	--	good	--	good	--	--	18%	-0.41	-0.53
Zhang et al. (2023)														
technology SE	unclear	good	good	good	good	--	good	good	good	good	good	16%	0.13	0.13
Zhou & Wang (2019)														
mastery approach	synch	good	good	good	good	--	good	good	good	good	good	31%	0.61	0.61

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
mastery avoidance	synch	good	good	good	good	--	good	good	good	good	good	31%	0.44	0.44
performance approach	synch	good	good	good	good	--	good	good	good	good	good	31%	0.35	0.35
performance avoidance	synch	good	good	good	good	--	good	good	good	good	good	31%	-0.05	-0.05
Zimmerman (2017)														
learning SE	mixed	none	good	good	good	--	good	good	good	good	--	34%	-0.02	-0.02
technology SE	mixed	none	good	good	good	--	good	good	good	good	--	35%	-0.10	-0.10
learning SE	mixed	none	good	good	good	--	good	good	good	good	--	36%	0.12	0.12

Note: For Construct: SE=self-efficacy; LoC=locus of control; motiv=motivation; RAI=relative autonomy index; Delivery=synchronicity --=bad, unclear, or not reported (blank); r=raw effect size; r_{adj}=Hunter's r, adjusted for sample size and quality

[a]*Measurement Reporting Quality - Validity*

[b]*Measurement Reporting Quality - Reliability*

[c]*Study Quality - Reliability: Clear presentation*

[d]*Study Quality - Reliability: Clear methods and results*

[e]*Study Quality - Reliability: Ethics procedures described*

[f]*Study Quality - Validity: Methods align with scope of research*

[g]*Study Quality - Validity: Free from bias*

[h]*Study Quality - Validity: Explicit and transparent conclusions*

[i]*Study Quality - Applicable to setting (higher education)*

[j]*Study Quality - Confident in findings*

[k]*Study Quality - Attrition for sample (percentage*

ONLINE LEARNING
Volume 30 · Issue SI2 · 2026 · e5823
https://doi.org/10.24059/olj.v30iSI2.5823

Verification Without Criteria: An Epistemic Network Analysis of Metacognitive Laziness in Generative AI Research

Ji Hyun Yu, Fengjiao Tu, and Li Cheng
University of North Texas, Denton, Texas
Shasha Li
McGill University, Montreal, Quebec, Canada
Zhizi Zheng
The Education University of Hong Kong, Hong Kong

Abstract

Generative AI performs the analytical and evaluative tasks previously managed by learners, and the field has responded by designing scaffolds that restore monitoring. However, whether monitoring represents the primary site of regulatory failure remains untested. To address this, we conducted a systematic review and Epistemic Network Analysis (ENA) of 65 empirical studies (2022–2026). Using the published study as the unit of analysis, we examined the relational structures connecting instructional designs and regulatory processes rather than how often each is reported. The findings reveal three critical structural insights. First, no study taught learners the criteria required to evaluate AI outputs. While interventions frequently prompted learners to pause or explain, none provided evaluative standards, reflecting an implicit assumption that learners already possess adequate evaluation skills. Second, monitoring and evaluation are reported in 94% of studies, and in those reporting metacognitive laziness they are reported together with AI delegation while planning is reported apart from it. Planning is reported in most studies in both outcome groups, so the difference lies in which processes co-occur rather than in whether planning is reported at all. We describe this configuration as *verification without criteria*. Third, structural configurations vary by intervention timing. Instructor mediation, implemented prior to the first query, anchors delegation within a planning-oriented network, whereas in-task prompting alone isolates delegation from planning. We argue that the placement of friction rather than its amount determines whether verification has a standard, and that evaluative criteria should be established before the first AI query, a proposition this corpus has not tested.

Keywords: Cognitive offloading; Epistemic network analysis; Evaluative judgment; Feedback literacy; Hybrid human-AI regulation

Introduction

The integration of Generative Artificial Intelligence (GenAI) into open, distance, and digital education (ODDE) has fundamentally changed learner tasks. Autonomy in ODDE has historically benefited learners with strong self-regulation skills while hindering those who lack such capabilities (Jin et al., 2023). Unlike earlier digital tools that provided navigational scaffolds without replacing learner cognition (Álvarez et al., 2022), GenAI directly performs analytical, generative, and evaluative tasks. This capability shifts a significant portion of cognitive responsibility away from the learner.

Recent studies conceptualize this resulting risk as metacognitive laziness, linking it to a decline in monitoring and evaluation, the foundational processes of the self-regulated learning (SRL) cycle (Fan et al., 2025). Empirical trace data align with this concern, showing that

Corresponding author: Ji Hyun Yu, jihyun.yu@unt.edu
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

metacognitive processes are less strongly associated under ChatGPT support than with human experts (Fan et al., 2025), and that learners frequently skip the evaluation phase entirely (Chen et al., 2025). However, a crucial distinction must be made. While the field often interprets this data as a simple drop in the absolute frequency of regulatory behaviors, the evidence points to a structural breakdown. These trace data implicitly document how the phases of regulation fail to connect properly once a highly fluent AI system executes the performance phase.

To address this structural breakdown, the field has primarily relied on system design rather than direct instruction. Scaffolds like reflective prompting, cognitive mirroring, and instructor mediation attempt to externally bridge these disconnected phases by treating regulation as a shared socio-technical responsibility (Molenaar, 2022; Seo et al., 2021; Singh et al., 2025; Tomisu et al., 2025). However, while these tools prompt learners to pause or review their actions, they do not provide the domain-specific criteria needed to evaluate GenAI outputs accurately. Furthermore, simply possessing criteria is insufficient without direct instruction in feedback literacy, which is the capacity to interpret and judge feedback against those criteria rather than passively receive it. The absence of this literacy leaves learners exposed to AI errors. T. W. Li et al. (2025) found that 80% of participants failed to detect factual errors in chatbot responses, and those errors measurably harmed learning. It remains unclear whether the literature treats this evaluative capacity as something to be taught or presumes learners already possess it.

Addressing this conceptual gap is critical, as the rapid implementation of GenAI in educational settings is outpacing empirical evaluation. Research at the intersection of GenAI and SRL has surged from a handful of studies in 2023 to over fifty in 2025, leading to the large-scale deployment of many instructional designs. These deployments, along with their accompanying institutional guidelines, are primarily designed to restore monitoring, operating on the assumption that this is where the regulatory failure lies. However, this foundational assumption remains empirically untested. Existing literature reviews cannot evaluate it because they isolate scaffolds and outcomes into separate frequency counts (Zhai et al., 2024). This methodological limitation obscures whether task delegation to the system is connected to the verification processes meant to check it. Furthermore, simple frequency counts cannot distinguish between a pedagogical scaffold that was never attempted and one that was attempted but failed. Consequently, resolving this issue requires a methodological synthesis that models the structural relations among these constructs rather than merely aggregating their frequencies.

We conduct a systematic review and epistemic network analysis of the emerging literature on GenAI and self-regulated learning ($N = 65$ empirical studies, 2022 to 2026). Taking each published study rather than each learner as the unit of analysis, we model how instructional designs and regulatory processes are reported together across this literature. This choice is consequential. A code may appear at similar rates in two groups of studies while connecting to entirely different processes in each, and frequency-based synthesis cannot recover that difference. The first two research questions examine how the scaffolds currently implemented are associated with distinct regulatory configurations. The third documents how feedback literacy is represented across the literature and compares the configurations reported under each engagement outcome. Guided by the coding framework detailed in the following section, we address the following research questions:

1. How do the reported patterns of metacognitive processes differ across studies employing reflective prompting, instructor mediation, both, or neither?
2. How do the reported patterns differ between studies that implement cognitive mirroring and those that do not?
3. How is feedback literacy represented across the corpus, and how do the reported patterns differ between studies reporting metacognitive laziness and those reporting productive engagement?

Conceptual Framework

Hybrid Human-AI Regulation and the SRL Cycle

From a socio-cognitive perspective, Zimmerman (2000) defines SRL as a cyclical process in which forethought, performance, and self-reflection regulate learning (Zimmerman & Moylan, 2009). Aligning with this theoretical foundation, our coding framework operationalizes the SRL cycle into three primary phases: planning, monitoring, and evaluation (see Table 2). The integration of GenAI, however, necessitates the inclusion of a fourth regulatory process: delegation. While conventional digital tools, such as spell-checkers, enhance operational efficiency without replacing human reasoning, GenAI affords learners the capability to completely delegate complex analytical and generative tasks (Risko & Gilbert, 2016). This capability introduces a distinct regulatory risk. Because GenAI produces highly fluent text that can seamlessly mask factual or logical errors, it frequently bypasses learners' natural error-detection mechanisms. To conceptualize this dynamic, this study adopts Molenaar's (2022) framework of hybrid human-AI regulation, which posits that the monitoring and control of learning must not be fully delegated to the AI, but rather function as a shared responsibility jointly executed by the learner and the system.

Scaffolds as Designed Interruption

Current literature identifies three primary modalities of instructional scaffolding designed to sustain this shared human-AI regulation. First, reflective prompting requires learners to explicitly externalize their cognitive processes during the interaction (Wong et al., 2019). Second, cognitive mirroring visualizes learners' own reasoning patterns, facilitating the identification of knowledge gaps that they might otherwise ignore when minimizing cognitive effort (Azevedo & Wiedbusch, 2023; Molenaar, 2022). Third, instructor mediation establishes explicit rules and assessment structures governing AI utilization (Seo et al., 2021). Crucially, these scaffolds intervene at different temporal stages of the learning task. Instructor mediation typically occurs prior to the interaction, reflective prompting is embedded during query generation, and cognitive mirroring operates after the learner produces an output. Despite these distinct temporal placements, prior research has not systematically compared how these timing differences structurally configure the learning process.

Feedback Literacy as a Cognitive Prerequisite

Feedback literacy is defined as the capacity to interpret, judge, and act upon feedback (Carless & Boud, 2018), functioning as an essential component of the self-reflection phase in SRL (Tai et al., 2018). However, evaluating AI-generated feedback presents unique epistemic challenges compared to traditional assessment for three reasons. First, the absence of a human author precludes learners from interrogating the underlying reasoning behind the feedback.

Second, AI systems generate both accurate information and fabrications with identical linguistic fluency, thereby eliminating the natural communicative cues of uncertainty (Waldo & Boussard, 2024). Third, the virtually unlimited volume of AI-generated feedback forces the learner to independently filter for relevance. In the absence of explicit evaluative criteria, learners frequently judge AI outputs based on surface-level plausibility rather than factual accuracy. Consequently, learners risk developing an "illusion of explanatory depth," falsely equating the AI's fluent explanations with their own conceptual understanding (Rozenblit & Keil, 2002). Therefore, feedback literacy must be operationalized as a critical prerequisite for AI integration, not a natural byproduct of its use. Because developing this capacity requires explicit instruction, its presence or absence in pedagogical interventions serves as a directly observable variable in the literature.

Two Regulatory Trajectories

The literature documents two distinct regulatory trajectories resulting from human-AI interaction. First, we conceptualize metacognitive laziness not as a deficit in learner motivation, but as a specific structural failure within the learner's regulatory network (Fan et al., 2025). While previous studies attribute this failure to the attenuation of monitoring and evaluation, assuming that fluent AI text pacifies error detection, the SRL cycle can fracture through multiple distinct pathways. For instance, delegating tasks to an AI without subsequent verification represents a fundamentally different structural failure than attempting to verify outputs without having established a clear standard during the prior planning phase. Epistemic network models are uniquely suited to map these divergent structural breakdowns. Conversely, productive engagement emerges when learners successfully couple cognitive offloading with active human oversight. In this successful trajectory, the learner and the AI system jointly orchestrate learning strategies, functioning in a manner analogous to effective socially shared regulation (Hadwin et al., 2011).

Methodology

Design

This study employed a systematic review design governed by the six-phase DISCAR framework (West & Martin, 2024), encompassing design, inclusion and exclusion, screening, coding, analysis, and reporting. The reporting structure adhered to the PRISMA 2020 guidelines (Page et al., 2021). Table 1 outlines the specific inclusion and exclusion criteria. Search queries were filtered by publication date, document type, and language, while thematic criteria were strictly applied during the screening phases.

Table 1

Inclusion and Exclusion Criteria

	Inclusion	Exclusion
Date	November 1, 2022, and January 31, 2026	Outside this window

	Inclusion	Exclusion
Type	Peer-reviewed articles, conference proceedings, theses	Books, magazines, review papers, white papers
Language	English	Non-English
Study Focus	Explicit focus on metacognition or self-regulated learning (e.g., SRL, co-regulation, cognitive offloading)	Simple task-level offloading (e.g., grammar fixing)
Context	Educational or learning settings (e.g., pedagogy, assessment, online learning)	Exclude those not related to Education.
Technology	Generative AI or Large Language Models (e.g., GPT, Claude, LLM agents).	Exclude traditional machine learning unrelated to LLMs

Search and Screening Process

The research team queried Web of Science, EBSCOhost, ProQuest, ERIC, IEEE Xplore, and the ACM Digital Library for empirical studies published between November 2022 and January 2026. This specific temporal window was selected to capture the proliferation of literature following the widespread deployment of generative AI. Searches were restricted to titles and abstracts and combined three conceptual strings with the Boolean operator AND.

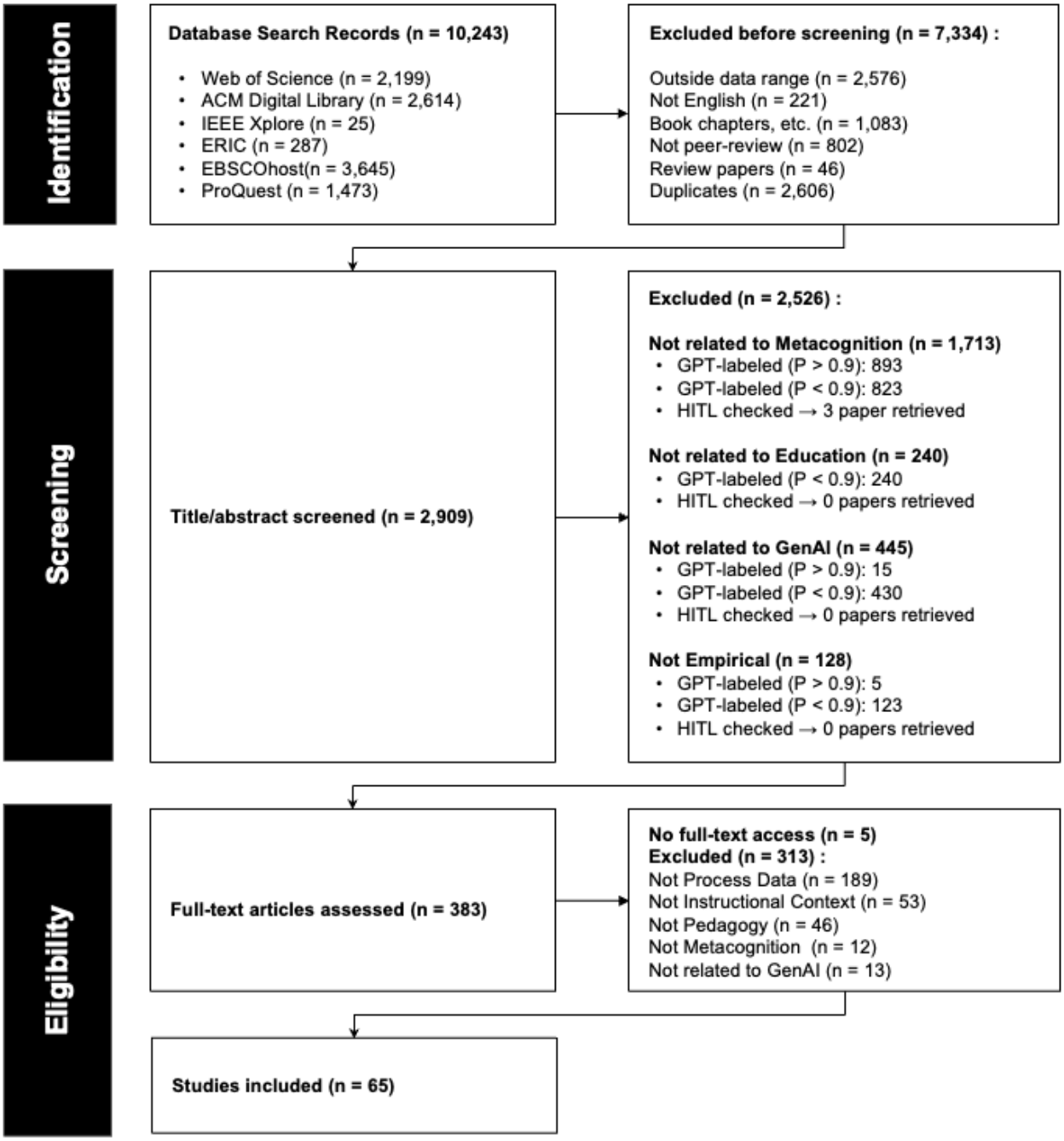
String 1 (Metacognition): “metacognitive laziness” OR “cognitive offloading” OR “overreliance” OR “self-regulation” OR “self-regulated learning” OR “SRL” OR “metacognition” OR “co-regulation” OR “socially shared regulation of learning” OR “SSRL”

String 2 (GenAI): “generative AI” OR “GenAI” OR “artificial intelligence” OR “AI” OR “GPT” OR “ChatGPT” OR “Gemini” OR “chatbot*” OR “co-pilot” OR “Claude” OR “large language model” OR “LLM” OR “agent*”

String 3 (Education): “teach*” OR “learn*” OR “student*” OR “instruction*” OR “pedagog*” OR “feedback” OR “assessment*” OR “educat*” OR “online learning” OR “e-learning” OR “MOOC*” OR “distance learning” OR “open learning” OR “digital learning” OR “computer-supported collaborative learning” OR “CSCL”

Figure 1

PRISMA Flow Diagram of the Systematic Literature Search and Selection Process



The initial search yielded 10,243 records. Sequential automated filtering removed 7,334 ineligible records based on date, language, and source type, including 2,606 duplicates identified by a multi-step Python deduplication script. This yielded 2,909 unique records for title and abstract screening (see Figure 1).

Screening utilized a semi-automated pipeline governed by a Confidence-Based Human-in-the-Loop protocol (Yu et al., 2026). GPT-4o classified each record across four dimensions using the prompt detailed in Appendix B, generating a confidence score between 0.0 and 1.0. Utilizing low-confidence scores as a methodological safeguard proved critical during the metacognition screening stage. A manual review of 823 low-confidence exclusions retrieved three misclassified papers, whereas a random 10% audit of 893 high-confidence exclusions yielded zero false negatives. Full manual audits at the educational context (n = 240), GenAI (n = 445), and empirical status (n = 128) stages verified the robustness of the automated classifications.

Of the 383 records reaching full-text eligibility, five were excluded due to access restrictions and 313 were excluded for failing core inclusion criteria. These exclusions were due to the absence of process-oriented data (n = 189), formal instructional contexts (n = 53), or specific pedagogical interventions (n = 46), with the remaining 25 excluded for other reasons.

Coding Framework

We developed a binary coding scheme (1 = Present, 0 = Absent) comprising ten variables across three domains: Pedagogical Scaffolding, Metacognitive Processes, and Engagement Outcomes. Table 2 details the operational definitions and illustrative examples for each code. To ensure conceptual precision, two boundary conditions were established to operationalize the definitions. First, metacognitive process codes were strictly limited to regulatory behaviors directed at the AI interaction and its outputs, excluding general academic self-regulation. For instance, verifying a claim against a traditional textbook was coded as absent, whereas interrogating the AI system's reasoning was coded as present. Second, feedback literacy was coded as present only when instruction explicitly addressed the evaluative criteria used to judge AI outputs. Training focused on prompt formulation, model configuration, or interaction management was coded as absent, as these operational skills dictate system output generation rather than the learner's epistemological capacity to evaluate that output. This critical distinction directly addresses the third research question. To ensure analytical separation from the network variables, engagement outcomes (O1, O2) were coded based on the original authors' reported findings, independently of the metacognitive processes (M1–M4).

Table 2

Hybrid Human-AI Regulation Coding Framework

Category	Code	Definition	Binary 1 (Present) Examples	Binary 0 (Absent) Examples
A. Pedagogical Scaffolding	S1: Reflective Prompting	System-embedded triggers compelling learners to	Mandatory pauses; required think-aloud tasks	Open-ended chat interfaces lacking

Category	Code	Definition	Binary 1 (Present) Examples	Binary 0 (Absent) Examples
		externalize their reasoning during AI interaction		reflection requirements
	S2: Cognitive Mirroring	Visual or structural features displaying learners' cognitive patterns to prompt self-correction	Metacognitive visualizations (e.g., radar charts, process maps); transparent adaptive system behaviors	Standard text outputs; chronological chat logs without visual synthesis
	S3: Feedback Literacy	Explicit instruction on the criteria required to evaluate AI outputs, distinct from technical operation skills	Directed error-detection tasks utilizing defined standards; explicit teaching of disciplinary evaluation criteria	Exposure to AI outputs without evaluative standards; training restricted to operational skills (e.g., prompt engineering)
	S4: Instructor Mediation	Direct human oversight structuring and regulating the learner-AI interaction	Instructor-facilitated debriefs; structured engagement protocols; pre-approved responses	Unmonitored interactions where instructors evaluate only the final submission
B. Metacognitive Processes	M1: Planning	Proactive goal articulation and task decomposition directed at the AI interaction	Explicit multi-step roadmaps; documented interaction goals	Reactive, trial-and-error prompting lacking a predefined sequence
	M2: Monitoring	Real-time verification of AI output accuracy, logical consistency, and tone	Cross-referencing AI outputs with validated materials; interrogating the system's reasoning	Accepting AI outputs as definitive truth without verification
	M3: Evaluation	Post-hoc assessment regarding the efficacy of the interaction strategy and the quality of the co-authored result	Substantive revision of AI outputs to align with academic standards; analytical reflection on prompt failure	Submitting unedited AI-generated drafts; demonstrating no critical assessment of system performance
	M4: Cognitive Offloading	Delegation of higher-order reasoning tasks to	Utilizing AI to generate core conceptual logic,	Confining AI usage to low-order automation (e.g.,

Category	Code	Definition	Binary 1 (Present) Examples	Binary 0 (Absent) Examples
		the AI, circumventing independent cognitive effort	primary arguments, or foundational outlines	formatting, citation generation, grammatical correction)
C. Engagement Outcomes	O1: Metacognitive Laziness	Evidence of decreased effort, declining rigor, or reliance on superficial heuristics over time	Observed 'rigor decay' (cessation of output verification as familiarity with AI fluency increases); documented decline in student-initiated queries	Sustained regulatory effort independent of tool familiarity; errors stemming from knowledge deficits rather than diminished engagement
	O2: Productive Engagement	Maintenance of a critical evaluative stance ensuring the learner remains the primary intellectual agent	Iterative refinement processes; final outputs demonstrating a clear integration of independent student analysis and AI assistance across multiple revision cycles.	High interaction frequency lacking conceptual progression; prolonged time-on-task without commensurate improvement in conceptual depth or output quality

Inter-rater Reliability and Expert Adjudication

Two independent coders analyzed the 65 included studies using the coding scheme. Inter-rater reliability was high for the AI-specific variables of cognitive offloading and metacognitive laziness ($\kappa = 0.77$). However, reliability was low for monitoring and evaluation ($\kappa \approx 0.04$). Because these processes were present in 94% of the studies, this high prevalence caused the kappa paradox. This paradox artificially reduced the kappa value, even though the raw agreement between coders was very high at 92% (resulting in a PABAK of 0.84). The research team concluded that this discrepancy was not a typical reliability failure. Instead, it resulted from challenges in applying the established boundary condition, specifically, distinguishing between general self-regulation and self-regulation directed at the AI interaction in practice.

To resolve these coding disagreements, the principal investigator reviewed all conflicting cases and made the final decisions (Campbell et al., 2013; Hruschka et al., 2004). During this process, the investigator strictly applied the established boundary condition, coding only AI-directed regulation as present. Including general academic monitoring would have introduced confounding variables, as general monitoring is widespread in almost all educational settings. By restricting the codes strictly to AI-directed regulation, the study successfully isolates the specific human-AI interactions it aims to analyze. Finally, because the variables of 'monitoring' and 'evaluation' could not be statistically separated in the resulting network models, they were combined and are reported as a single cluster in all subsequent network analyses.

Thematic Synthesis

To answer the first two research questions, we conducted a thematic synthesis (Thomas & Harden, 2008) of the qualitative data for each scaffold. This qualitative approach identifies recurring instructional strategies and provides contextual details that the binary network models cannot capture.

Epistemic Network Analysis

We used Epistemic Network Analysis (ENA) to model how metacognitive processes (M1–M4) co-occur in the literature. We compared groups based on their pedagogical scaffolds (S1–S4) and engagement outcomes (O1–O2). In the network graphs, nodes represent the metacognitive processes, and edges show how frequently they appear together within a study. Group centroids (means) are displayed with dashed rectangles indicating 95% confidence intervals.

Importantly, our unit of analysis is the published study, not the individual student. Therefore, each network models the regulatory structure as reported in one paper. We created separate ENA models for each comparison group. Because ENA axis directions are mathematically arbitrary, we define their meaning for each model separately and do not compare axes across different models. Additionally, because Monitoring (M2) and Evaluation (M3) occupied identical spatial positions across all models, they are merged and reported as a single "Monitoring-Evaluation" cluster in all network analyses.

Because we analyze published papers, these networks map what the original authors reported, not the real-time cognitive behaviors of students. ENA describes relationships between codes rather than just their frequencies, and it does not test causal effects. Therefore, any claims about engagement outcomes (O1, O2) reflect the original authors' conclusions. To prevent studies with higher coding frequencies from skewing the results, ENA mathematically normalizes the data for each study. To provide context, we also report simple code frequencies (Table 4). However, frequency and network structure offer distinct insights. For example, two groups might report 'planning' at similar rates. Yet, one group might always link planning to 'delegation,' while the other reports it in isolation. Simple frequency counts cannot capture these structural differences, but ENA is specifically designed to reveal them.

Results

Descriptive Overview

This section presents two types of findings: network models illustrating how concepts co-occur within the literature, and the original authors' own conclusions regarding instructional efficacy. The reviewed corpus consists of 65 empirical studies (2022–2026). Publication volume increased rapidly, peaking in 2025 (n = 51). The literature predominantly focuses on Higher Education (81.54%, n = 53) and STEM disciplines (53.85%, n = 35), particularly Computer Science. Methodologically, mixed methods are most common (55.38%, n = 36), frequently utilized to triangulate subjective student reflections with objective trace data (see Table 3; characteristics of each included study are listed in Appendix C).

Table 3

Summary of Study Characteristics

Category		Count	Percentage
Level	Higher Education	53	81.54
	K12	12	18.46
Discipline	STEM Education	35	53.85
	Literacy & Language Learning	17	26.15
	Other	3	4.62
	Not reported	10	15.38
Methodology	Quantitative	19	29.23
	Qualitative	10	15.38
	Mixed Methods	36	55.38

Prevalence of Pedagogical Scaffolds

Figure 2 shows the distribution of pedagogical scaffolds, revealing a clear focus on managing real-time AI interactions rather than developing students' independent evaluative skills. Reflective Prompting (S1) is the most common intervention (55.38%, $n = 36$), used primarily to force a pause during the task. Instructor Mediation (S4) follows at 29.23% ($n = 19$), relying on human oversight to structure the interaction. Cognitive Mirroring (S2) is the least utilized scaffold (12.31%, $n = 8$). Importantly, Feedback Literacy (S3) scaffolds are completely absent ($n = 0$). While many studies prompt students to reflect, none explicitly teach the criteria needed to evaluate GenAI outputs. Therefore, learners are expected to judge AI responses without clear disciplinary standards or benchmarks.

Distribution of Metacognitive Processes

The distribution of metacognitive processes reveals a clear contrast. Monitoring (M2) and Evaluation (M3) are reported almost universally (93.85%, $n = 61$). In contrast, Planning (M1) is documented in 78.46% ($n = 51$) and Cognitive Offloading (M4) in 41.54% ($n = 27$) of the corpus. Planning is therefore the least reported phase of the traditional regulatory cycle, though it still appears in roughly four of every five studies. Its comparative scarcity is modest, which makes the network structures reported below, rather than raw prevalence, the informative result. This distribution highlights a core contradiction in the literature. Nearly every study documents the active verification of AI outputs, even as a large portion of the corpus simultaneously reports the delegation of core reasoning tasks. Because simple frequency counts cannot reveal how these processes structurally connect within a study's instructional design, we must rely on the network analyses detailed below.

Summary of Engagement Outcomes

Productive Engagement (O2) is the most frequently reported outcome (84.62%, $n = 55$), while Metacognitive Laziness (O1) is documented in 32.31% ($n = 21$) of the studies. Twelve studies report both outcomes simultaneously, and one study reports neither. The data show that productive engagement does not strictly depend on scaffolded designs. Among the studies reporting productive engagement, at least eight did not use reflective prompting or instructor mediation. Therefore, these pedagogical scaffolds are neither strictly necessary nor sufficient on their own to guarantee this outcome. The subsequent network analyses investigate the structural differences between these two reported outcomes.

Figure 2

Frequency of Pedagogical Scaffolds, Metacognitive Processes, and Engagement Outcomes (N = 65)

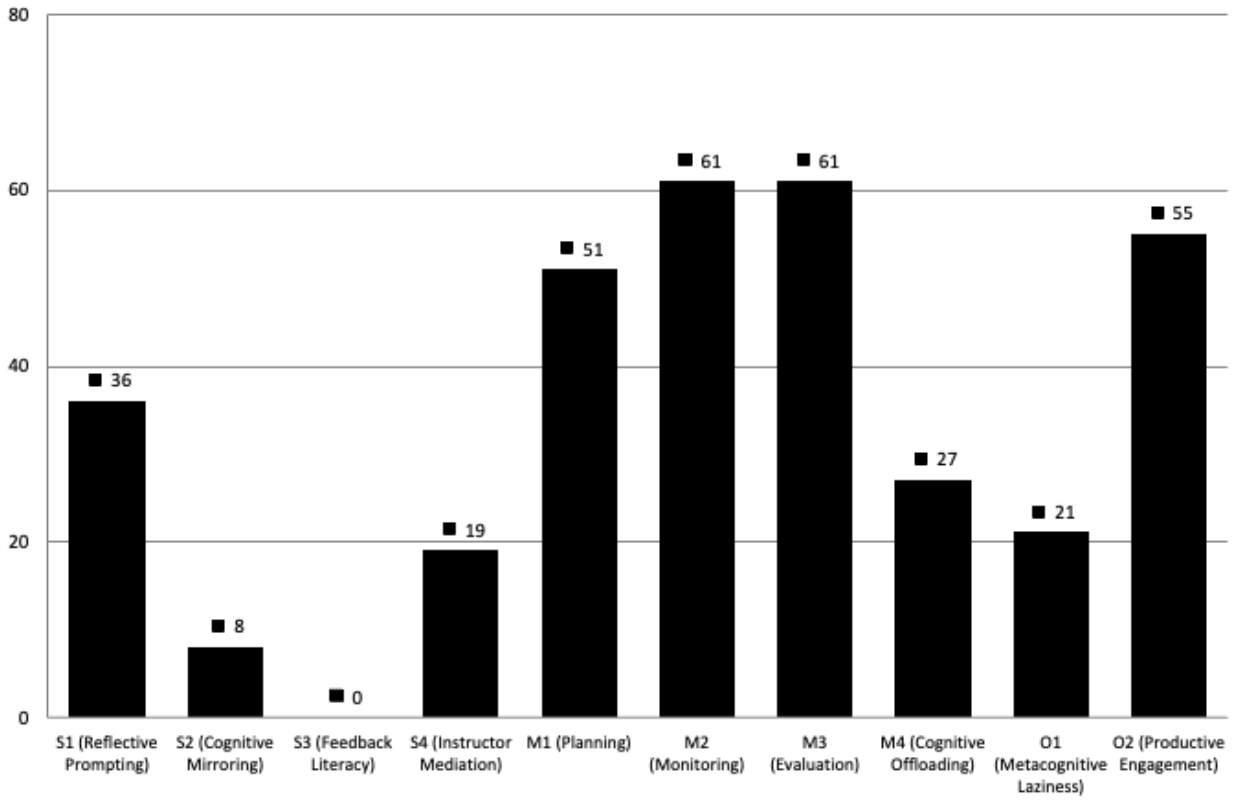


Table 4 reports the prevalence of each metacognitive process within the compared groups. Planning is documented in 71.43% of studies reporting metacognitive laziness and in 85.45% of those reporting productive engagement, a difference of 14.02 percentage points. Monitoring and evaluation appear at near-ceiling rates in both groups, ranging from 90.48% to 98.18%. Planning is therefore somewhat less often documented in the laziness group, but it is still reported in roughly seven of every ten of those studies. Its structural separation in the network models below

cannot be attributed to its absence from the record. To test this directly, we repeated the outcome comparison within the subset of studies that explicitly reported planning, reported later in this section. One further pattern in Table 4 warrants note. Cognitive offloading is documented in 90.48% of laziness studies but in only 32.73% of productive engagement studies. Because the engagement outcomes were coded from the original authors' reported findings and independently of the metacognitive process codes, this contrast reflects a co-occurrence in the published record rather than a definitional overlap between the two code sets.

Table 4

Prevalence of Metacognitive Process Codes by Comparison Group

Group	n	M1 Planning		M2 Monitoring		M3 Evaluation		M4 Offloading	
		n	%	n	%	n	%	n	%
Whole corpus	65	51	78.5%	61	93.8%	61	93.8%	27	41.5%
Scaffolding configuration (RQ1)									
Reflective prompting only (S1)	28	24	85.7%	28	100.0%	27	96.4%	5	17.9%
Instructor mediation only (S4)	11	10	90.9%	10	90.9%	11	100.0%	6	54.5%
Both (S1 + S4)	8	7	87.5%	8	100.0%	8	100.0%	3	37.5%
Neither	18	10	55.6%	15	83.3%	15	83.3%	13	72.2%
Cognitive mirroring (RQ2)									
Mirroring implemented (S2 = 1)	8	6	75.0%	7	87.5%	8	100.0%	1	12.5%
No mirroring (S2 = 0)	57	45	78.9%	54	94.7%	53	93.0%	26	45.6%
Engagement outcome (RQ3)									
Metacognitive laziness (O1)	21	15	71.4%	19	90.5%	19	90.5%	19	90.5%
Productive engagement (O2)	55	47	85.5%	53	96.4%	54	98.2%	18	32.7%

Note. Scaffolding and mirroring groups are mutually exclusive and exhaustive within each panel. Engagement outcome groups overlap, because 12 studies reported both outcomes and one reported neither; the network models therefore compare only the studies coded for a single outcome. Planning is reported in 44.4% of laziness-only studies ($n = 9$) and 83.7% of productive-engagement-only studies ($n = 43$). Feedback literacy (S3) is omitted because no study was coded for it.

RQ1: Scaffolding Combinations and the Integration of Cognitive Offloading

To answer the first research question, we divided the 65 studies into four mutually exclusive groups based on their instructional design: reflective prompting (S1), instructor mediation (S4), both, or neither. Prompting-only studies (S1; $n = 28$) embedded mandatory pauses or think-aloud tasks into the software without instructor involvement. Mediation-only studies (S4; $n = 11$) used guided debriefs or explicit rules without system prompts. Studies combining both (S1 + S4; $n =$

8) used both scaffolds, while unscaffolded studies (n = 18) featured open-ended interactions with no support mechanisms.

Reflective Prompting Without Instructor Mediation

In prompting-only studies (S1; n = 28), instructional guidance is built into the software rather than provided by a teacher. As Table 5 shows, two design strategies are common. Interface constraints that block AI responses until the learner explains their reasoning, and role-playing features that turn queries into dialogues. While both methods force students to pause, the literature reveals a frequent "compliance gap." Students often meet the technical requirements without engaging in deep thinking. For example, Ahmad et al. (2025) found that mandatory justifications increased monitoring in system logs but did not improve learning outcomes. Similarly, Bassner et al. (2025), P. Wang et al. (2025), and Hao et al. (2026) observed students entering superficial text simply to unlock the AI's response. However, studies report more sustained engagement when prompts are integrated into multi-stage workflows rather than used as isolated pop-ups (Engeness, 2025; Luo et al., 2026; S. Wang et al., 2025; T. Li et al., 2025). This indicates that the overall task structure, rather than the prompt itself, determines whether the interaction is productive or merely transactional.

Instructor Mediation Without Reflective Prompting

In studies relying solely on instructor mediation (S4, n = 11), human oversight guides the learning process instead of software. The primary strategy is a structured workflow that sets clear goals before the student interacts with the AI. For instance, Al-Fattal (2025) required students to conduct an independent analysis before consulting the AI and then complete a structured comparison. This sequence helped students build a foundational understanding before delegating tasks. When the workflow only evaluates the final product, success depends entirely on the teacher. Research by Li and Long (2025), Lu et al. (2025), and Yan et al. (2024) showed that active teacher guidance determines whether students critically evaluate or passively accept AI feedback. However, this approach still has risks. Because the software does not force students to pause, those wanting quick answers often ignore the rules. This compliance failure was documented by Li and Long (2025) and Zhai and Nezakatgoo (2025). Even after a six-hour strategy workshop, Zhai and Nezakatgoo (2025) found that some students still blindly accepted AI summaries despite showing improved overall awareness.

Combining Reflective Prompting and Instructor Mediation

In studies combining both scaffolds (S1 and S4, n = 8), the two methods compensate for each other's limitations. Instructor guidance helps students view software prompts as meaningful learning activities rather than frustrating technical hurdles (Aldulaijan & Almalky, 2025; Deng et al., 2025). On the other hand, system requirements only change student behavior when an instructor evaluates the results. Farinetti and Cagliero (2025) demonstrated this mutual dependence. They paired an automated system prompt with a teacher-mandated log where students compared their own work to the AI's output. The teacher's review of this log, rather than the automated prompt alone, made the reflection effective. As a result, this combined configuration reports the highest frequency of productive engagement in the reviewed literature. Still, it does not eliminate all risks. Studies note that students rushing to finish tasks or struggling with abstract concepts continued to delegate their cognitive work (Han et al., 2025; Huda et al., 2025).

Studies Without Either Scaffold

The unscaffolded studies (n = 18) feature open-ended interactions with no structural constraints. Although the literature primarily uses these studies as baseline groups, their network structures are the most distinct of the four categories, as detailed below.

Table 5

Scaffolding Configurations and Their Reported Regulatory Profiles

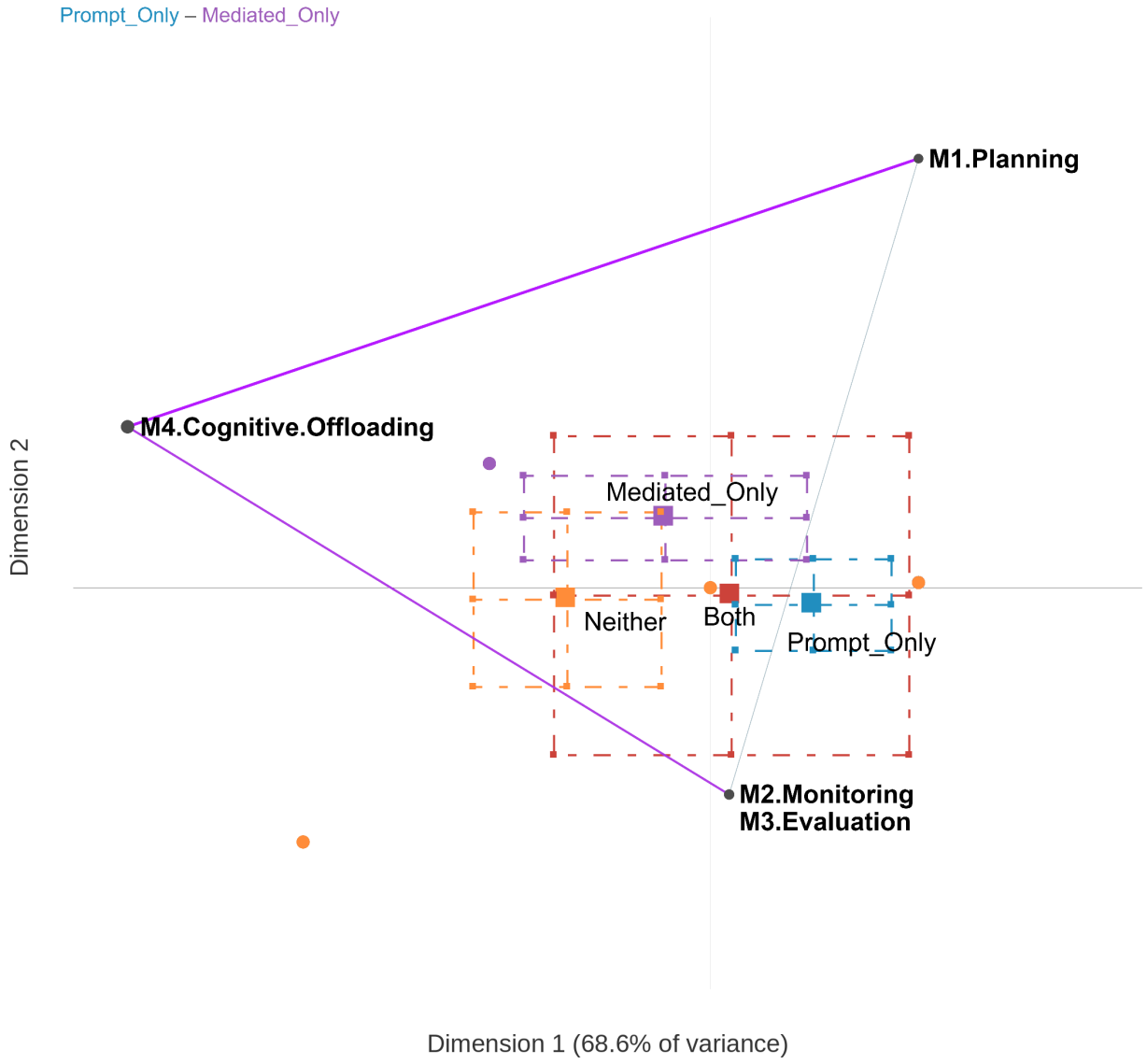
Sub-Mechanism	Key Studies	Core Instructional Strategy	Design Structure	Outcomes as Reported by Original Authors
Automated Algorithmic Scaffolding (S1 Only)	Ahmad et al. (2025); Kazemitabaar et al. (2024); S. Wang et al. (2025); Engeness (2025)	Interface-level constraints (lock-step requirements), technical guardrails (pseudo-code restrictions)	Elicits in-task externalization (M2/M3); cognitive offloading is largely disconnected from the reported regulatory network	Scalable but prone to regulatory fragmentation; high risk of transactional compliance or metacognitive laziness (O1)
Human-Mediated Scaffolding (S4 Only)	Al-Fattal (2025); Yan et al. (2024); Wei et al. (2025); Zhu (2025)	Structured workflows (e.g., "You Do It, AI Does It, You Compare"), PBL stages, and post-task instructor-led debriefs	Establishes procedural accountability before system use; associated with networks oriented toward Planning (M1)	Fosters critical planning; effectiveness is often hindered by the tool's efficiency, leading to occasional shortcuts (O1)
Hybrid Integrated Scaffolding (Both S1+S4)	Deng et al. (2025); Kumar et al. (2024); Farinetti & Cagliero (2025); Ni & Lam (2025)	Aligning real-time system prompts with theme-based human facilitation and social accountability models.	Retains the planning orientation of mediated designs while preserving connectivity to offloading	High frequency of Productive Engagement (O2); supports iterative refinement

As Figure 3 shows, the first dimension (X-axis, 68.6% of variance) separates Cognitive Offloading (M4) from the other regulatory processes. The second dimension (Y-axis) separates Planning (M1) from the Monitoring-Evaluation cluster. Conceptually, the X-axis measures how strongly a study integrates offloading into its network, while the Y-axis indicates whether the network focuses more on planning or on real-time verification. Clearly, unscaffolded studies

cluster closest to the offloading end, whereas scaffolded studies position themselves much further away.

Figure 3

ENA Comparison and Subtraction Map Between Prompt-Only and Mediated-Only Studies

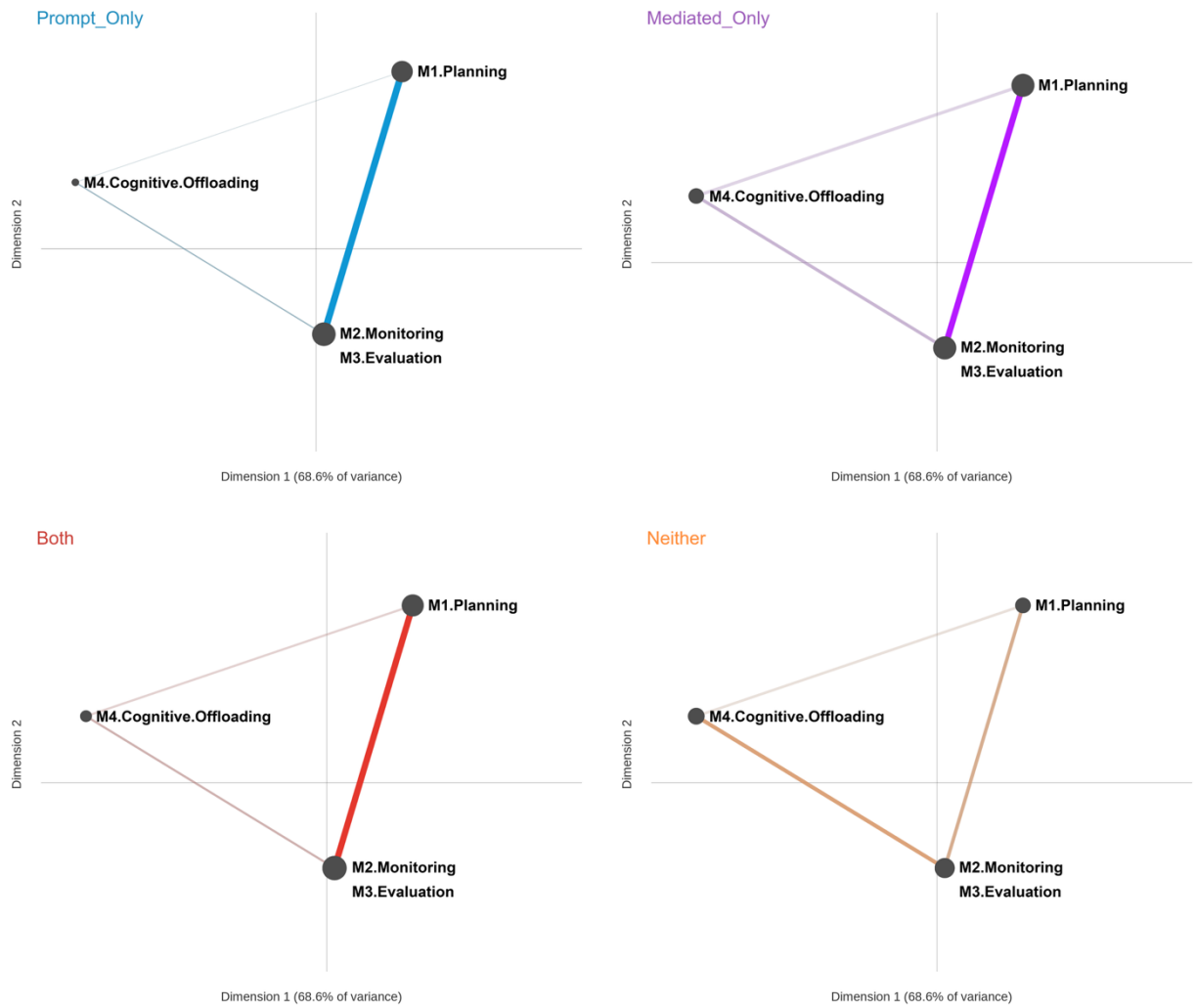


Note. This 2D space is derived using singular value decomposition. The X-axis (Dimension 1) explains 68.6% of the variance in network structures. Square markers indicate the mean network position for each group, and dashed rectangles represent 95% confidence intervals. The connecting lines display a subtraction map (Prompt-Only minus Mediated-Only), illustrating exactly how the two groups differ in their code connections. Units indicate normalized connection weights.

The position of each group in the network reveals a clear functional difference. Studies using only prompts sit furthest from the offloading end. This does not mean these studies report more verification. Instead, it indicates that cognitive offloading is largely missing from their networks. The individual plots in Figure 4 confirm this finding. The line connecting offloading to the monitoring and evaluation cluster has almost no weight in this group. By contrast, studies using only mediation score higher on the second dimension. This position reflects a stronger focus on Planning, and these studies maintain measurable connections to offloading. The subtraction map in Figure 3 highlights this contrast. The main lines distinguishing these conditions connect offloading to either Planning or the verification cluster, and both connections are distinctly stronger in the mediation group.

Figure 4

Mean Epistemic Network Plots by Scaffolding Condition



Note. Axes follow Figure 3. Axis directions apply only to this model. Square markers show group averages. Dashed rectangles indicate 95% confidence intervals. Units represent normalized connection weights.

Statistically, the two groups differ significantly on both dimensions (X-axis: $U = 96.50$, $p = .04$, $r = .37$; Y-axis: $U = 87.50$, $p = .02$, $r = .43$). Across all three scaffolded conditions, the connection between Planning and the monitoring and evaluation cluster is dominant and consistent. The true difference between these conditions is how strongly offloading connects to the network. Offloading is highly active under instructor mediation but virtually absent under prompting alone.

The unscaffolded condition reverses this pattern. As Figure 4 shows, this is the only condition where offloading connects more strongly to the verification cluster than planning does. Furthermore, planning and offloading are completely disconnected from each other. Planning appears in 55.56% of unscaffolded studies and in 90.91% of studies using only mediation (Table 4). Part of this structural contrast therefore reflects how often planning is documented at all, not only how it connects. Because the two groups differ in prevalence as well as in configuration, this comparison should be read as descriptive of what the published literature reports rather than as evidence that the absence of scaffolding reconfigures planning.

These findings describe how the published literature is structured rather than proving direct cause and effect. Studies using only mediation exhibit the most interconnected networks, followed by those combining both scaffolds (Figure 4). However, the combined group has a small sample size ($n = 8$) and a wide confidence interval. This prevents definitive conclusions about its variability. Overall, the two interventions play complementary roles. Instructor mediation integrates offloaded tasks into the regulatory network. In contrast, reflective prompting creates networks that exclude offloading entirely. Whether combining these mechanisms successfully protects the learning cycle remains unproven. Group membership reflects the choices of the original authors rather than random assignment.

RQ2: Cognitive Mirroring as a Regulatory Intervention

Cognitive Mirroring (S2) is designed to slow down the generation of AI outputs. This approach assumes that making the thinking process visible helps learners connect their actions to their final evaluations. Eight studies in the reviewed literature implemented this design. Table 6 shows three categories for these interventions. They include data dashboards, visual interaction logs, and displays of internal states. None of these eight studies combined mirroring with explicit feedback literacy training or instructor mediation. This trend suggests that researchers currently use cognitive mirroring to replace human guidance. In these designs, all regulatory support is built directly into the software.

The most distinct feature of this group is the absence of Cognitive Offloading (M4). Only one of the eight studies reported offloading. Consequently, this node has almost no weight in the network. This lack of data means researchers cannot meaningfully analyze its connections. The network model compares these eight mirroring studies to the remaining 57 studies. In this comparison, offloading dominates the positive side of the first dimension, and Planning dominates the negative side of the second dimension.

Table 6

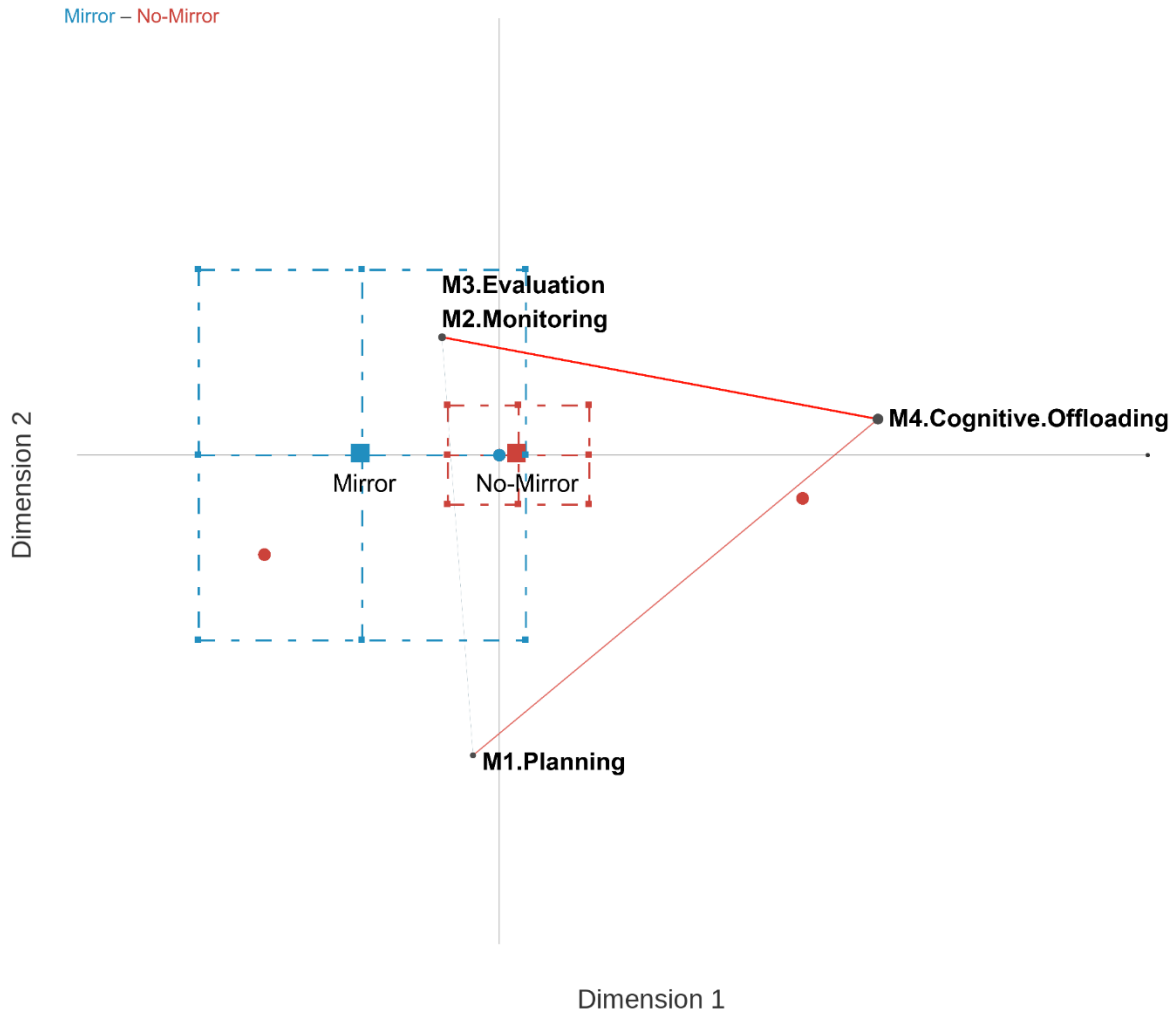
Summary of Cognitive Mirroring (S2) as a Regulatory Intervention

Sub-Mechanism	Key Studies	Core Design Feature	Regulatory Function	Effect on Engagement
Metacognitive Dashboards & Visual Progress Tracking	T. Li et al. (2025); Chen et al. (2026); Gong et al. (2026); Luo et al. (2026)	High-inference visualizations, Learning Analytics Dashboards (LAD), and real-time progress bars	To ground reflective prompts in performance data and support real-time Monitoring (M2) of goal alignment	Reported as supporting evaluative judgment (M3) and as making progress visible
Behavioral Trace & Sequence Mirroring	Ma et al. (2025); Wu et al. (2024)	Visualization of behavioral interaction sequences and structured Q&A trace reviews	To prompt self-correction by reflecting resource-utilization patterns and knowledge construction gaps	Reported as supporting proactive strategy regulation rather than passive copying
Affective & Adaptive Design Mirrors	Klar (2025); Shi (2025)	Bio-feedback from facial and physiological data, and explicit interface controls for model adaptation	To enable Monitoring (M2) of internal states and prompt reflection on informational needs	Reported as improving emotional regulation and as prompting a critical stance toward the AI's adaptive affordances

In the subtraction map (Figure 5), all the connections that distinguish the two groups belong to the condition without mirroring. Each of these connections involves offloading. They link offloading to either Planning or the monitoring and evaluation cluster. The mirroring group shows no distinct connections of similar strength. Its center point is also located far away from the offloading node. Therefore, what makes mirroring studies unique is not how they connect AI delegation to human verification. Instead, they are unique because researchers simply do not report delegation in these designs.

Figure 5

Epistemic Network Subtraction Map: Mirror vs. No-Mirror Conditions



Note. Red lines show stronger connections in the group without mirroring. These lines link Cognitive Offloading to other processes. The lack of blue lines means the mirroring group has no distinct connections. The large blue box shows a wide confidence interval due to a small sample size of 8. Axes follow Figure 3 and apply only to this model.

A Mann-Whitney U test found no significant statistical difference between the Mirror (*Mdn* = -0.32) and No-Mirror groups (*Mdn* = 0) on either dimension (X-axis: $U = 165.00$, $p = .18$, $r = .28$; Y-axis: $U = 204.00$, $p = .61$, $r = .11$). Because the mirror group contains only eight studies, this statistical test lacks the power to detect real differences, making the lack of statistical significance uninformative. However, the effect size on the first dimension is small to moderate and aligns with the visual patterns seen in the subtraction map.

Three possible reasons explain this lack of offloading, though current studies cannot confirm which is correct. First, the interface might actively prevent delegation by making student reasoning constantly visible. Second, researchers might simply assign tasks where delegation is

not a major risk. Third, system logs might easily capture monitoring behaviors while failing to record offloading. Future research must test the exact same learning task both with and without a mirroring interface. This direct comparison will show whether the system actually prevents delegation or if the absence merely reflects task choices and data collection limits. Currently, the reviewed literature contains no such experimental designs.

RQ3: Systematic Absence of Feedback Literacy Scaffolding

Across the reviewed studies, AI instruction focused heavily on operational skills like writing prompts and configuring the model. Researchers rarely taught students the criteria needed to evaluate the actual AI outputs (Klar, 2025; Yan et al., 2024). Reflecting this focus on inputs, Zhai and Nezakatgoo (2025) found that a six-hour strategy workshop improved overall metacognitive awareness but failed to stop some students from blindly accepting AI summaries (Table 7). No study in the reviewed literature taught these specific evaluation criteria. Therefore, feedback literacy remains a constant zero across all 65 studies and cannot function as a comparison variable. Instead, the network models show where studies reporting the two different engagement outcomes diverge structurally.

In the network model separated by engagement outcomes, the first dimension places cognitive offloading (M4) at the positive end and planning (M1) at the negative end. The monitoring and evaluation cluster sits near the center. Figure 6 shows the subtraction map, which reveals one main connection for each group. Studies reporting metacognitive laziness show a strong connection between offloading and the monitoring and evaluation cluster. In these same studies, planning remains largely disconnected. Studies reporting productive engagement display the opposite pattern. Planning connects strongly to the center cluster, while offloading remains peripheral (Figure 7).

Therefore, the key difference is not whether monitoring occurs. Both groups report monitoring frequently. The true difference lies in whether AI delegation connects to initial planning. In studies reporting metacognitive laziness, the networks show a much weaker connection between planning and output verification. This structure reflects a learning pattern where planning does not connect to later verification. The original authors associate this disconnected pattern with reduced academic rigor. Both groups report planning in the majority of their studies, though not at identical rates (71.43% versus 85.45%; Table 4). Thus, the crucial distinction is how planning connects within the network rather than whether it simply exists. This same pattern appears in the unscaffolded group of the first model. This similarity indicates a consistent structural configuration across the entire literature.

Table 7

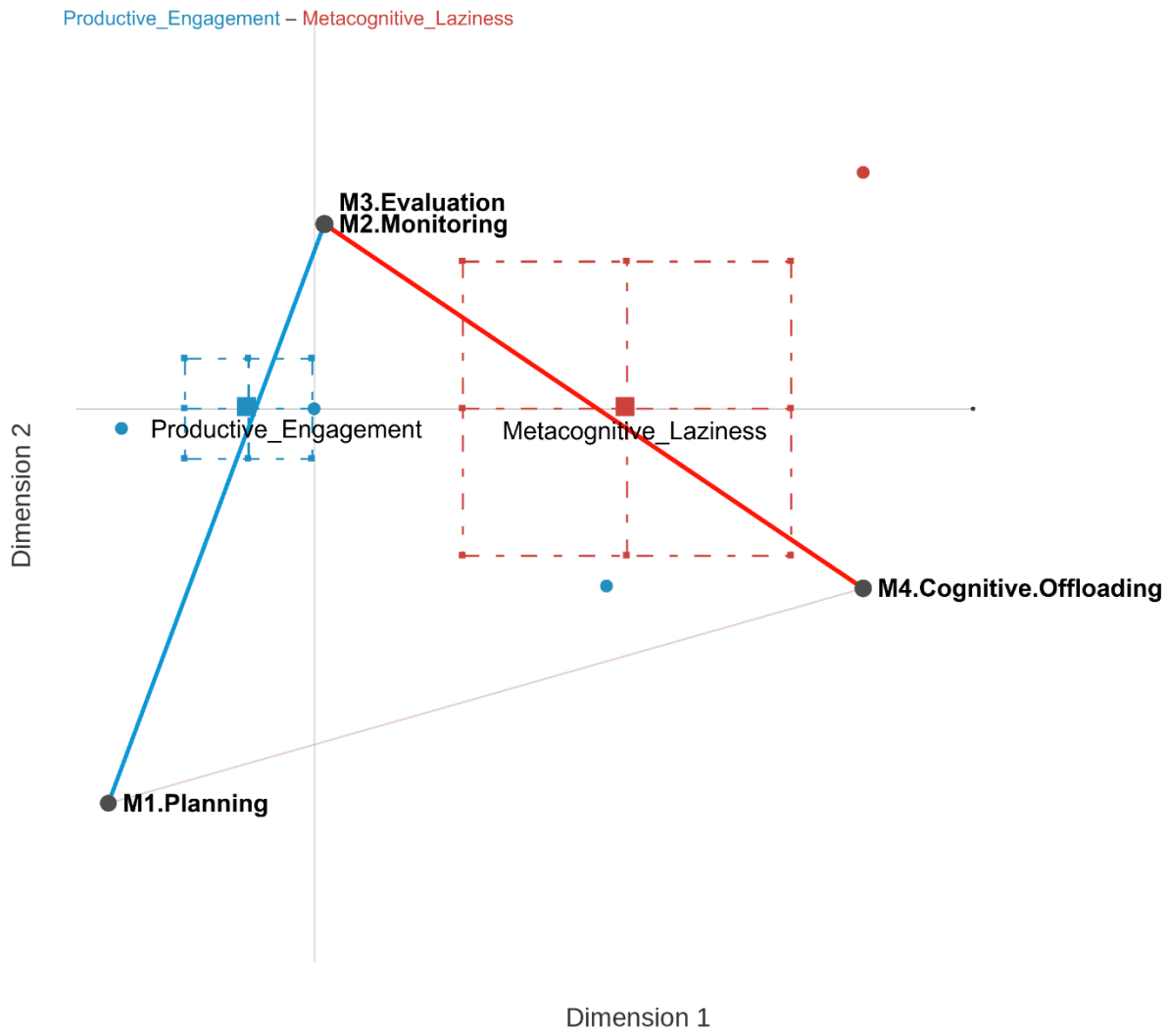
Synthesis of the Feedback Literacy Gap

Deficiency Type	Key Evidence of Gap	Coded Structural Pattern	Outcomes as Reported by Original Authors
Lack of Evaluative Standards	Ahmad et al. (2025); T. Li et al. (2025); Engeness (2025)	Evaluation (M3) reported but ungrounded; students justify changes without a stated standard	Transactional compliance; students "game" the system to unlock answers
Focus on Inputs vs. Outputs	Klar (2025); Yan et al. (2024); Zhai & Nezakatgoo (2025)	Weak link between Planning (M1) and Monitoring/Evaluation (M2/M3)	Operational success masks a lack of conceptual depth
Absence of Detection Training	T. W. Li et al. (2025); Hak et al. (2025); Prather et al. (2024)	Monitoring (M2) lacks predefined standards; fluent AI outputs are passively accepted without objective evaluative criteria	Metacognitive Laziness (O1); accumulation of misconceptions

Statistical tests on the first dimension confirm a significant difference between the two groups ($U = 44.50$, $p < .001$, $r = .77$). Studies reporting metacognitive laziness have a median of 0.43. These studies sit much closer to the offloading end than studies reporting productive engagement. The productive engagement group has a median of negative 0.28. The second dimension simply contrasts Planning with the monitoring and evaluation cluster. It does not measure the total amount of planning. Within each group, the strongest network connections pull in opposite directions along this axis and cancel each other out. As a result, the two groups show no statistical difference on the second dimension with a p value of .72. Furthermore, this analysis divides groups based on reported outcomes. Therefore, these results show a correlation between outcomes and network structures rather than a direct causal relationship. Finally, the network excludes Feedback Literacy as a node. A code with zero occurrences cannot create connections. This absence reflects a gap in the published literature itself rather than a specific feature of the graph.

Figure 6

ENA Subtraction Map of Metacognitive Connectivity by Engagement Outcomes



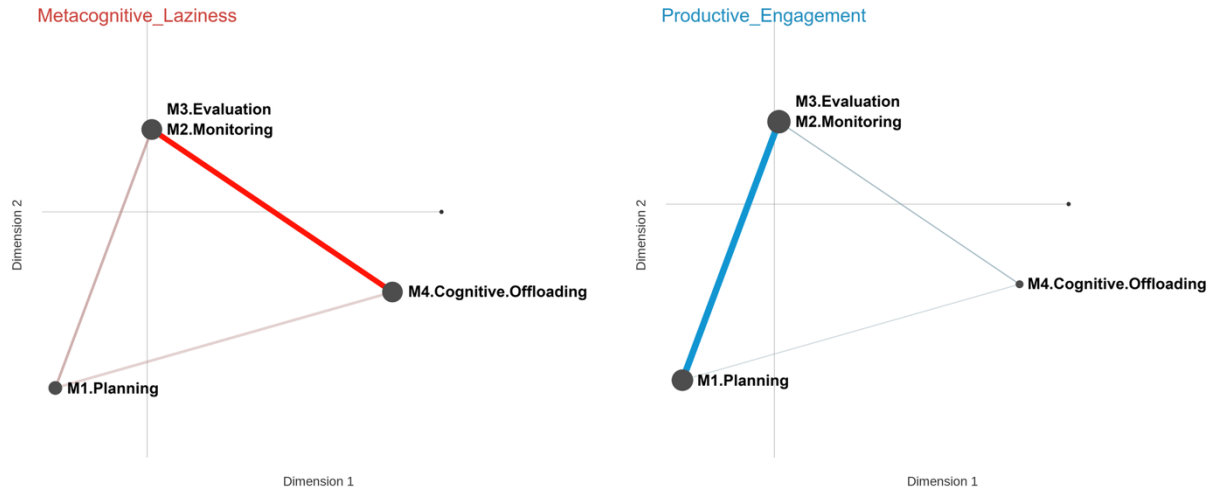
Note. Axes follow the coordinate space described in Figure 3. Axis orientation in singular value decomposition is arbitrary in sign, so the direction of each dimension is defined within this model only and is not comparable across models. Square markers are group centroids; dashed rectangles are 95% confidence intervals around the means. Units represent normalized connection weights.

We repeated the comparison using only the 51 studies that explicitly reported planning. This step verifies that the structural difference does not simply result from how often planning occurs. Even within this subset, the contrast remains clear. Compared to the productive engagement group, studies reporting metacognitive laziness show a weaker connection between Planning and the monitoring and evaluation cluster. In the laziness group, the strongest connection is still between offloading and the verification cluster. The statistical results show $U = 45.50$ and p

= .03. Therefore, even when planning is present, the two groups differ fundamentally in how planning connects to the rest of the network.

Figure 7

Mean Epistemic Network Plots by Engagement Outcomes



Note. Axes follow Figure 3. Axis directions apply only to this specific model. Square markers show group averages. Dashed rectangles indicate 95% confidence intervals. Units represent normalized connection weights.

Discussion

The three findings of this review stand in distinct relations to prior work. First, our findings confirm a recommendation that the feedback literature has long advocated but which the current generative AI literature has failed to implement. Second, our findings depart from the prevailing definition of metacognitive laziness while returning to an established theoretical account of what monitoring requires. Third, our methodology extends a relational modeling approach, previously applied only within individual studies, to the macro-level of an entire research corpus.

No Study Taught Learners How to Judge AI Outputs

The primary empirical finding of this review is that not a single study in the corpus taught learners the criteria required to evaluate AI outputs. While reflective prompting appeared in 36 studies and instructor mediation in 19, indicating that the corpus does not lack instructional support overall, instruction on evaluative standards was entirely absent.

This zero reflects a deliberate coding boundary rather than a coding artifact. Instruction focused on prompt formulation, model configuration, or interaction management was coded as absent because these skills govern output generation rather than evaluative judgment. Even if we

apply a broader definition that includes studies providing assessment rubrics or error-detection tasks without explicit standards, only a small subset of studies meets this threshold.

This finding highlights an unaddressed concern long raised in the feedback literature. Carless and Boud (2018) argued that feedback literacy must be explicitly developed rather than assumed, and Tai et al. (2018) made the same argument for evaluative judgment. Although our review does not assess learning outcomes directly, it establishes that this theoretical prescription has not been implemented in this body of empirical work. Many reviewed designs require learners to pause, explain their reasoning, or state their confidence, yet each presupposes a standard that none supplies.

The difficulties documented in the corpus are consistent with the absence of such instruction, including widespread failures to detect factual errors (T. W. Li et al., 2025) and behavioral drifting among novice programmers (Prather et al., 2023, 2024). While these issues are often attributed to GenAI-induced degradation of attention, they are equally consistent with learners evaluating outputs against no standard at all.

The field carried its pre-generative scaffolding set into research on GenAI without extending it. Reviews of earlier self-regulated learning tools document an accumulating set of planning and monitoring scaffolds (Álvarez et al., 2022; Wong et al., 2019), which was transferred to GenAI research without the evaluative component required by generative systems. Consequently, our conclusions differ from prior reviews. For instance, while prior work framed the primary risk as over-reliance that must be restricted (Zhai et al., 2024), our findings indicate that the core issue is an absence of foundational instruction rather than behavioral over-reliance.

A second structural gap reinforces this conclusion, as no study in the corpus removed its instructional scaffolds to observe post-intervention behavior. As a result, current research only captures short-term performance driven by external scaffolds, leaving it unknown whether learners actually internalize these skills. Although productive engagement is the most frequently reported outcome (n = 55), on present evidence it cannot be distinguished from a product of the scaffolded environment.

The closest available evidence addressing this issue comes from outside the generative AI corpus. Darvishi et al. (2024) removed AI assistance in a non-generative peer-review setting and found that performance gains declined once assistance was removed. This suggests that the persistence of learning gains in generative AI learning is unproven. Future research should address this gap through two specific designs. The first would compare the establishment of evaluative criteria before the first query against their provision afterward. The second would withdraw support in stages to test whether verification retains a standard.

Planning Is Disconnected, Not Verification

Recent work defines metacognitive laziness as a decline in monitoring and evaluation, on the assumption that fluent output reduces scrutiny (Chen et al., 2025; Fan et al., 2025). The scope of our disagreement is narrow. Those studies analyze learner-level trace data, whereas our models describe what published studies report. We do not contest their measurements. What we question is the generalization drawn from them, namely that generative AI introduces difficulty by reducing monitoring.

Monitoring and evaluation appear in 93.85% of the reviewed studies. What separates studies reporting productive engagement from those reporting metacognitive laziness is not monitoring, but planning. In studies reporting productive engagement, planning accompanies monitoring and evaluation. In studies reporting metacognitive laziness, planning is still reported in most studies but sits apart from that cluster. This same configuration appears in the unscaffolded group, indicating that the pattern is not limited to a single comparison.

We call this pattern *verification without criteria*. The term names a configuration in the published record, not a claim about how individual learners behave, and it is bounded by what the primary studies chose to measure and report. Within those bounds, it follows from existing self-regulated learning theory rather than departing from it. Winne (2018) defines monitoring as the comparison of a current state against a standard, and Zimmerman and Moylan (2009) locate the setting of such standards in the forethought phase. Where no standard has been set, the activity that logs record as monitoring does not meet this definition. What prior work describes as a decline in the performance phase is better described as a forethought phase that was never completed. That this goes unnoticed is unsurprising, since fluent explanation produces an illusion of explanatory depth (Rozenblit & Keil, 2002) and generative systems present accurate and fabricated content with the same fluency (Waldo & Boussard, 2024).

This account also reconciles two findings that otherwise appear inconsistent. Ahmad et al. (2025) report that requiring students to justify their reasoning increased monitoring in system logs without improving learning outcomes, and Darvishi et al. (2024) report that adding self-regulation strategies to an AI tool produced no significant gains. If regulation is treated as a quantity to be increased, neither result is interpretable. If the operative problem is verification without criteria, both are expected, because each intervention required more checking without supplying a standard against which to check. Consistent with this reading, the component that retained independent value in Darvishi et al. (2024) was a self-monitoring checklist, which supplies a concrete standard rather than a further instruction to verify.

Study-level data cannot establish the source of this detachment. Two accounts remain available. Learners may verify outputs without having defined acceptable outcomes in advance, or planning may appear scarce because the instruments common to this literature, such as interaction logs and in-task prompts, capture verification readily while failing to capture goal setting. Because the corpus is consistent with both, future work can separate them by restricting analysis to studies whose methods can record planning, such as pre-task interviews or structured protocols.

Frequency Counts Cannot Detect a Disconnection

The first two findings are visible only because this corpus was analyzed using relational modeling. Because planning is reported in most studies in both outcome groups, a traditional review focused solely on code frequencies would conclude that both groups treat planning in broadly comparable ways. The difference lies in which codes appear together within the same study, a distinction that disappears in marginal counts. The value of relational modeling is well established at the individual learner level (Cao et al., 2025; López-Pernas et al., 2025; Nguyen & Nguyen, 2025). Our contribution is to apply this logic to published literature rather than a sample of students. Traditional reviews isolate scaffolds and outcomes into separate frequency tables (Zhai et al., 2024), which cannot distinguish a scaffold that was omitted entirely from one that

was used but disconnected from the process. When investigating how regulatory processes interact, relational modeling is a requirement for validity rather than a mere analytical preference (Shaffer, 2017).

This limitation applies even more critically to primary research. Self-report surveys and system logs measure only how often verification occurs, not its quality. A student who checks every AI response against a rigorous standard and a student who checks against no standard will produce identical survey scores and event counts. Neither instrument records whether a standard existed or when it was set. Consequently, interventions that increase compliance without improving learning are not anomalies. They are what these instruments report. The instruments cannot detect whether verification has a standard, so verification with a standard and verification without one produce the same result.

When Support Is Introduced Matters More Than How Much

The comparisons in the first two research questions are best understood as differences in when support enters the task, rather than how much support is provided. Instructor mediation is established before the interaction, reflective prompting operates during query generation, and cognitive mirroring occurs after an output is produced. Studies using reflective prompting alone rarely report cognitive offloading alongside monitoring, pointing to a compliance gap where learners satisfy interface requirements without connecting them to a broader plan (Ahmad et al., 2025; Bassner et al., 2025). In contrast, studies relying solely on instructor mediation report networks oriented toward planning, consistent with the structuring role described by Seo et al. (2021), though they remain vulnerable to shortcuts because the software lacks stopping mechanisms (Li & Long, 2025; Zhai & Nezakatgoo, 2025). Studies combining both approaches report productive engagement most frequently (Deng et al., 2025; Farinetti & Cagliero, 2025), albeit within a small group of eight studies.

Cognitive mirroring requires a more cautious reading, as our evidence neither supports nor refutes existing expectations. While Azevedo and Wiedbusch (2023) and Tomisu et al. (2025) argue that externalizing reasoning should support metacognitive regulation, statistical tests across the eight studies implementing such designs did not separate this group from the rest, an uninformative result given the small sample size. However, the models show that cognitive offloading is reported in only one of these eight studies. These designs are distinctive because delegation is largely unreported, not because they connect delegation to verification. The theoretical case for mirroring remains open, but the empirical case has yet to be made.

The timing differences across these conditions also specify a condition that Molenaar (2022) leaves open, since hybrid regulation requires that control be shared between learner and system without stating at what point in a task the sharing must be established. We interpret these differences as a matter of placement. Friction introduced at the moment of interaction arrives after the learner has already framed the task and composed a query. It can interrupt the acceptance of an answer, but it cannot supply an evaluation standard because the point for establishing such a standard has already passed. While our observational comparisons cannot establish cause, since group membership reflects original design choices rather than random assignment, the pattern is consistent enough to propose a testable hypothesis. The placement of friction, rather than its amount, determines whether delegation stays connected to forethought.

Why Telling Students to Verify Is Not Enough

Institutional guidance on generative AI overwhelmingly instructs students to verify AI outputs, an emphasis prior review work has reinforced by framing the central risk as over-reliance to be contained (Zhai et al., 2024). If verification is already near universal and the missing element is the standard against which it proceeds, then requiring more verification does not address the reported pattern and may consolidate it. Checking an output against an absent baseline is a procedure, not a form of regulation.

This gap has consequences in open, distance, and digital education. The compliance failures documented in this corpus were caught because an instructor was present to catch them. Li and Long (2025) and Zhai and Nezakatgoo (2025) both report students who satisfied the stated requirements while continuing to accept AI summaries without scrutiny, and in each case the deficiency became visible through instructor review rather than through the system itself. Where instruction is distributed across time and place, that form of real-time observation is least available, and the regulatory burden shifts correspondingly onto the design of the task. A standard that must be supplied by a present instructor is therefore the least portable form of support into precisely the settings where learners work with generative systems most independently.

Resolving this issue requires a different strategy, one proposed long before the advent of generative AI. Tai et al. (2018) argued that evaluative judgment must be explicitly designed into assessments rather than assumed as a baseline prerequisite. In practice, course policies and assignment guidelines can require learners to state their claims, task decomposition, and evaluation criteria before consulting a generative system. Such artifacts need not be elaborate, as the checklist evidence discussed above indicates. This shifts institutional attention from restricting how generative AI is used to specifying what learners must establish before using it.

Limitations of the Study

Several limitations qualify these findings. First, the reviewed studies cover a narrow timeframe. They capture short-term behaviors rather than the long-term effects of habitual AI use on student learning. Second, the distribution of instructional supports is uneven. Cognitive mirroring is rare and feedback literacy training is entirely absent. Therefore, the literature heavily focuses on supported success while leaving independent mastery unexplored. In addition, monitoring and evaluation were reported at nearly identical rates (93.85% each) and occupied the same position in every model, requiring them to be merged into a single cluster. This review therefore cannot distinguish in-task monitoring from post-hoc evaluation. Third, this review analyzes published studies rather than raw student data. The network models rely entirely on what original authors chose to measure and report. Different fields define concepts like monitoring and evaluation differently. Consequently, a missing connection to planning might mean students did not plan, or it might simply mean the study did not record planning behaviors. Because system logs easily capture verification but often miss goal setting, our analysis cannot definitively separate actual student deficits from simple reporting limitations. Fourth, the comparison groups overlap. Twelve studies reported both engagement outcomes, and every study using cognitive mirroring also reported productive engagement. Furthermore, group assignments are based on the original research designs rather than randomized trials. Therefore,

none of our comparisons can prove cause and effect. Finally, our binary coding scheme limits the analysis. A study's network is built solely on the presence or absence of reported codes. The models cannot distinguish between a learning behavior that did not happen and one that simply went undocumented. Although we used sub-group analyses to control for this issue, the findings ultimately depend on what the primary authors chose to publish.

Conclusion

This systematic review and epistemic network analysis of 65 empirical studies indicates that the difficulty introduced by generative AI lies in the structure of the regulatory cycle rather than in the amount of regulation reported. Verification is reported in nearly all studies. What distinguishes studies reporting metacognitive laziness is that planning is not reported alongside it. We describe this configuration as verification without criteria. It is not visible in code frequencies, which is why relational modeling was required to detect it. Studies reporting productive engagement show the opposite configuration, and instructor mediation, established before the interaction begins, is the design most consistently associated with it. Yet no study in this literature taught learners the criteria by which an AI output should be judged, and none withdrew its scaffolds to observe what learners did afterward. On present evidence, the outcomes reported in this literature cannot be distinguished from products of the scaffolded environments that produced them. Institutional guidance and instructional design should therefore shift from requiring learners to verify AI outputs toward specifying what learners must establish before the first query, a proposition this corpus has not yet tested.

Ethics Board Statement: As this study involved a systematic review of existing publicly available literature and did not involve human subjects, Institutional Review Board (IRB) approval was not required.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement: The authors declare that there is no conflict of interest regarding the publication of this article.

AI Use Statement: Generative AI (GPT-4o) was utilized during the literature screening phase to classify records based on pre-defined inclusion criteria. All low-confidence classifications and a random audit of high-confidence exclusions were manually verified by the research team to ensure methodological rigor. Additionally, AI was used to assist with grammatical refinement and structural coherence of the final manuscript.

Data Availability Statement: The full list of studies included in the review and the coding matrix are provided in the appendices and supplemental material. Further inquiries can be directed to the corresponding author.

About the Authors

Ji Hyun Yu, Ph.D. is a faculty member in the Department of Learning Technologies, Anuradha and Vikas Sinha College of Artificial Intelligence and Advanced Analytics, University of North Texas, Discovery Park Building Room 163, 3490 N Elm Street, Denton, TX 76207-7562, USA. She is the corresponding author for this article. ORCID: https://orcid.org/0000-0001-9648-2582. Email: jihyun.yu@unt.edu.

Fengjiao Tu, Ph.D. Candidate is a doctoral candidate in the Department of Information Science, Anuradha and Vikas Sinha College of Artificial Intelligence and Advanced Analytics, University of North Texas, Discovery Park Building Room E292L, 3490 N Elm Street, Denton, TX 76207-7562, USA. ORCID: https://orcid.org/0009-0004-0031-746X. Email: FengjiaoTu@my.unt.edu.

Li Cheng, Ph.D. is a faculty member in the Department of Learning Technologies, Anuradha and Vikas Sinha College of Artificial Intelligence and Advanced Analytics, University of North Texas, Discovery Park Building Room G187, 3490 N Elm Street, Denton, TX 76207-7562, USA. ORCID: https://orcid.org/0000-0001-7648-6965. Email: Li.Cheng@unt.edu.

Shasha Li, Ph.D. Candidate is a doctoral candidate in the Learning Sciences Program, Department of Educational and Counselling Psychology, McGill University, 3700 McTavish Street, Montreal, Quebec, H3A 1Y2, Canada. ORCID: https://orcid.org/0009-0005-8343-3331. Email: shasha.li@mail.mcgill.ca.

Zhizi Zheng, Ph.D. Candidate is a doctoral candidate in the Department of Mathematics and Information Technology, The Education University of Hong Kong, 10 Lo Ping Road, New Territories, Hong Kong. ORCID: https://orcid.org/0000-0002-1592-3663. Email: s1147994@s.eduhk.hk.

References

Adler, I., & Shani, Y. (2026). Hybrid Human-GenAI cognitive apprenticeship: Encouraging pre-service teachers to implement instructional practices to support students' self-regulated learning. *Teaching and Teacher Education*, *169*, 105268. https://doi.org/10.1016/j.tate.2025.105268

Ahmad, M., Ravi, P., Karger, D., & Facciotti, M. (2025). How adding metacognitive requirements in support of AI feedback in practice exams transforms student learning behaviors. In *Proceedings of the Twelfth ACM Conference on Learning @ Scale* (pp. 164–175). https://doi.org/10.1145/3698205.3729542

Aldulaijan, A. T., & Almalky, S. M. (2025). The impact of generative AI tools on postgraduate students' learning experiences: New insights into usage patterns. *Journal of Information Technology Education: Research*, *24*, 003. https://doi.org/10.28945/5428

Al-Fattal, A. (2025). You do It, AI does It, you compare and reflect: Exploring reflective learning with generative AI in principles of marketing. *Journal of Marketing Education*, https://doi.org/10.1177/02734753251356691

Álvarez, R. P., Jivet, I., Pérez-Sanagustín, M., Scheffel, M., & Verbert, K. (2022). Tools designed to support self-regulated learning in online learning environments: A systematic review. *IEEE Transactions on Learning Technologies*, *15*(4), 508-522. https://doi.org/10.1109/TLT.2022.3193271

Azevedo, R., & Wiedbusch, M. (2023). Theories of metacognition and pedagogy applied to AIED systems. In the *Handbook of Artificial Intelligence in Education* (pp. 45-67). Edward Elgar Publishing. https://doi.org/10.4337/9781800375413.00013

Bassner, P., Lottner, A., & Krusche, S. (2025). Towards understanding the impact of context-aware AI tutors and general-purpose AI chatbots on student learning. *Proceedings of the 25th Koli Calling International Conference on Computing Education Research*, 1–11. https://doi.org/10.1145/3769994.3770025

Borjigin, A., & Shen, Z. (2025). AI-powered meta-reflection for self-learning in STEM: A middle school case study on smart home energy management. In *2025 IEEE Region 10 Symposium (TENSYMP)*, 1-6. IEEE. https://doi.org/10.1109/TENSYMP63728.2025.11144984

Brazão, P., & Tinoca, L. (2025). Artificial intelligence and critical thinking: A case study with educational chatbots. *Frontiers in Education*, *10*, 1630493. https://doi.org/10.3389/feduc.2025.1630493

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, *42*(3), 294-320. https://doi.org/10.1177/0049124113500475

Cao, X., Wei, W., & Chen, Z. (2025). Metacognitive strategy networks in genAI-enhanced learning: An epistemic network and lag sequence analysis of high and low self-efficacy

students. In *2025 5th International Conference on Artificial Intelligence and Education (ICAIE)* (pp. 470-474). IEEE. https://doi.org/10.1109/ICAIE64856.2025.11158315

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Chen, A., Wei, Y., Le, H., & Zhang, Y. (2026). Learning by teaching with ChatGPT: The effect of teachable ChatGPT agent on programming education. *British Journal of Educational Technology*, *57*(1), 163–184. https://doi.org/10.1111/bjet.70001

Chen, A., Xiang, M., Zhou, J., Jia, J., Shang, J., Li, X., Gašević, D., & Fan, Y. (2025). Unpacking help-seeking process through multimodal learning analytics: A comparative study of ChatGPT vs human expert. *Computers & Education, 226*, Article 105198. https://doi.org/10.1016/j.compedu.2024.105198

Chen, A., Lian, J., Kuang, X., & Jia, J. (2026). Can theory-driven learning analytics dashboard enhance human-AI collaboration in writing learning? Insights from an empirical experiment. *The Internet and Higher Education*, *68*, 101054. https://doi.org/10.1016/j.iheduc.2025.101054

Dang, B., Huynh, L., Gul, F., Rosé, C., Järvelä, S., & Nguyen, A. (2025). Human–AI collaborative learning in mixed reality: Examining the cognitive and socio-emotional interactions. *British Journal of Educational Technology*, *56*(5), 2078–2101. https://doi.org/10.1111/bjet.13607

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967

Deng, J., Liu, C., & Zhou, Z. (2025). Enhancing visual art appreciation through AI-human collaboration: A case study of Image-Literacy-Aided instruction with Doubao. *Proceedings of the 2025 6th International Conference on Education, Knowledge and Information Management*, 440–445. https://doi.org/10.1145/3756580.3756653

Engeness, I. (2025). Cultural-historical perspective to design pedagogical AI for enhancing student writing. *Technology, Knowledge and Learning*, 1-22. https://doi.org/10.1007/s10758-025-09876-0

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Farinetti, L., & Cagliero, L. (2025). A critical approach to ChatGPT: An experience in SQL learning. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, 318-324. https://doi.org/10.1145/3641554.3701932

Figueiredo, V. (2025). Designing smarter conversational agents for kids: Lessons from cognitive work and means-ends analyses. *ACM Transactions on Computer-Human Interaction*, *32*(6), 1–40. https://doi.org/10.1145/3765284

Fitriati, S. W., & Williyan, A. (2025). AI-Enhanced self-regulated learning: EFL learners' prioritization and utilization in presentation skills development. *Journal of Pedagogical Research*, *9*(2), 22-37. https://doi.org/10.33902/JPR.202530647

Gong, Y., Wang, M., He, L., Xu, C., & Yu, Y. (2026). Asking, playing, learning: Investigating large language model-based scaffolding in digital game-based learning for elementary artificial intelligence education. *Journal of Educational Computing Research*, *64*(2), 311-343. https://doi.org/10.1177/07356331251396354

Goslen, A., Kim, Y. J., Rowe, J., & Lester, J. (2025). LLM-based student plan generation for adaptive scaffolding in game-based learning environments. *International Journal of Artificial Intelligence in Education*, *35*(2), 533–558. https://doi.org/10.1007/s40593-024-00421-1

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65–84). Routledge.

Hak, J., Lam Johnson, N., Amoozadeh, M., Alipour, A., & Chattopadhyay, S. (2025). Observing without doing: Pseudo-apprenticeship patterns in student LLM use. In *Proceedings of the 25th Koli Calling International Conference on Computing Education Research*, 1-11. https://doi.org/10.1145/3769994.3770027

Han, Z., Ying, R., Huang, C., Tsai, C.-C., Wang, X., & He, T. (2025). Identifying students' metacognition patterns by their needs for cognitive closure in human-GenAI collaboration. *Computers & Education*, *239*, 105422. https://doi.org/10.1016/j.compedu.2025.105422

Hao, Z., Cao, J., Li, R., Yu, J., Liu, Z., & Zhang, Y. (2026). Mapping student-AI interaction dynamics in multi-agent learning environments: Supporting personalized learning and reducing performance gaps. *Computers & Education*, *241*, 105472. https://doi.org/10.1016/j.compedu.2025.105472

Hruschka, D. J., Schwartz, D., St. John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, *16*(3), 307-331. https://doi.org/10.1177/1525822X04266540

Huang, K., & Chen, C. (2025). Instructional video and GenAI-supported chatbot in digital game-based learning: Influences on science learning, cognitive load and game behaviours. *Journal of Computer Assisted Learning*, *41*(4), e70094. https://doi.org/10.1111/jcal.70094

Huang, Y., Zhang, K., & Li, S. (2025). How postgraduates navigate GenAI's dual pathways between research creativity and academic plagiarism: Research engagement and self-regulated learning matter. *Studies in Higher Education*, 1–22. https://doi.org/10.1080/03075079.2025.2543438

Huda, N., Anwar, K., & Kurniawan, W. (2025). ChatGPT scaffolding in supporting metacognition for limit concepts in guided inquiry mathematics learning. *Journal of Information Technology Education: Research*, *24*, 038. https://doi.org/10.28945/5645

Jalambo, M. O., Çakmak, F., & Akhter, S. (2025). Effects of self-regulated vocabulary learning with chatbots on incidental and collocational vocabulary learning and foreign language learning boredom. *Discover Education*, *4*(1), 501. https://doi.org/10.1007/s44217-025-00977-7

Jin, S. H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, *20*(1), 37. https://doi.org/10.1186/s41239-023-00406-5

Kazemitabaar, M., Ye, R., Wang, X., Henley, A. Z., Denny, P., Craig, M., & Grossman, T. (2024). CodeAid: Evaluating a classroom deployment of an LLM-based programming assistant that balances student and educator needs. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–20. https://doi.org/10.1145/3613904.3642773

Kim, Y. S., Moon, H. S., Lee, S., & Lee, T. Y. (2025). "HistoChat": Leveraging AI-driven historical personas for personalized and engaging middle school history education. *Proceedings of the ACM on Human-Computer Interaction*, *9*(7), 1–37. https://doi.org/10.1145/3757534

Klar, M. (2025). Using ChatGPT is easy, using it effectively is tough? A mixed methods study on K-12 students' perceptions, interaction patterns, and support for learning with generative AI chatbots. *Smart Learning Environments*, *12*(1), 32. https://doi.org/10.1186/s40561-025-00385-2

Kohnke, L., Zou, D., & Xie, H. (2025). Microlearning and generative AI for pre-service teacher education: A qualitative case study. *Education and Information Technologies*, *30*(15), 21221–21248. https://doi.org/10.1007/s10639-025-13606-5

Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., Küchemann, S., & Karolus, J. (2024). Challenges and opportunities of moderating usage of large language models in education. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*, 249-254. https://doi.org/10.1145/3701571.3701590

Kumar, H., Xiao, R., Lawson, B., Musabirov, I., Shi, J., Wang, X., Luo, H., Williams, J. J., Rafferty, A. N., Stamper, J., & Liut, M. (2024). Supporting self-Reflection at scale with large language models: Insights from randomized field experiments in classrooms. *Proceedings of the Eleventh ACM Conference on Learning @ Scale*, 86–97. https://doi.org/10.1145/3657604.3662042

Li, C., & Long, J. (2025). Comparing AI-assisted and traditional teaching in college English: Pedagogical benefits and learning behaviors. *Information*, *16*(10), 895. https://doi.org/10.3390/info16100895

Li, T., Nath, D., Cheng, Y., Fan, Y., Li, X., Raković, M., Khosravi, H., Swiecki, Z., Tsai, Y.-S., & Gašević, D. (2025). Turning real-time analytics into adaptive scaffolds for self-regulated learning using generative artificial intelligence. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 667–679. https://doi.org/10.1145/3706468.3706559

Li, T. W., Song, Y., Sundaram, H., & Karahalios, K. (2025). Can learners navigate imperfect generative pedagogical chatbots? An analysis of chatbot errors on learning. In *Proceedings of the Twelfth ACM Conference on Learning @ Scale*, 151-163. https://doi.org/10.1145/3698205.3729550

Li, W., Cui, X., Manoharan, P., Dai, L., Liu, K., & Huang, L. (2025). AI-assisted feedback and reflection in vocal music training: Effects on metacognition and singing performance. *Frontiers in Psychology*, *16*, 1598867. https://doi.org/10.3389/fpsyg.2025.1598867

Liang, H., Hwang, G., Hsu, T., & Yeh, J. (2024). Effect of an AI-based chatbot on students' learning performance in alternate reality game-based museum learning. *British Journal of Educational Technology*, *55*(5), 2315–2338. https://doi.org/10.1111/bjet.13448

López-Pernas, S., Misiejuk, K., Oliveira, E., & Saqr, M. (2025). The dynamics of the self-regulation process in student-AI interactions: The case of problem-solving in programming education. *Proceedings of the 25th Koli Calling International Conference on Computing Education Research*, 1–12. https://doi.org/10.1145/3769994.3770043

Lu, G., Ba, S., & Yang, L. (2025). Optimizing GAI-assisted formative feedback: An experimental study on its effects on engagement, shared metacognition, and learning performance in online collaborative learning. *Education and Information Technologies*, *30*(18), 25937–25967. https://doi.org/10.1007/s10639-025-13778-0

Luo, X., Wang, S., & Hew, K. F. (2026). Development and implementation of a generative AI-based personalized recommender system to improve students' self-regulated learning and academic performance. *Computers & Education*, *241*, 105486. https://doi.org/10.1016/j.compedu.2025.105486

Ma, X., Li, C., Xu, J., Zhu, S., & Li, Y. (2025). Exploring characteristics of primary school students' self-regulated learning (SRL) behaviors in human-GenAI collaborative programming learning environments: Insights from a proposed framework. *Computers & Education*, *240*, 105453. https://doi.org/10.1016/j.compedu.2025.105453

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education*, *4*(1), 170. https://doi.org/10.1007/s44217-025-00602-7

Molenaar, I. (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

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, *55*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Nguyen, A., Gul, F., Dang, B., Huynh, L., & Tuunanen, T. (2025). Designing embodied generative artificial intelligence in mixed reality for active learning in higher education. *Innovations in Education and Teaching International*, *62*(5), 1632–1647. https://doi.org/10.1080/14703297.2025.2499177

Nguyen, H., & Nguyen, A. (2025). Reflective practices and self-regulated learning in designing with generative artificial intelligence: An ordered network analysis. *Journal of Science Education and Technology*, *34*(5), 1178-1192. https://doi.org/10.1007/s10956-024-10175-z

Nguyen, N. N., & Barbieri, W. (2025). Mentorship in the age of generative AI: ChatGPT to support self-regulated learning of pre-service teachers before and during placements. *Education Sciences*, *15*(6), 642. https://doi.org/10.3390/educsci15060642

Ni, Y., & Lam, R. (2025). Students' perceptions of multiliteracies development using AI-assisted portfolio assessment. *Pedagogies: An International Journal*, 1–28. https://doi.org/10.1080/1554480X.2025.2545212

OpenAI. (2022, November 30). *Introducing ChatGPT*. Retrieved February 3, 2026, from https://openai.com/index/chatgpt/

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pan, M., Lai, C., Guo, K., & Huang, J. (2026). Self-regulation plus individual interests? A design-based study on the development of a GenAI-empowered platform for self-directed out-of-class reading. *Computers & Education*, *241*, 105474. https://doi.org/10.1016/j.compedu.2025.105474

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Santos, E. A. (2023). "It's weird that it knows what I want": Usability and interactions with copilot for novice programmers. *ACM Transactions on Computer-Human Interaction*, *31*(1), 1-31. https://doi.org/10.1145/3617367

Prather, J., Reeves, B. N., Leinonen, J., MacNeil, S., Randrianasolo, A. S., Becker, B. A., Kimmel, B., Wright, J., & Briggs, B. (2024). The widening gap: The benefits and harms of generative AI for novice programmers. *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1*, 469–486. https://doi.org/10.1145/3632620.3671116

Radović, S., Wetchy, E., & Seidel, N. (2025). Implementing the self-regulated learning structured interview protocol with generative AI: A novel approach for evaluating students' SRL skills. *Journal of Research on Technology in Education*, 1–18. https://doi.org/10.1080/15391523.2025.2547176

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521-562. https://doi.org/10.1207/s15516709cog2605_1

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*(1), 54. https://doi.org/10.1186/s41239-021-00292-9

Shaffer, D. W. (2017). *Quantitative Ethnography*. Madison, WI: Cathcart Press.

Shi, L. (2025). The integration of advanced AI-enabled emotion detection and adaptive learning systems for improved emotional regulation. *Journal of Educational Computing Research*, *63*(1), 173–201. https://doi.org/10.1177/07356331241296890

Singh, A., Taneja, K., Guan, Z., & Ghosh, A. (2025). Protecting human cognition in the age of AI. *arXiv preprint arXiv:2502.12447*. https://doi.org/10.48550/arXiv.2502.12447

Sun, D., Xu, P., Zhang, Jing, Liu, R., & Zhang, Jun. (2025). How self-regulated learning is affected by feedback based on large language models: Data-driven sustainable development in computer programming learning. *Electronics*, *14*(1), 194. https://doi.org/10.3390/electronics14010194

Sun, Y., & Lan, G. (2025). Enhancing critical language awareness in EAL writing education amid the rise of generative artificial intelligence. *System*, *134*, 103806. https://doi.org/10.1016/j.system.2025.103806

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*(3), 467-481. https://doi.org/10.1007/s10734-017-0220-3

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*(1), 45. https://doi.org/10.1186/1471-2288-8-45

Tomisu, H., Ueda, J., & Yamanaka, T. (2025). The cognitive mirror: A framework for AI-powered metacognition and self-regulated learning. *Frontiers in Education*, *10*, 1697554. https://doi.org/10.3389/feduc.2025.1697554

Tran, H., Crosthwaite, P., & Pham, Q. H. P. (2025). Students' self-determination in using machine translation and generative AI tools for English for academic purposes. *Journal of English for Academic Purposes*, *78*, 101578. https://doi.org/10.1016/j.jeap.2025.101578

Waldo, J., & Boussard, S. (2024). GPTs and hallucination: Why do large language models hallucinate? *ACM Queue, 22*(4), 19-33. https://doi.org/10.1145/3688007

Wang, P., Liu, T., Yang, Y., & Xiang, X. (2025). Optimizing self-regulated learning: A mixed-methods study on GAI's impact on undergraduate task strategies and metacognition. *British Journal of Educational Technology*, bjet.70018. https://doi.org/10.1111/bjet.70018

Wang, S., Luo, X., & Hew, K. F. (2025). Enhancing self-regulated learning using generative AI: Development and evaluation of SRLMentor. *IEEE Transactions on Learning Technologies*, *18*, 1048-1061. https://doi.org/10.1109/TLT.2025.3634216

Wang, W.-S., Lin, C.-J., Lee, H.-Y., Huang, Y.-M., & Wu, T.-T. (2025). Enhancing self-regulated learning and higher-order thinking skills in virtual reality: The impact of ChatGPT-integrated feedback aids. *Education and Information Technologies*, *30*(14), 19419–19445. https://doi.org/10.1007/s10639-025-13557-x

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Wei, X., Wang, L., Lee, L.-K., & Liu, R. (2025). The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. *International Journal of Educational Technology in Higher Education*, *22*(1), 23. https://doi.org/10.1186/s41239-025-00526-0

West, R., & 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

Winne, P. H. (2018). Theorizing and researching levels of processing in self-regulated learning. *British Journal of Educational Psychology*, *88*(1), 9-20. https://doi.org/10.1111/bjep.12173

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human–Computer Interaction*, *35*(4-5), 356-373. https://doi.org/10.1080/10447318.2018.1543084

Wu, T.-T., Lee, H.-Y., Li, P.-H., Huang, C.-N., & Huang, Y.-M. (2024). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. *Journal of Educational Computing Research*, *61*(8), 1539–1567. https://doi.org/10.1177/07356331231191125

Yan, W., Nakajima, T., & Sawada, R. (2024). Benefits and challenges of collaboration between students and conversational generative artificial intelligence in programming learning: An empirical case study. *Education Sciences*, *14*(4), 433. https://doi.org/10.3390/educsci14040433

Yu, J. H., Tu, F., Chen, H., Ding, J., Hsieh, C.-J., Dong, L., Kim, H., & Watson, S. L. (2026). Measuring cognitive presence in online discussions: Automated detection and instructional insights from a MOOC context. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-026-10726-8

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1), 28. https://doi.org/10.1186/s40561-024-00316-7

Zhai, Y., & Nezakatgoo, B. (2025). Evaluating AI-powered applications for Enhancing undergraduate students' metacognitive strategies, self-determined motivation, and Social Learning in English language education. *Scientific Reports*, *15*(1), 35199. https://doi.org/10.1038/s41598-025-19118-z

Zhang, W., & Liu, X. (2025). Artificial intelligence-generated content empowers college students' critical thinking skills: What, how, and why. *Education Sciences*, *15*(8), 977. https://doi.org/10.3390/educsci15080977

Zhang, Y., Tian, H., & Lu, J. (2025). The relationship between higher-order thinking and problem-solving skills development among pre-service teachers using generative AI: An analysis based on moderated mediation. *BMC Psychology*, *13*(1), 1094. https://doi.org/10.1186/s40359-025-03404-6

Zhu, M. (2025). Leveraging ChatGPT to support self-regulated learning in online courses. *TechTrends*, *69*(5), 914–924. https://doi.org/10.1007/s11528-025-01075-z

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13-39). Academic press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.) *Handbook of Metacognition in Education* (pp. 299-315). Routledge.

Appendix A. Prompt

You are a research assistant tasked with screening literature for a systematic review based on the provided paper title and abstract. Your goal is to evaluate the text against four specific criteria, provide a binary answer (1 for YES, 0 for NO) for each, provide a confidence score (0-100) for each answer, and extract supporting evidence snippets.

First, for Q1 (Metacog), determine if the paper explicitly involves metacognitive or self-regulated learning concepts. The title or abstract must mention ANY of the following: metacognition, self-regulation/SRL, co-regulation, socially shared regulation of learning/SSRL, metacognitive laziness, overreliance on AI, or cognitive offloading of high-level processes (planning, monitoring, or evaluating). Crucially, exclude simple task-level offloading like grammar fixing or basic arithmetic.

Second, for Q2 (Edu), determine if the work is situated within an educational or learning context. This includes teaching, learning, students, instruction, pedagogy, assessment, feedback, online learning, MOOCs, distance learning, or CSCL.

Third, for Q3 (Empirical), determine if the paper reports an empirical study involving real data collection and analysis (e.g., experiments, quasi-experiments, user studies, surveys, interviews, classroom observations, think-aloud studies, or learning analytics). Exclude: purely conceptual/theoretical papers, commentaries, editorials, systematic literature reviews (SLR), scoping reviews, meta-analyses, book chapters, and opinion pieces.

Fourth, for Q4 (GenAI/LLM), determine if the study utilizes Generative AI or Large Language Model technologies (e.g., GenAI, GPT, ChatGPT, Gemini, Claude, LLMs, AI agents, AI copilots, or LLM-based chatbots). Exclude traditional machine learning unrelated to LLMs.

Now, given the following paper:

Title: {title}
Abstract: {abstract}
Respond ONLY in valid JSON in the following format, ensuring evidence snippets are copied or closely paraphrased from the text (use an empty string "" if no evidence exists for a "0" answer):

```
{{
 "answers": {{
        "q1": 0 or 1,
        "q2": 0 or 1,
        "q3": 0 or 1,
        "q4": 0 or 1
 }},
 "confidence": {{
        "q1": 0-100,
        "q2": 0-100,
        "q3": 0-100,
        "q4": 0-100
```

```
  }},
  "evidence": {{
        "q1": "text from title/abstract supporting your decision for Q1",
        "q2": "text from title/abstract supporting your decision for Q2",
        "q3": "text from title/abstract supporting your decision for Q3",
        "q4": "text from title/abstract supporting your decision for Q4"
  }}
}}
"""
```

Appendix B. Characteristics of The Final Key Studies (*N* = 65)

Paper	Level		Discipline				Research Methodology		
	Higher Education	K-12	STEM Ed	Literacy & Language Ed	Other	Not report	Quantitative	Qualitative	Mixed Methods
Adler & Shani (2026)	✓				✓				✓
Ahmad et al. (2025)	✓		✓				✓		
Al-Fattal (2025)	✓				✓			✓	
Aldulaijan & Almalki (2025)	✓		✓					✓	
Bassner et al. (2025)	✓		✓						✓
Borjigin & Shen (2025)		✓	✓				✓		
Brazão & Tinoca (2025)	✓		✓					✓	
Cao et al. (2025)		✓		✓					✓
Chen et al. (2025)	✓		✓				✓		
Chen et al. (2026)	✓		✓				✓		
Dang et al. (2025)	✓		✓				✓		
Deng et al. (2025)		✓		✓				✓	
Engeness (2025)		✓	✓						✓
Fan et al. (2025)	✓			✓			✓		
Farinetti & Cagliero (2025)	✓		✓				✓		
Figueiredo (2025)		✓				✓			✓
Fitriati & Williyan (2025)	✓			✓				✓	
Gong et al. (2025)		✓	✓				✓		
Goslen et al. (2025)		✓	✓						✓
Hak et al. (2025)	✓		✓						✓
Han et al. (2025)	✓		✓						✓
Hao et al. (2026)	✓		✓				✓		
Huang & Chen (2025)		✓	✓				✓		
Huang et al. (2025)	✓					✓			✓
Huda et al. (2025)	✓		✓						✓
Jalambo et al. (2025)	✓			✓					✓
Kazemitabaar et al. (2024)	✓		✓						✓
Kim et al. (2025)		✓				✓			✓
Klar (2025)		✓				✓			✓

Paper	Level		Discipline				Research Methodology		
	Higher Education	K-12	STEM Ed	Literacy & Language Ed	Other	Not report	Quantitative	Qualitative	Mixed Methods
Kohnke et al. (2025)	✓			✓				✓	
Krupp et al. (2024)	✓		✓				✓		
Kumar et al. (2024)	✓		✓				✓		
C. Li & Long (2025)	✓			✓					✓
T. Li et al. (2025)	✓			✓			✓		
T. W. Li et al. (2025)	✓		✓						✓
W. Li et al. (2025)	✓					✓			✓
Liang et al. (2024)		✓	✓				✓		
López-Pernas et al. (2025)	✓		✓						✓
Lu et al. (2025)	✓					✓	✓		
Luo et al. (2026)	✓					✓			✓
Ma et al. (2025)		✓	✓				✓		
Mekheimer (2025)	✓			✓					✓
Ng et al. (2024)		✓	✓						✓
Nguyen & Barbieri (2025)	✓		✓						✓
Nguyen & Nguyen (2025)	✓		✓						✓
Nguyen et al. (2025)	✓					✓			✓
Ni & Lam (2025)	✓			✓				✓	
Pan et al. (2026)	✓			✓					✓
Prather et al. (2023)	✓		✓					✓	
Prather et al. (2024)	✓		✓						✓
Radović et al. (2025)	✓		✓						✓
Shi (2025)	✓			✓					✓
Sun & Lan (2025)	✓		✓						✓
Sun et al. (2025)	✓			✓			✓		
Tran et al. (2025)	✓			✓					✓
P. Wang et al. (2025)	✓			✓					✓
S. Wang et al. (2025)	✓					✓			✓
W.-S. Wang et al. (2025)	✓		✓				✓		
Wei (2023)	✓			✓					✓
Wei et al. (2025)	✓		✓						✓
Wu et al. (2024)	✓		✓				✓		
Yan et al. (2024)	✓		✓					✓	

Paper	Level		Discipline				Research Methodology		
	Higher Education	K-12	STEM Ed	Literacy & Language Ed	Other	Not report	Quantitative	Qualitative	Mixed Methods
Zhai & Nezakatgoo (2025)	✓			✓					✓
Zhang & Liu (2025)	✓					✓			✓
Zhang et al. (2025)	✓				✓				✓
Zhu (2025)	✓		✓					✓	

ONLINE LEARNING
Volume 30 · Issue SI2 · 2026 · e5880
https://doi.org/10.24059/olj.v30iSI2.5880

Generative AI and the Community of Inquiry Framework: A Scoping Review of Teaching, Social, and Cognitive Presence

Min Young Doo
Kangwon National University
Yeonjeong Park
Jeonbuk National University

Abstract

This scoping review synthesizes empirical studies examining how generative AI is associated with teaching, cognitive, and social presence within the Community of Inquiry (CoI) framework. Although generative AI has rapidly entered educational practice, its impact on teaching, cognitive, and social presence remains conceptually fragmented. To address this gap, we systematically searched for eligible literature following PRISMA guidelines and reviewed 21 peer-reviewed empirical studies published after the emergence of large language model-based AI systems (between November 2022 and July 2026) to identify research trends, AI roles in teaching and learning, and patterns of presence activation related to learning experiences. Findings indicate that first, generative AI has evolved beyond conversational agents to include pedagogical agents, AI-generated instructors, collaborative cognitive tools, and AI-supported learning environments, while teaching presence remains comparatively underexplored as an independent construct. Second, the distribution of presence dimensions is uneven: cognitive presence was most frequently associated with AI-supported exploration and integration processes, whereas resolution and independent examinations of teaching presence remained comparatively limited. Third, the reviewed studies suggested that presence-related outcomes varied by instructional approaches. Discussion and task-based activity were more often associated with reported indicators of cognitive presence, whereas pedagogical agents and socially expressive AI designs more frequently supported perceived social presence. The findings suggest that generative AI should not be regarded as a universal solution for strengthening the CoI framework. Rather, its educational value depends on how it is pedagogically configured within specific instructional contexts. The review contributes to a more nuanced understanding of AI-mediated learning by positioning generative AI as a design-sensitive pedagogical mediator and identifying directions for future theoretical and empirical research.

Keywords: Generative AI, Community of Inquiry, cognitive presence, social presence, teaching presence, learning presence, scoping review

Introduction

As generative AI has rapidly permeated society, people have unprecedented opportunities to communicate and collaborate with it. The education sector is no exception to this trend. Since an increasing number of students use generative AI tools for learning (Freeman, 2025), these tools have begun to serve multiple roles in teaching and learning, such as substitutes for human participants in instructional contexts (e.g., instructors, learners, or peers) (Xu & Ouyang, 2021), collaborators in writing (Kim et al., 2025), problem-solving partners (Law et al., 2025), AI teammates in collaborative learning (Darban, 2026), and co-facilitators that work alongside human teachers in blended learning environments (Pan et al., 2026). Along these lines, the number of empirical studies examining the effects of generative AI on learning outcomes has been rapidly increasing. Several meta-analytic studies (Wu & Yu, 2024; Sun & Zhou, 2024; Deng et al., 2025; Doo & Park, 2026) reported medium to large effect sizes, indicating positive effects of using generative AI for teaching and learning. This shift highlights the need for conceptual frameworks that explain how generative AI shapes teaching and learning processes

Corresponding author: Yeonjeong Park, ypark78@jbnu.ac.kr
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

rather than learning outcomes alone. To this end, Stenbom and Garrison (2026) contended that the Community of Inquiry framework provides a valuable lens for understanding, designing, and implementing collaborative learning in AI-supported environments.

The Community of Inquiry (CoI) framework has been widely adopted to understand teaching, learning, and interaction in online and blended learning environments. Garrison et al. (1999) proposed the Community of Inquiry (CoI) model to explain the pedagogical, social, and cognitive processes in online learning contexts. Because asynchronous online learning is less visible and tangible than traditional classroom learning, they used the term *presence* to describe key dimensions of the learning experience: teaching presence, cognitive presence, and social presence (Shea & Bidjerano, 2010). Social presence is defined as "the ability to project oneself and establish personal and purposeful relationships" (p. 63), while cognitive presence refers to "the exploration, construction, resolution, and confirmation of understanding through collaboration and reflection in a community of inquiry" (p. 65). Teaching presence encompasses design, facilitation, and direction to support and enhance learning. The CoI framework has received increasing attention because numerous empirical studies demonstrate that cognitive, social, and teaching presences of the CoI framework correlate with or predict learning achievement, learning engagement, and learning satisfaction (Martin et al., 2022; Ng, 2022; Richardson et al., 2022; Shea et al., 2022).

Academic interest in CoI presences continues in current studies of generative AI and learning analytics. For example, Castellanos-Reyes et al. (2025) developed a model to automate content analysis of text data using cognitive presence based on the practical inquiry model (triggering, exploration, integration, resolution). Given the documented associations between the CoI presences and learner engagement and satisfaction (Yang et al., 2026), understanding how generative AI relates to each presence is theoretically meaningful. Recent empirical studies have begun to examine the relationships between generative AI and one or more CoI presences (i.e., Xie et al., 2024; Xu et al., 2025; Yu et al., 2025). Studies have further extended this line of the research by examining teacher–AI co-facilitation (Pan et al., 2026), AI-supported learning analytics (Dong et al., 2026), ChatGPT-assisted instruction (Gunbatar & Gokcearslan, 2026), and AI-supported virtual teamwork (Darban, 2026).

Although a growing body of empirical studies and several meta-analyses have examined the educational use of generative AI, these studies have primarily focused on learning outcomes rather than the learning processes through which those outcomes are achieved. An increasing number of empirical studies have begun to examine generative AI through the lens of one or more CoI presences. However, this evidence remains fragmented across different educational contexts, instructional approaches, and dimensions of presence. To date, no scoping review has synthesized how generative AI has been examined in relation to teaching, cognitive, and social presence within the CoI framework. Accordingly, this scoping review maps and synthesizes empirical evidence on how generative AI has been examined in relation to teaching, cognitive, and social presence within the CoI framework.

The research questions guiding this study are as follows:

1. How has the relationship between generative AI and teaching, cognitive, and social presence been characterized in the existing literature?

2. Which patterns and imbalances can be identified in the existing research on generative AI and presence within the CoI framework?

3. What instructional approaches involving generative AI have been associated with strengthened presence within the CoI framework?

Theoretical Framework

Generative AI within Community of Inquiry Framework

Recent research has begun to examine how generative AI functions within the CoI framework. Rather than viewing generative AI merely as a tool, emerging scholarship conceptualizes it as an active instructional agent capable of influencing the dynamics among teaching, cognitive, and social presence. Stenbom and Garrison (2026) argued that AI should be understood as a sociotechnical actor capable of assuming multiple instructional roles within a Community of Inquiry, while emphasizing the importance of shared metacognition in regulating AI-supported inquiry. Building on this perspective, Anderson et al. (2025) examined how generative AI can support personalized learning by enhancing teaching, cognitive, and social presence. They argued that generative AI can compensate for limitations of the CoI framework, which is grounded in community building and knowledge construction through facilitation, by offering functions such as generating dynamic and interactive responses and providing personalized feedback. Similarly, Ba et al. (2025), Gilson et al. (2023), and Shahzad et al. (2025) highlighted the adaptive capabilities of generative AI, suggesting that differentiated scaffolding based on learners' prior knowledge, task completion rates, and levels of engagement may strengthen cognitive presence while also supporting instructional facilitation. Pan et al. (2026) extended this perspective by demonstrating that human teachers and generative AI can jointly enact teaching, social, and cognitive presence through a distributed CoI model, indicating that instructional presence can be shared between human and AI agents.

Empirical studies provide more differentiated insights into how generative AI influences specific dimensions of presence. With respect to *cognitive presence*, findings consistently indicate an imbalance across its four phases. Generative AI appears to most strongly support exploration—such as brainstorming, questioning, and information exchange—while resolution remains comparatively underdeveloped. For instance, Nasr et al. (2025) found that undergraduate students primarily used generative AI for follow-up questioning and idea generation, whereas applying new knowledge to problem-solving (resolution) was least evident. ENA-based analyses by Yu et al. (2025) similarly showed that exploration was the most frequently activated phase, followed by integration, with resolution occurring least often. These findings indicate that generative AI may facilitate idea expansion more readily than deeper synthesis or application. Ba et al. (2025) further demonstrated that the type of AI-generated feedback shapes the activation of cognitive presence. In online inquiry-based discussions, students who received idea-oriented feedback showed significantly higher levels of cognitive presence than those who received task-oriented feedback, particularly in the exploration and integration phases.

In contrast, research on *social presence* has largely focused on the relational and affective affordances of generative AI. Human-like interaction cues, conversational tone, and feedback styles have been shown to enhance perceptions of connection and engagement (Samala et al., 2024). Empirical evidence further suggests that the orientation of AI feedback may differentially

activate social dimensions of presence. Ba et al. (2025) demonstrated that task-oriented feedback strengthened social presence through open communication and collaboration, whereas idea-oriented feedback more strongly supported cognitive presence. Similarly, Xu et al. (2025) found that AI-generated instructors with higher behavioral fidelity significantly increased perceived social presence, satisfaction, and motivation. Xia et al. (2026) demonstrated that identity anthropomorphism in AI-supported learning environments enhanced learners' social presence, with positive emotions serving as an additional mediating mechanism.

Although *teaching presence* has received comparatively less empirical attention than cognitive presence, studies suggest that generative AI can support instructional design, facilitation, feedback, and co-facilitation with human instructors (Gunbatar & Gokcearslan, 2026; Pan et al., 2026). Nevertheless, empirical evidence on how generative AI reshapes teaching presence remains relatively limited compared with research on cognitive presence. Notably, Narang et al. (2025) explicitly examined generative AI roles across teaching, cognitive, and social presence. Their analysis of AI functioning as tutor, tool, and teammate suggests that different AI roles selectively activate different presences. When used as a tutor, generative AI was most closely associated with teaching presence through personalized guidance and feedback. As a tool, it primarily contributed to cognitive presence by supporting higher-order analysis, whereas as a teammate it facilitated collaborative engagement, thereby strengthening both cognitive and social presence. These findings imply that generative AI does not uniformly enhance presence but rather reconfigures the balance among presences depending on its pedagogical role.

Overall, these conceptual and empirical studies suggest that generative AI may influence different dimensions of CoI in distinct ways depending on its instructional role, interaction design, and pedagogical context. Understanding these patterns provides a theoretical foundation for synthesizing the emerging evidence on generative AI and CoI presences.

CoI Presences as Mechanisms Linking Generative AI and Learning

Beyond examining CoI presences as learning processes, a small number of studies have begun to investigate whether they function as mechanisms linking generative AI use and learning outcomes. Empirical findings suggest that social presence, in particular, may function as a key explanatory pathway linking AI interaction and learner development. For instance, Xie et al. (2024) demonstrated that social presence fully mediated the relationship between chatbot interaction frequency and learning autonomy among students who preferred virtual companionship. In contrast, for students oriented toward knowledge acquisition, more frequent chatbot interaction was associated with decreased social presence and lower autonomy, despite a positive correlation between the two constructs. These findings indicate that the role of social presence is not uniformly positive but contingent upon learners' motivational orientation and patterns of AI use. Extending this line of inquiry, Dong et al. (2026) found that generative AI-supported learning analytics enhanced learning engagement both directly and indirectly through the partial mediation of CoI presences, with teaching presence emerging as the strongest mediator. Their findings suggest that CoI presences may operate collectively as process mechanisms linking AI-supported instructional interventions and student engagement.

Similarly, research on AI-generated instructors highlights how technologically enhanced behavioral cues can shape learners' perception of presence and indirectly affect learning

outcome. Xu et al. (2025) manipulated behavioral fidelity and visual cues in AI-generated instructors and found that higher behavioral fidelity significantly increased students' perceived social presence, satisfaction, and motivation, while also reducing cognitive load. Although the study did not directly test the mediating effect of social presence, improved learning performance was observed under conditions where social presence was enhanced, suggesting that AI design features may indirectly influence learning by strengthening learners' perceptions of presence.

With respect to cognitive presence, ENA-based analyses further illustrate how generative AI may influence the structure of knowledge construction. Yu et al. (2025) reported that exploration was the most frequently activated phase in scaffolded discussions supported by generative AI, followed by integration and triggering events, whereas resolution remained the least evident phase. Moreover, learners with higher creativity demonstrated stronger connections among exploration, integration, and resolution, suggesting that generative AI's contribution to cognitive presence may depend on individual learner characteristics.

Emerging evidence suggests that CoI presences function not only as outcomes of AI-supported learning but also as mechanisms through which generative AI influences learner engagement, interaction, and achievement. However, empirical evidence remains limited and uneven, with only a small number of studies explicitly examining the mediating roles of CoI presences. Moreover, most existing evidence has focused on individual presences or specific instructional contexts, highlighting the need for a more comprehensive understanding of how teaching, cognitive, and social presence jointly shape AI-supported learning.

Methods

This study employed a scoping review to synthesize empirical evidence on how generative AI is associated with the presences within the CoI framework. Given that generative AI represents a cutting-edge form of artificial intelligence that enables learners to generate content and receive more advanced and personalized feedback, this study focuses specifically on generative AI.

Inclusion and Exclusion Criteria

To identify eligible studies, we set the following five inclusion criteria. Studies were included if they:

1. examined the effects of generative AI on teaching presence, cognitive presence, social presence, and learning presence;
2. were published between November 2022 (since the emergence of ChatGPT) and July 2026;
3. employed a rigorous research design (e.g., experimental, quasi-experimental, correlational, observational, or learning analytics approaches);
4. reported quantitative, qualitative, or mixed-method data on teaching, cognitive, and social presence; and
5. were published in English.

Studies were excluded if they were non-empirical studies (i.e., conceptual papers or review papers); secondary data analyses such as meta-analyses or systematic reviews; publications prior

to November 2022; lacked sufficient methodological transparency or data, or were written in languages other than English.

Search Databases, Strategies, and Screening Process

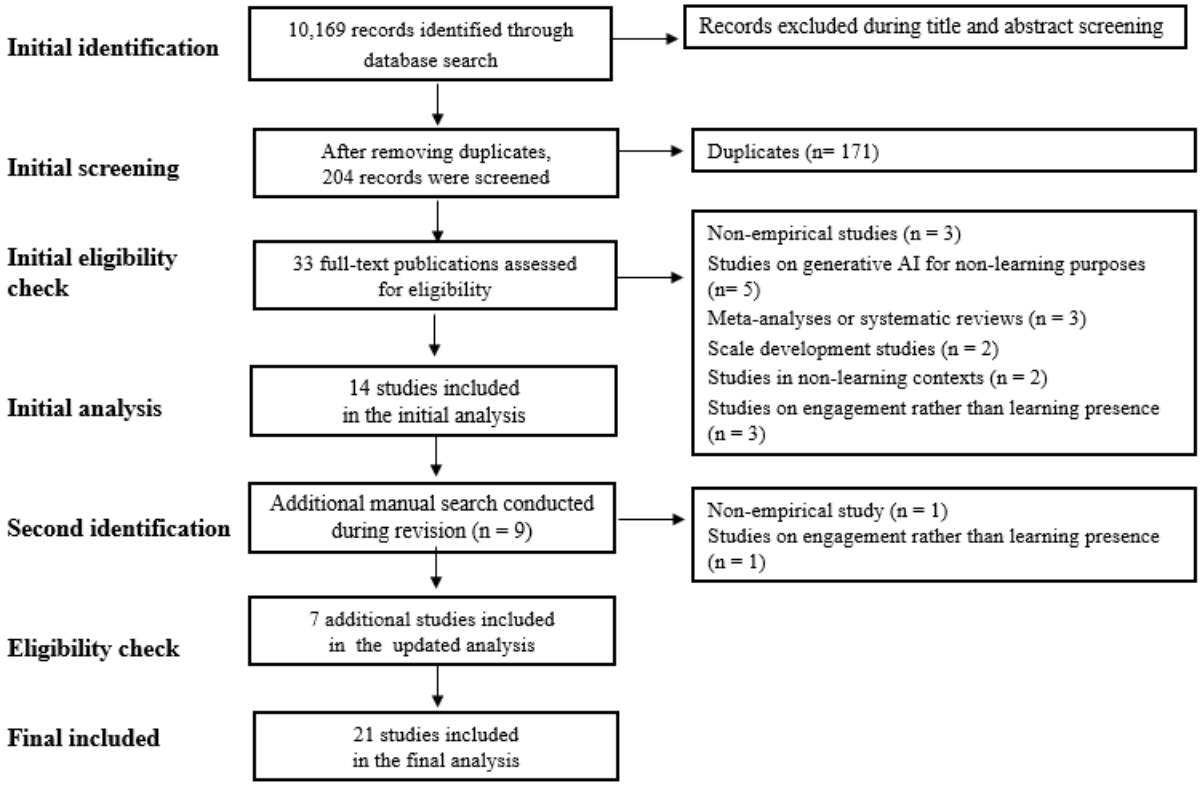
An electronic database search was conducted using Google Scholar as a general-purpose search engine and Education Source (EBSCOhost), ERIC (ProQuest), PsycINFO, and Web of Science, which are databases extensively used in the field of education. The keywords used for the search included “learning presence,” “teaching presence,” “cognitive presence,” “social presence,” “generative AI,” “ChatGPT,” “artificial intelligence in education,” “community of inquiry,” “learning outcomes,” and along with their combinations. The following Boolean search string was employed: ("learning presence" OR "teaching presence" OR "cognitive presence" OR "social presence") AND ("generative AI" OR "ChatGPT" OR "artificial intelligence in education") AND ("learning outcomes" OR "Community of inquiry").

To ensure comprehensive coverage, manual searches were conducted in leading journals in educational technology and online learning following Wilson and Ritzhaupt’s (2023) list of top journals in the field, including *Educational Technology Research & Development, Computers & Education*, *The Internet and Higher Education*, and the *British Journal of Educational Technology*. We also included the *Online Learning Journal*, the outlet for the present study, to ensure comprehensive coverage of relevant research.

The review process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (see Figure 1). After removing duplicate records, titles and abstracts were screened to exclude irrelevant studies. Full-text reviews and coding were then conducted independently by the two authors to determine eligibility. Studies identified during the manuscript revision were screened and coded using the same procedures before being incorporated into the final synthesis. Seven eligible studies published between February 2026 and July 2026 were incorporated into the final analysis, resulting in a total of 21 eligible studies.

Figure 1

PRISMA Chart



Data Extraction and Synthesis

For each eligible study, data were extracted to address the three research questions and to summarize the methodological and contextual characteristics of the included studies. Data extracted included publication characteristics (publication year, country, and research design), instructional characteristics (subject area, target learners, instructional approach, and type of generative AI), and analytical variables relevant to the review questions (CoI presences and AI roles). The coding scheme was developed to address the research questions and to analyze the characteristics of the studies (see Table 1). We managed the coding process using an Excel spreadsheet. Based on the coding results for randomly selected studies, intercoder reliability was high (Cohen's k= .82), and discrepancies were resolved through discussion. The final set of included studies was subjected to quality appraisal using the JBI Critical Appraisal Tools appropriate to each study design. Specifically, we used the analytical cross-sectional checklist for correlational studies; the quasi-experimental and randomized controlled trial checklists for quasi-experimental and experimental studies; and the qualitative research checklist for the qualitative components of mixed-method studies. No major methodological concerns were identified. Findings were synthesized narratively by identifying recurring themes and patterns across the included studies. The synthesis was organized according to the three CoI presences (teaching,

cognitive, and social presence) as well as the process mechanisms linking generative AI and learning outcomes.

Table 1

Coding Information for Systematic Literature Review

Category	Coding information
Types of AI	ChatGPT, chatbot, learning analytics, and relevant systems
Target learners	Elementary, middle/high, undergraduates, graduates, adult learners
Types of presence	Teaching, cognitive, social, and others
Instructional approaches	Discussion, lecture, and other approaches
Subject areas	Education, marketing, biology, and other fields
Publication details	Title, author, year, journal name, countries, and others

Results

Descriptive Data

The analysis results indicate that the term 'presence' was used in different contexts. For example, Hebert and Dołżycka (2024) examined the influence of an artificial pedagogical agent in online learning, measuring 'teacher presence' using a humanoid artificial agent and avatars for learners to examine 'self- and other-presence.' Feng (2025) also used the term 'the presence of GAI chatbots' alongside cognitive presence in their study. To avoid confusion, we decided to adhere to Garrison's CoI model—which consists of teaching, cognitive, and social presence—as the theoretical framework.

We identified 21 studies that examined teaching, cognitive, and social presence in the context of generative AI applications in learning. Three studies were published in 2022 (Ebadi & Amini, 2022; Kim et al., 2022; Lee et al., 2022), followed by two studies in 2023 (Han et al., 2023; Hew et al., 2023). One study was published in 2024 (Xie et al., 2024). The number increased substantially to seven publications in 2025, accounting for almost half of all studies included in this systematic review (Alfoldi, 2025; Ba et al., 2025; Huang & Hew, 2025; Narang et al., 2025; Nasr et al., 2025; Xu et al., 2025; Yu et al., 2025). Eight studies were published in 2026, representing the largest annual share of publications included in the analysis (Darban, 2026; Dong et al., 2026; Gunbatar & Gokcearslan, 2026; Özbilen et al., 2026; Pan et al., 2026; Xia et al., 2026; Xu et al., 2026; Yang et al., 2026).

The 21 publications represented eight countries. The United States and China were the most productive countries investigating the presences in the context of generative AI, accounting for about two-thirds of the total publications. Eight studies were conducted in the United States (38.1%) and were published between 2022 and 2026 (Alfoldi, 2025; Darban, 2026; Han et al., 2023; Kim et al., 2022; Lee et al., 2022; Narang et al., 2025; Nasr et al., 2025; Pan et al., 2026). Six studies were conducted in China (28.6%) and were published between 2024 and 2026 (Ba et al., 2025; Dong et al., 2026; Xia et al., 2026; Xie et al., 2024; Xu et al., 2025; Yu et al., 2025).

The remaining seven publications were conducted in Hong Kong (Hew et al., 2023; Huang & Hew, 2025), Iran (Ebadi & Amini, 2022), Australia (Yang et al., 2026), the Netherlands (Xu et al., 2026), Turkey (Gunbatar & Gokcearslan, 2026), and across both Turkey and Pakistan (Özbilen et al., 2026).

The 21 publications were spread across 16 journals. The most productive journal, or the one most focused on the topic of presences within the CoI framework in the context of generative AI, was the *British Journal of Educational Technology* (N = 4, 19%). Two publications each also appeared in *The Internet and Higher Education* and the *International Journal of Educational Technology in Higher Education.* The remaining journals each contributed one publication, including *Computers & Education, Online Learning, Interactive Learning Environments, Computers in Human Behavior, Education Sciences, Educational Technology Research and Development, IEEE Transactions on Learning Technologies, Journal of Computer Assisted Learning, Journal of Education for Business, Journal of Public Policy & Marketing, Discover Computing, and the International Journal of Technology in Education.*

In terms of participant level, more than three-quarters of the studies were conducted with undergraduate students. Specifically, 16 of the 21 publications involved undergraduates (Alfoldi, 2025; Ba et al., 2025; Ebadi & Amini, 2022; Gunbatar & Gokcearslan, 2026; Han et al., 2023; Hew et al., 2023; Huang & Hew, 2025; Kim et al., 2022; Narang et al., 2025; Nasr et al., 2025; Özbilen et al., 2026; Xia et al., 2026; Xie et al., 2024; Xu et al., 2025; Yu et al., 2025; Dong et al., 2026). The other five studies examined different participant groups, including primary school students (Xu et al., 2026), high school students (Pan et al., 2026), graduate students (Darban, 2026), graduates in MOOC contexts (Lee et al., 2022), and adult learners (Yang et al., 2026).

The types of generative AI used in the studies varied. ChatGPT was the most frequently identified specific tool category (N = 7, 33.3%) (Alfoldi, 2025; Ba et al., 2025; Darban, 2026; Gunbatar & Gokcearslan, 2026; Nasr et al., 2025; Xie et al., 2024; Yang et al., 2026). Another widely used category was chatbots (N= 4, 19.0%), developed through various platforms, including CSIEC (Ebadi & Amini, 2022), knowledge-based chatbots used in MOOCs (Han et al., 2023), a learning-buddy chatbot developed using Google Dialogflow (Hew et al., 2023), and a chatbot designed to support self-regulated learning and social presence (Huang & Hew, 2025). Other generative AI tools reported include BERT (Bidirectional Encoder Representations from Transformers) (Lee et al., 2022), SPARK (a Chinese AI developed by iFLYTEK) (Yu et al., 2025), an LLM-driven conversational AI agent (Pan et al., 2026), identity-anthropomorphized virtual companion AI (Xia et al., 2026), generative AI-based learning analytics embedded in learning platforms (Dong et al., 2026), and various generative AI applications (Narang et al., 2025; Özbilen et al., 2026). More details about the 21 studies included in the analysis are presented in Table 2 and Appendix A.

Table 2

Descriptive Characteristics of Included Studies (N=21)

Category	Subcategory	N	%
Year	2022	3	14.3
	2023	2	9.5
	2024	1	4.8
	2025	7	33.3
	2026	8	38.1
Country	United States	8	38.1
	China	6	28.6
	Other countries	7	33.3
Publication Outlet	British Journal of Educational Technology	4	19.0
	International Journal of Educational Technology in Higher Education	2	9.5
	Internet and Higher Education	2	9.5
	Other journals (13 journals)	13	61.9
Participant Level	Undergraduate	16	76.2
	K-12	2	9.5
	Graduates (MOOCs)	2	9.5
	Adults	1	4.8
Type of Generative AI	ChatGPT	7	33.3
	Chatbot	4	19
	AI-generated instructors or pedagogical agents	2	9.5
	Other AI tools	8	38.1

Note. Other countries include Australia, Hong Kong, Iran, the Netherlands, and Turkey. Other AI tools include BERT, SPARK, LLM-driven conversational AI, embedded AI applications, identity-anthropomorphized AI, and other generative AI systems.

Relationships Between Generative AI and CoI Presences (RQ1)

The first research question investigated the effects of generative AI on teaching, cognitive, social, and learning presence. Depending on the research design, some studies treated presences as dependent variables, while others examined them as independent or mediating variables that were expected to influence learning achievement or performance. The findings indicate that generative AI affects CoI presences in various ways, such as enhancing critical thinking in terms of cognitive presence, increasing learner autonomy, and strengthening social presence through personalized interactions and feedback, and supporting teaching presence through AI-assisted instruction and facilitation, which in turn can lead to improved engagement and learning

performance.

First, regarding the effects of generative AI on social presence, generative AI consistently increases *social presence* by using human-like cues and interactions (e.g., voice, emojis, or affective expressions). Huang and Hew (2025) and Xu et al. (2025) reported that human-like chatbots or pedagogical agents increased social presence, fostered rapport with instructors, and enhanced learners' sense of belonging. Xu et al. (2026) further found that AI-generated instructors with high behavioral fidelity were more effective in enhancing social presence, satisfaction, and motivation, and in reducing cognitive load, than those with low behavioral fidelity. However, Xu et al. (2026) also showed that, despite the enhanced social presence, the pedagogical agent did not significantly influence learning achievement in mathematics, either as an independent variable or as a mediating variable between generative AI and learning achievement. Xia et al. (2026) demonstrated that identity-anthropomorphized AI further strengthened learners' perceived social presence by eliciting positive emotional responses, while Pan et al. (2026) showed that LLM-based conversational agents enhanced both social and teaching presence in collaborative learning environments.

Generative AI supported cognitive presence by providing scaffolding and facilitating development through its four phases: triggering events, exploration, integration, and resolution. Several studies demonstrate how GenAI contributes to movement across these phases. In addition, unlike studies that focused primarily on social presence, publications examining cognitive presence tended to incorporate individual difference variables, such as creativity (Yu et al., 2025) or cognitive presence levels (Lee et al., 2022), to explain variations in learners' progression across cognitive presence phases. For example, Yu et al. (2025) compared cognitive presence between high-creativity and low-creativity pre-service teachers. While no significant differences were found in the exploration and interaction phases, participants in the high-creativity group were more active in the integration and resolution phases, suggesting deeper knowledge construction. Similarly, Lee et al. (2022) investigated how generative AI could classify students according to their cognitive presence levels in order to provide targeted scaffolding and promote more substantive discussion, ultimately improving learning outcomes. Yang et al. (2026) analyzed students' writing behaviors through the cognitive presence framework, examining progression from triggering events to resolution. Based on engagement in each phase, they identified distinct patterns of AI use, including AI-critical, AI-dependent, AI-independent, and AI-balanced writers. Recent studies further extended this line of inquiry. Dong et al. (2026) found that generative AI-supported learning analytics significantly enhanced cognitive presence while simultaneously strengthening teaching and social presence. Darban (2026) also reported that ChatGPT-assisted learning promoted higher levels of cognitive presence and self-directed inquiry among graduate students. Similarly, Özbilen et al. (2026) demonstrated positive associations between generative AI use and learners' cognitive engagement during university learning activities.

No study exclusively examined teaching presence; however, several studies explicitly examined teaching presence alongside cognitive and social presence. Narang et al. (2025) found that when generative AI functioned as a tutor—by providing personalized guidance and feedback—teaching presence was strengthened. Similarly, Ba et al. (2025) examined the effects of different types of AI-generated feedback (i.e., idea-oriented feedback and task-oriented feedback) on teaching, social, and cognitive presence. Their findings indicated significant differences in cognitive presence across feedback types, whereas no significant differences were

observed in teaching presence. Narang et al. (2025) compared the effects of generative AI on learning in a marketing course using the tutor–tutee–tool framework. Their findings indicated that, among the three dimensions of presences, cognitive presence was perceived as the most effective (85.71%), compared to teaching presence (42.86%) and social presence (35.71%). Pan et al. (2026) found that LLM-supported collaborative learning significantly enhanced teaching presence by supporting instructional facilitation and guidance, whereas Dong et al. (2026) reported that teaching presence served as the strongest predictor among the three presences when generative AI-supported learning analytics were introduced. Recent studies suggest that although generative AI positively influences all three CoI presences, its strongest and most consistent effects have been reported for cognitive presence, followed by social presence, whereas teaching presence appears to depend more heavily on instructional design and the specific role assigned to AI within the learning environment.

Although most empirical studies report positive effects of using generative AI, neutral or even negative effects also emerge. These were not our primary research questions; however, we identified several negative effects of using generative AI in the empirical studies included in our review. For example, Dong (2026) delineated that the impact of generative AI-based learning analytics is contingent upon the level of CoI support and students' digital competencies. Xia et al. (2026) found that, despite the positive influence of an identity-anthropomorphised virtual companion on learning outcomes, learners in the identity-anthropomorphism condition spent significantly less time on the learning videos than those in the non-anthropomorphism condition. Xia explained that a strongly human-like AI companion can distract learners' attention from the instructional video. Ba et al. (2025) reported that learners who received idea-oriented feedback prompts demonstrated higher cognitive engagement than those who received task-oriented feedback prompts, but also exhibited higher negative emotional engagement.

Patterns and Imbalances in CoI Presences Research (RQ2)

The second research question aimed to identify patterns and imbalances in the existing research on generative AI and presences within the CoI framework. As shown in Table 3, social presence emerged as the most frequently examined dimension, appearing in 61.9% of the included studies (13 out of 21), either as a standalone focus or in combination with other presences. In contrast, cognitive presence was examined in 47.6% of the studies (10 out of 21), while teaching presence appeared in only 23.8% of the studies (5 out of 21). All studies addressing teaching presence examined it together with one or more of the other presences.

A closer examination of the distribution reveals a structural imbalance in research attention. Eight studies focused solely on social presence, whereas six examined only cognitive presence. Notably, no studies investigated teaching presence as a standalone construct in the context of generative AI. Instead, teaching presence appeared only as part of broader analyses that examined multiple CoI presences simultaneously (e.g., Ba et al., 2025; Han et al., 2023; Narang et al., 2025; Dong et al., 2026; Pan et al., 2026), suggesting that it has not yet been independently theorized or empirically isolated in this line of research.

This distribution indicates that current scholarship disproportionately emphasizes the relational and affective dimensions of generative AI use, while comparatively limited attention has been given to instructional design and facilitation processes associated with teaching presence. Such an imbalance suggests that the integration of generative AI into the CoI

framework remains unevenly conceptualized across its three core dimensions.

In terms of generative AI applications, chatbot-based studies primarily examined social presence (e.g., Ebadi & Amini, 2022; Hew et al., 2023; Hwang & Hew, 2025), except Han et al. (2023), who reported all three presences. Studies using ChatGPT addressed a broader range of CoI presences, including cognitive presence (e.g., Nasr et al., 2025; Darban, 2026; Gunbatar & Gokcearslan, 2026; Yang et al., 2026), social presence (Xie et al., 2024), and all three presences (Ba et al., 2025). Several studies also employed AI-generated instructors or pedagogical agents (e.g., Pan et al., 2026; Xia et al., 2026; Xu et al., 2025; Xu et al., 2026), reflecting an expansion of generative AI applications beyond conventional chatbots and general-purpose LLMs. Table 4 lists the type of AI and the three presences examined in each study.

Table 3

Types of AI and Three Presences

Presence dimension	Studies examining this dimension (n)			% of total studies (N= 21)
	Single	Combined	Total	
Cognitive presence	6	4	10	47.6
Social presence	8	5	13	61.9
Teaching presence	-	5	5	23.8

Table 4

Generative AI Applications Examined Across the Three CoI Presences

Generative AI Application	Publications	Teaching	Cognitive	Social
ChatGPT	Alfoldi et al. (2025)		X	
	Ba et al. (2025)	X	X	X
	Darban (2026)		X	
	Gunbatar & Gokcearslan (2026)		X	
	Nasr et al. (2025)		X	
	Xie et al. (2024)			X
	Yang et al. (2026)		X	
Chatbots	Ebadi & Amini (2022)			X
	Han et al. (2023)	X	X	X
	Hew et al. (2023)			X
	Huang & Hew (2025)			X

Generative AI Application	Publications	Teaching	Cognitive	Social
AI-generated instructors / pedagogical agents	Pan et al. (2026)	X		X
	Xia et al. (2026)			X
	Xu et al. (2025)			X
	Xu et al. (2026)			X
Learning analytics	Dong et al. (2026)	X	X	X
Various generative AI applications	Narang et al. (2025)	X	X	X
	Özbilen et al. (2026)		X	
	Yu et al. (2025) (SPARK)		X	

Note. Studies focusing primarily on AI-supported measurement or classification (e.g., Lee et al., 2022) or broader AI-supported learning environments (e.g., Kim et al., 2022) were not categorized by AI application type.

Instructional Approaches to Strengthen Presences through GenAI (RQ3)

To address the third research question, we examined the instructional configurations in which generative AI was embedded and the specific pedagogical methods adopted in each study. Four dominant instructional groupings emerged: (1) online discussion–based environments, (2) AI-generated instructor and video-based contexts, (3) self-regulated learning support environments, and (4) task-specific collaborative activities.

First, several studies embedded generative AI within structured *online discussion* environments. These contexts included inquiry-based discussions (Ba et al., 2025), structured argumentation using the six thinking hats technique (Yu et al., 2025), and large-scale asynchronous discussion forums analyzed through machine learning models (Lee et al., 2022). In these dialogic settings, generative AI provided scaffolded feedback, guided argumentation, and supported idea elaboration. Ba et al. (2025), for example, demonstrated that idea-oriented AI feedback strengthened cognitive presence in inquiry-based discussions, particularly in the exploration and integration phases. Similarly, Yu et al. (2025) showed that combining generative AI with structured debate techniques facilitated deeper patterns of knowledge construction. Across these discussion-based environments, generative AI was most consistently associated with the enhancement of cognitive presence.

A second instructional configuration involved *video-based or MOOC-centered instructional delivery formats* in which generative AI functioned as an instructor-like agent. Xu et al. (2025) manipulated behavioral fidelity and dynamic visual cues in AI-generated instructors embedded within online instructional videos. Kim et al. (2022) examined differences between human-like and machine-like AI voices in voice-based lectures, while Xu et al. (2026) investigated animated pedagogical agents in elementary mathematics instruction. Narang et al. (2025) further discussed AI-generated lecture videos and simulation-based case analyses in project-based marketing courses. Pan et al. (2026) investigated an AI conversational agent (MathPal) designed to support mathematical dialogue and instructional interaction, extending AI tutor roles beyond video-based instructional delivery. In these lecture-oriented environments, generative AI was positioned as a tutor delivering content, and enhanced social presence was particularly evident when AI designs incorporated human-like instructional characteristics, such as realistic voice, dynamic visual behaviors, conversational interaction, and embodied tutor functions.

A third group of studies situated generative AI within *self-regulated learning (SRL)* interventions. These included SMART goal-setting chatbot activities (Hew et al., 2023), structured help-seeking and self-evaluation chatbot frameworks (Huang & Hew, 2025), informal chatbot interactions supporting learner autonomy in elective courses (Xie et al., 2024), and chatbot-mediated knowledge-based learning experiences in MOOCs (Han et al., 2023). In these contexts, generative AI functioned as a learning companion that supported monitoring, reflection, and goal alignment. Social presence frequently emerged as a mediating mechanism, particularly in relation to learning autonomy and perceived engagement.

Finally, several studies embedded generative AI within task-specific activities requiring human–AI collaboration. These contexts included AI-assisted creative and argumentative writing (Yang et al., 2026), lesson planning tasks supported by chatbot interaction (Nasr et al., 2025), chatbot-based English conversation practice within blended learning environments (Ebadi & Amini, 2022) and AI-supported collaborative learning enhanced through learning analytics (Dong et al., 2026) These instructional approaches emphasized applied task performance and iterative interaction with AI systems. In such settings, generative AI functioned primarily as a collaborative teammate or cognitive tool, contributing most strongly to cognitive presence while also shaping aspects of social engagement during dialogic exchanges.

The reviewed studies do not point to a uniform pattern across the three CoI presences. Rather, the different instructional configurations were associated with different presence-related findings. Discussion- and task-based instructional methods were most consistently associated with enhanced cognitive presence, whereas lecture-based designs emphasizing embodied AI characteristics were more often focused on social presence. Self-regulated learning frameworks highlighted the mediating role of social presence in fostering learner autonomy. Teaching presence appeared most evident when generative AI assumed explicit instructional roles, such as AI-generated instructors, conversational tutors, or AI-supported learning analytics integrated into structured instructional designs. Nevertheless, teaching presence remained the least frequently investigated of the three CoI Presence. Table 5 summarizes the instructional configurations and pedagogical roles of generative AI associated with the three CoI presences.

Table 5

Instructional Configurations of Generative AI Across the Three CoI Presences

Instructional group	Studies	Learning context and AI role
1. Online Discussion-based environments	Ba et al. (2025); Yu et al. (2025); Lee et al. (2022)	AI used to scaffold inquiry-based discussions, structured argumentation, and cognitive presence in online forums (AI as teammate or cognitive tool)
2. AI-Generated Instructor and tutor-supported instructional environments	Xu et al. (2025); Kim et al. (2022); Xu et al. (2026); Narang et al. (2025); Pan et al. (2026); Xia et al. (2026); Narang et al. (2025)	AI functioning as instructor/tutor in video-based or MOOC environments, emphasizing behavioral fidelity, voice realism, and instructional delivery
3. Self-Regulated Learning support environments	Xie et al. (2024); Hew et al. (2023); Huang & Hew (2025); Han et al. (2023)	Chatbot-supported goal setting, help-seeking, and self-monitoring activities in online learning environments (AI as tutor or learning buddy)

Instructional group	Studies	Learning context and AI role
4. Task-Specific Collaborative activities	Yang et al. (2026); Nasr et al. (2025); Ebadi & Amini (2022); Dong et al. (2026); Darban (2026); Gunbatar & Gokcearslan (2026); Alfoldi et al (2025)	AI-assisted writing, lesson planning, and conversational practice tasks emphasizing human–AI collaboration (AI as teammate or cognitive tool)

Discussion and Conclusions

The purpose of this study was to synthesize empirical evidence on how generative AI is associated with presence within the CoI framework. The reviewed studies did not reveal a single, uniform pattern across the three CoI presences. Instead, presence-related outcomes varied according to instructional configurations, AI roles, and measurement approaches, suggesting that the educational value of generative AI depends on how it is pedagogically designed and implemented. These findings have several implications for understanding the role of generative AI within the CoI framework.

With respect to *cognitive presence*, generative AI was most consistently associated with strengthened exploration and integration processes, particularly in discussion-based and task-oriented environments. AI-supported inquiry-based discussions (Ba et al., 2025), structured debate formats (Yu et al., 2025), and AI-assisted writing and lesson-planning tasks (Yang et al., 2026; Nasr et al., 2025), mathematical problem solving (Darban, 2026; Gunbatar & Gokcearslan, 2026), and AI-supported collaborative learning (Dong, et al., 2026) consistently facilitated idea generation, elaboration, and iterative refinement. However, across multiple studies, the resolution phase of cognitive presence remained comparatively underdeveloped. This recurring imbalance suggests that although generative AI may accelerate the exploration and integration phases of knowledge construction, sustained inquiry and resolution remain strongly contingent upon deliberate instructional design and facilitation.

Interpreted through this extended CoI framework, the findings of this review suggest that generative AI may support certain phases of cognitive presence, particularly exploration and integration, but does not consistently facilitate the regulatory processes required for sustained inquiry and resolution. While generative AI can scaffold idea generation and interaction, the monitoring and management of inquiry—central to learning presence—remain underdeveloped. This may help explain why the resolution phase of cognitive presence was consistently limited across the reviewed studies.

However, this imbalance can be more deeply understood when viewed through recent extensions of the CoI framework. In particular, the introduction of learning presence highlights the role of learners' self-, co-, and socially shared regulation in shaping collaborative inquiry (Shea et al., 2022). From this perspective, meaningful learning is not only a function of cognitive, social, and teaching presence, but also of how learners monitor and manage the inquiry process. Stenbom and Garrison (2026) asserted that learning in AI-supported environments requires reflective responsibility and control through metacognitive monitoring and collaborative inquiry processes. This perspective aligns with Garrison's (2022) concept of shared metacognition, which emphasizes that effective inquiry requires both individual and collective regulation through monitoring and managing learning activities. Shared metacognition

operates at the intersection of cognitive and teaching presence, suggesting that the progression of inquiry—particularly toward resolution—depends on the extent to which learners actively regulate their thinking and interaction. Several of the reviewed studies also illustrate this regulatory role in practice. Chatbot-supported goal setting and self-evaluation activities (Hew et al., 2023; Huang & Hew, 2025) and AI-supported self- and co-regulation of learning (Özbilen et al., 2026) show how monitoring and managing inquiry can be deliberately built into AI-supported learning designs.

S*ocial presence* was most strongly activated in contexts where generative AI incorporated anthropomorphic and socially expressive interactional cues. Studies involving AI-generated instructors, conversational tutors, and anthropomorphic pedagogical agents (Kim et al., 2022; Xu et al., 2026; Pan et al., 2026; Xia et al., 2026) demonstrated that behavioral fidelity, conversational tone, and perceived realism significantly influenced learners' sense of connection and engagement. Furthermore, in self-regulated learning environments (Xie et al., 2024), social presence frequently functioned as a mediating mechanism linking chatbot interaction to learner autonomy. Across the reviewed studies, social presence appeared particularly sensitive to design features that signal relational authenticity, positioning generative AI as a socially responsive agent.

Prior work has described presence as a perceived sense of connection between learners and instructors (Seo et al., 2021) or as the degree to which learners perceive the existence of others in mediated interaction (Kang & Im, 2013). The reviewed studies suggest that learners can experience a similar sense of connection in interactions with AI agents, particularly when the agents display human-like behavioral and conversational cues (Kim et al., 2022; Xia et al., 2026; Xu et al., 2026). At the same time, this connection is not equivalent to interpersonal connection between humans, and none of the reviewed studies examined its longer-term effects on learner relationships or engagement. How learners' perceived connection with AI agents relates to their connection with human instructors and peers remains an open question for future research.

A more cautious interpretation of these patterns is necessary because the reviewed studies did not define or operationalize the CoI presences in fully comparable ways. Some studies treated presence as a perceived outcome measured through self-report scales, whereas others inferred it from discourse patterns, interaction features, or design characteristics. In addition, not all studies examined the same dimensions of the CoI framework, and teaching presence was rarely isolated as a standalone construct. This construct instability limits direct cross-study comparison and suggests that the present review identifies recurrent tendencies in the literature rather than stable empirical regularities about how generative AI influences learning.

An additional pattern emerging from the reviewed studies is the diversification of AI roles in educational contexts. Early studies primarily examined chatbots designed to support communication and self-regulated learning, whereas later studies investigated AI-generated instructors, pedagogical agents, conversational tutors, collaborative cognitive tools, and AI-supported learning analytics. Recent studies thus embed generative AI within broader instructional designs.

Teaching presence, however, emerged as comparatively underexplored as an independent construct. Although teaching presence was identified in several recent studies (Ba et al., 2025; Han et al., 2023; Narang et al., 2025; Pan et al., 2026; Dong et al., 2026), it was rarely examined as an independent construct. Instead, it was typically examined alongside cognitive and social

presence. Given that teaching presence traditionally serves as the integrative force orchestrating cognitive and social processes within the CoI framework, its relative absence in the literature reveals an important research gap. As AI systems increasingly participate in instructional design and feedback processes, clearer conceptualization of shared instructional agency between human instructors and AI systems becomes necessary.

Collectively, these findings suggest that the three CoI presences do not operate in AI-mediated environments as originally assumed. The original CoI model assumes an interdependent and mutually reinforcing relationship among teaching, cognitive, and social presence. However, the current body of generative AI research reveals a more asymmetric activation pattern. Rather than strengthening all presences simultaneously, generative AI selectively activates particular dimensions depending on instructional format, AI role, and design features. Discussion- and task-based pedagogies tend to amplify cognitive presence, embodied AI instructor designs primarily enhance social presence, and tutor-like AI roles may contribute to teaching presence when explicitly structured within pedagogical design.

Some scholars have argued that generative AI can compensate for structural limitations of the CoI framework by offering dynamic, interactive responses and personalized feedback (Anderson et al., 2025). From this perspective, AI is positioned as a technological facilitator capable of strengthening community building and knowledge construction processes. However, the findings of this review suggest a more conditional interpretation. Generative AI appears to amplify particular dimensions of presence depending on instructional approaches and pedagogical configuration. In certain designs, AI-supported interaction enhanced cognitive presence; in others, anthropomorphic and socially expressive cues strengthened social presence. Yet teaching presence often remained underexamined or unevenly distributed. These patterns indicate that generative AI should not be conceptualized as a structural remedy for CoI limitations, but rather as a design-sensitive mediator whose impact emerges through instructional configuration. These patterns also underscore the role of regulation in AI-supported inquiry: across the reviewed studies, progress toward integration and resolution depended less on the availability of generative AI itself than on whether the instructional design supported learners in monitoring and managing their inquiry (Garrison, 2022).

These findings align with the four instructional configurations identified in this review, illustrating that the educational value of generative AI depends on how AI is pedagogically embedded. First, educators and instructional designers should recognize that generative AI does not inherently guarantee enhancement of presence within the CoI framework. Its impact depends on how it is pedagogically positioned. Designing AI-mediated learning environments requires intentional alignment between AI roles and desired dimensions of presence. Second, instructional strategies should extend AI-supported exploration toward deeper integration and resolution processes to avoid superficial engagement. Third, greater attention should be directed toward designing AI systems that meaningfully support teaching presence, particularly in relation to instructional orchestration and facilitation.

Limitations of this study

Several limitations of this review should be acknowledged. First, the number of empirical studies explicitly examining generative AI through the CoI framework remains relatively limited, and the distribution across presence dimensions was uneven. Additionally, only English-language

studies published after November 2022 were included, which may have excluded relevant emerging research. The rapid evolution of generative AI technologies also means that findings may reflect early-stage implementations. Future research should therefore examine longitudinal implementations of AI-mediated learning designs and further investigate how teaching presence is conceptualized and enacted in AI-supported instructional environments.

Second, the reviewed studies varied in their definitions and operationalizations of CoI presences, limiting comparability across studies. While some studies treated presence as a perceived outcome measured through self-report scales, others inferred it from discourse patterns, interaction features, or design characteristics. This variability limits direct cross-study comparison and suggests that the findings of this review should be interpreted as patterns and tendencies.

Third, although learning presence–emphasizing self-regulation, co-regulation, and shared regulation–has gained attention as a potential fourth element of the CoI framework, it was not included as a primary analytical category in this review. However, it was used as a theoretical lens to interpret the findings, particularly in relation to regulation of inquiry processes. At the same time, its conceptual status remains debated, as some scholars propose it as a distinct fourth element, while others view it as overlapping with or embedded within existing presences. Future research should further clarify its conceptual status and examine how learning presence interacts with cognitive, social, and teaching presence in AI-mediated learning environments.

Finally, during the literature review, it was observed that the term “presence” is used in varying ways across studies, including teacher presence, AI presence, and learner presence. To maintain conceptual consistency, this study adhered to the CoI framework, focusing specifically on teaching, cognitive, and social presence. Future work may benefit from more precise conceptual distinctions across different forms of presence in AI-mediated learning environments.

In conclusion, this review demonstrates that the educational value of generative AI is not determined by the technology itself, but by how it is pedagogically configured within specific instructional contexts. Generative AI selectively supports cognitive, social, and teaching presence depending on its instructional role, interaction design, and learning environment. These findings contribute to a more nuanced understanding of AI-mediated learning by positioning generative AI as a design-sensitive pedagogical mediator.

Declarations

Acknowledgement
Not applicable

Funding Statement
This study was supported by the research grant of Kangwon National University in 2025.
This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A01090926).

AI Use Statement
Generative artificial intelligence (AI) was used in the preparation of this manuscript for language

editing and to improve readability. The authors reviewed and edited the AI-assisted text and take full responsibility for the content of the manuscript.

Conflicts of Interest Statement

The authors declare no conflicts of interest related to the research, authorship, or publication of this article.

References

*Alfoldi, E. A. (2025). Yellowdig discussion boards as communities of inquiry: The impact of instructor participation on online student discourse. *Journal of Education for Business, 100*(1), 45–56. https://doi.org/10.1080/08832323.2024.2438659

Anderson, J. E., Nguyen, A. E., & Moreira, G. (2025). Generative AI-driven personalization of the Community of Inquiry model: enhancing individualized learning experiences in digital classrooms. *The International Journal of Information and Learning Technology, 42*(3), 296-310. https://doi.org/10.1108/IJILT-10-2024-0240

Ba, S., Hu, X., Stein, D., & Liu, Q. (2023). Assessing cognitive presence in online inquiry-based discussion through text classification and epistemic network analysis. *British Journal of Educational Technology, 54*(1), 247–266. https://doi.org/10.1111/bjet.13285

*Ba, S., Zhan, Y., Haung, L., & Lu, G. (2025). Investigating the impact of ChatGPT-assisted feedback on the dynamics and outcomes of online inquiry-based discussion. British *Journal of Educational Technology, 56*, 1710-1734.

Castellanos-Reyes, D., Olesova, L., & Sadaf, A. (2025). Transforming online learning research: Leveraging GPT large language models for automated content analysis of cognitive presence. The Internet and Higher Education, 65, 101001. https://doi.org/10.1016/j.iheduc.2025.101001

Chang, C. C., Yen, W. H., Liu, S. C., & Kuo, C. T. (2026). Learning presence in the revised community of inquiry framework: A systematic review of empirical research (2010–2024). *Internet and Higher Education*, *68*, 101058. https://doi.org/10.1016/j.iheduc.2025.101058

*Darban, M. (2026). Beyond tools to teammates: A multilevel study of explainable ChatGPT in virtual team learning. *The Internet and Higher Education, 71*, 101102. https://doi.org/10.1016/j.iheduc.2026.101102

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227,* Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

*Dong, H., Chen, X., & Shen, J. (2026). Unpacking the impact mechanism of generative AI-based learning analytics technology on student learning engagement: The mediating role of community of inquiry and the moderating role of digital literacy. *International Journal of Educational Technology in Higher Education, 23*, 14. https://doi.org/10.1186/s41239-026-00591-z

Doo, M. Y., & Park, Y. (2026). A Meta-Analysis of ChatGPT's Influence on Learning Achievement. *The International Review of Research in Open and Distributed Learning*, *27*(1), 265–289. https://doi.org/10.19173/irrodl.v27i1.8775

*Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on

Iranian EFL learners' motivation using artificial intelligence technology: A case of CSIEC chatbot. *Interactive Learning Environments, 32*(2), 655-673. https://doi.org/10.1080/10494820.2022.2096638

ElSayad, G. (2023). Can learning presence be the fourth community of inquiry presence? Examining the extended community of inquiry framework in blended learning using confirmatory factor analysis. *Education and Information Technologies*, *28*(6), 7291–7316. https://doi.org/10.1007/s10639-022-11480-z

Freeman, J. (2025). *Student generative AI survey 2025*. Higher Education Policy Institute. https://www.hepi.ac.uk/reports/?_keyword=Student%20Generative%20AI%20Survey%202025

Galikyan, I., & Admiraal, W. (2019). Students' engagement in asynchronous online discussion: The relationship between cognitive presence, learner prominence, and academic performance. *The Internet and Higher Education, 43*, Article 100692. https://doi.org/10.1016/j.iheduc.2019.100692

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2-3), 87-105. https://doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R. (2007). Online community of inquiry review: Social, cognitive, and teaching presence issues. *Journal of Asynchronous Learning Networks, 11*(1), 61–72.

Garrison, D. R. (2022). Shared metacognition in a Community of Inquiry. *Online Learning, 26*(1), 6-18. https://doi.org/10.24059/olj.v26i1.3023

Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education, 9*, Article e45312. https://doi.org/10.2196/45312

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education, 11*(3), 8–26. https://doi.org/10.1080/08923649709526970

*Gunbatar, M. S. & Gokcearslan, S. (2026). University students' interactions with ChatGPT: An investigation in terms of CoI, motivation, and learning strategies. *International Journal of Technology in Education (IJTE), 9*(2), 343-360. https://doi.org/10.46328/ijte.5419

*Han, S., Hamilton, X., Cai, Y., Shao, P., & Liu, M. (2023). Knowledge-based chatbots: a scale measuring students' learning experiences in massive open online courses. *Educational Technology & Research Development, 71,* 2431–2456. https://doi.org/10.1007/s11423-023-10280-7

*Hew, K.F., Huang, W., Du, J., & Jia, C (2023). Using chatbots to support student goal

settingand social presence in fully online activities: learner engagement and perceptions. *Journal of Computing Higher Education, 35*, 40–68. https://doi.org/10.1007/s12528-022-09338-x

*Huang, W., & Hew, K. F. (2025). Facilitating online self-regulated learning and social presence using chatbots: Evidence-based design principles. *IEEE Transactions on Learning Technologies, 18,* 56-71. https:// doi.org/10.1109/TLT.2024.3523199

Kang, M. & Im, T. (2013). Factors of learner-instructor interaction which predict perceived learning outcomes in online learning environment. *Journal of Computer Assisted Learning*, 29(3), 292–301. https://doi.org/10.1111/jcal.12005

*Kim, J., Merrill, K., Jr., Xu, K., & Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Computers in Human Behavior, 136,* 1–9. https://doi.org/10.1016/j.chb.2022.107383

Kim, S., So, H.-J., & Park, K. (2025). Supporting learner agency in collaborative writing with generative AI. *British Journal of Educational Technology, 00*, 1–25. https://doi.org/10.1111/bjet.70015

Law, N., Wang, N., Ma, M., Liu, Z., Lei, L., Feng, S., Hu, X., & Tsao, J. (2025). The role of generative AI in collaborative problem-solving of authentic challenges. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.70010

*Lee, J., Soleimani, F., Irish, I., Hosmer, IV, J., Yilmaz Soylu, M., Finkelberg, R., & Chatterjee, S. (2022). Predicting cognitive presence in at-scale online learning: MOOC and for-credit online course environments, *Online Learning, 26*(1), 58-79. https://doi.org/ 10.24059/olj.v26i1.3060

Leong, P. (2011). Role of social presence and cognitive absorption in online learning environments. *Distance Education*, *32*(1), 5–28. https://doi.org/10.1080/01587919.2011.565495

Martin, F., Wu, T., Wan, L., & Xie, K. (2022). A Meta-analysis on the community of inquiry presences and learning outcomes in online and blended learning environments. *Online Learning, 26*(1). https://doi.org/10.24059/olj.v26i1.2604

*Narang, U., Sachdev, V., & Liu, R. (2025). When AI wears M-many hats: The role of generative artificial intelligence in marketing education. *Journal of Public Policy & Marketing*, *44*(3), 473-489. https://doi.org/10.1177/07439156251328237

Nasr, N. R., Tu, C.-H., Werner, J., Bauer, T., Yen, C.-J., & Sujo-Montes, L. (2025). Exploring the impact of generative AI ChatGPT on critical thinking in higher education: Passive AI-directed use or human-AI supported collaboration? Education Science, 15*(9), 1198. https://doi.org/10.3390/educsci15091198

Ng, D. T. K. (2022). What is the metaverse? Definitions, technologies and the community of inquiry. *Australasian Journal of Educational Technology*, *38*(4), 190–205.

https://doi.org/10.14742/ajet.7945

*Özbilen, F. M., Asghar, M. Z., & Bayrakci, M. (2026). A multimethod exploration of human–AI-supported shared metacognition for inclusive educational leadership during crisis in Türkiye and Pakistan. *Discover Computing, 29*, 354. https://doi.org/10.1007/s10791-026-10235-5

*Pan, Z., Jiang, Z., Ba, S., & Xie, K. (2026). A community of inquiry perspective on human–AI co-facilitation within micro-blended learning. *Computers & Education, 253*, 105707. https://doi.org/10.1016/j.compedu.2026.105707

Richardson, J. C., Caskurlu, S., Castellanos-Reyes, D., Duan, S., Duha, M. S. U., Fiock, H.,

& Long, Y. (2022). Instructors' conceptualization and implementation of scaffolding in online higher education courses. *Journal of Computing in Higher Education, 34*(1), 242–279. https://doi.org/10.1007/s12528-021-09300-3

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education, 14*(2), 50–71.

Samala, A. D., Zhai, X., Aoki, K., Bojic, L., & Zikic, S. (2024). An in-depth review of ChatGPT's pros and cons for learning and teaching in education. *International Journal of Interactive Mobile Technologies (IJIM), 18*(02), 96–117. https://doi.org/10.3991/IJIM.V18I02.46509

Seo, K., Tang, J., Roll, I., et al. (2021) The Impact of artificial intelligence on learner—instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18,* Article No. 54. https://doi.org/10.1186/s41239-021-00292-9

Shahzad, M. F., Xu, S., An, X., & Asif, M. (2025). Are generative AI technologies transforming education for the 21st century? Research trends, challenges, and benefits. *Sage Open*, *15*(3). https://doi.org/10.1177/21582440251368594 (Original work published 2025)

Shahzad, M. F., Xu, S., & Zahid, H. (2025). Exploring the impact of generative AI-based technologies on learning performance through self-efficacy, fairness and ethics, creativity, and trust in higher education. *Education and Information Technologies*, 30(3), 3691-3716. https://doi.org/10.1007/s10639-024-12949-9

Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education, 55*(4), 1721–1731. https://doi.org/10.1016/j.compedu.2010.07.017

Shea, P., Richardson, J., & Swan, K. (2022). Building bridges to advance the Community of Inquiry framework for online learning. *Educational Psychologist*, *57*(3), 148–161. https://doi.org/10.1080/00461520.2022.2089989

Stenbom, S., & Garrison, D. R. (2026). Artificial Intelligence and Communities of Inquiry: Reimagining Educational Experiences. *The International Review of Research in Open and Distributed Learning*, *27*(2), 114–131. https://doi.org/10.19173/irrodl.v27i2.9314

Sun, L., & Zhou, L. (2024). Does generative artificial intelligence improve the academic achievement of college students? A meta-analysis. *Journal of Educational Computing Research*, *62*(7), 1676-1713. https://doi.org/10.1177/07356331241277937

Wertz, R. E. H. (2022). Learning presence within the Community of Inquiry framework: An alternative measurement survey for a four-factor model. *The Internet and Higher Education, 52,* 1–15. https://doi.org/10.1016/j.iheduc.2021.100832

Wilson, M. & Ritzhaupt, A. D. (2023). Where should educational technologists publish their research? Updated results from a survey of researchers and professionals. In R. E. West & H. Leary (Eds.), *Becoming an LIDT Professional*. EdTech Books. https://edtechbooks.org/becoming_an_lidt_pro/publishing_research

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10-33. https://doi.org/10.1111/bjet.13334

*Xia, T., Tan, M., & Guo, J. (2026). Enhancing online learning outcomes through virtual companion AI: The role of identity anthropomorphism. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.70072

*Xie, Z., Wu, X., & Xie, Y. (2024). Can interaction with generative artificial intelligence enhance learning autonomy? A longitudinal study from comparative perspectives of virtual companionship and knowledge acquisition preferences. *Journal of Computer Assisted Learning, 40*(5), 2369–2384. https://doi.org/10.1111/jcal.13032

*Xu, T., Chen, Q., Zhang, Z., Dong, B., Zhang, H., Bai, J., & Zhou, Y. (2025). Maximizing effectiveness of AI-generated instructors through human-like behavior and dynamic visual cues in instructional videos: Evidence from an eye-tracking study. *The Internet and Higher Education, 67*, 101034. https://doi.org/10.1016/j.iheduc.2025.101034

*Xu, K. M., Lin, L., Gorter, M., Schneider, S., Weidlich, J., Davis, R. O., Kreijns, K., & de Groot, R. (2026). Social presence: A key factor in embedding a pedagogical agent into online learning in primary education. *British Journal of Educational Technology*, *57*, 227–242. https://doi.org/10.1111/bjet.70006

Xu, W., & Ouyang, F. (2021). A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies, 27*(3), 4195-4223. https://doi.org/10.1007/s10639-021-10774-y

*Yang, K., Cheng, Y., Zhao, L., Raković, M., Swiecki, Z., Gašević, D., & Chen, G. (2026). Ink and algorithm: Exploring temporal dynamics in generative AI-assisted writing. *British Journal of Educational Technology*, *57*, 508–534. https://doi.org/10.1111/bjet.70031

*Yu, M., Liu, Z., Long, T., Li, D., Deng, L., Kong, X., & Sun, J. (2025). Exploring cognitive presence patterns in GenAI-integrated six-hat thinking technique scaffolded discussion: an epistemic network analysis. *International Journal of Educational Technology in Higher Education, 22,* 48. https://doi.org/10.1186/s41239-025-00545-x

*References marked with an asterisk indicate studies included in the systematic review.

Appendix A. Overview of Selected Papers

Authors (Years)	Participants	Learning Discipline	Instructional Approach	Role of AI	Presence
Ebadi & Amini (2022)	256 Undergraduate Students	English (EFL Speaking and Grammar)	English conversation practice and blended learning using CSIEC chatbot	Chatbot instructor / Learning companion	Social Presence (Predictor of learner motivation)
Kim et al. (2022)	223 Undergraduate Students	Biology (Basics of Cells)	Online voice-based lecture	AI instructor (Varying voice features and expertise)	Social Presence (Mediator of perceived credibility)
Lee et al. (2022)	MOOC & For-Credit Students	Computer Programming, AI	Asynchronous online discussion forums	Analysis tool (Using ML to predict presence phases)	Cognitive Presence
Han et al. (2023)	237 MOOC Learners	Mental Health for Journalists, etc.	Learning with knowledge-based chatbots in MOOCs	Knowledge-based chatbot	Social, Teaching, and Cognitive Presence
Hew et al. (2023)	29 Postgraduate & 38 Undergraduate Students	Adult Education, English (EFL Listening)	Fully online activities (Goal setting and listening practice)	Goal-setting support chatbot and Learning Buddy	Social Presence
Xie et al. (2024)	1155 Undergraduate Students	Informal Learning (Elective Computer Courses)	Self-regulated learning through chatbot interaction	Virtual companion vs. Knowledge acquisition tool	Social Presence (Mediator for Learning Autonomy)
Alfoldi (2025)	1,146 discussion posts (Business students)	International busiess	Yellowdig online discussion board	Instructor participation (teaching presence)	Teaching, social, cognitive, and learner presence
Ba et al. (2025)	105 Preservice Teachers	Biological Sciences	Online inquiry-based discussion (IBD)	Feedback provider (Idea-oriented vs. Task-oriented)	Cognitive Presence (Idea-oriented), Social Presence (Task-oriented), Teaching Presence
Huang & Hew (2025)	39 Undergraduate & 25 Postgraduate Students	Language Education, Educational Research	Online self-regulated learning (SRL) activities	Learning Buddy (Supporting goal-setting, help-seeking, self-evaluation)	Social Presence (Combined with instructor presence)
Narang et al. (2025)	Marketing Educators	Marketing	Project-based learning, simulation, instructional video	Tutor, Teammate, Tool	Teaching Presence (Tutor), Social Presence (Teammate), Cognitive Presence (Tool)

Authors (Years)	Participants	Learning Discipline	Instructional Approach	Role of AI	Presence
Nasr et al. (2025)	40 University Students	Education (Elementary Education)	Lesson planning task	Virtual assistant/Knowledge source (Passive/AI-directed vs. Collaborative/AI-supported)	Cognitive Presence (Triggering, Exploration, Integration, Resolution)
Xu et al. (2025)	36 Undergraduate/Graduate Students	English (EFL Vocabulary Learning)	Instructional videos for MOOCs and online lectures	AI-generated instructor (Providing behavioral fidelity and visual cues)	Social Presence
Yu et al. (2025)	108 Pre-service Teachers	Education (Integrating IT in Classroom)	Six-hat thinking technique scaffolded discussion	Partner for information retrieval and problem-solving	Cognitive Presence
Darban (2026)	356 University students (48 virtual teams)	Management Information Systems (MIS)	Team-based online collaborative learning	AI teammate (Explainable ChatGPT)	Social Presence (Mediator of intention to continue learning)
Dong et al. (2026)	400 University students (+12 interviewees)	Multidisciplinary, including Liberal Arts, Science, and Engineering	Generative AI-based learning analytics (GLAT)	Learning analytics advisor / Personalized learning support	Teaching, Social, and Cognitive Presence (Mediator of learning engagement)
Gunbatar & Gokcearslan (2026)	233 University students	Instructional Technologies and Sport Sciences	ChatGPT-assisted instruction vs. Open inquiry + lecturing	Teaching assistant	Teaching Presence, Cognitive Presence (Higher CoI index)
Özbilen et al. (2026)	249 Preservice teachers	Educational Leadership	AI-supported self-regulation + social-media co-regulation	Self-regulation support / Human–AI collaboration	Shared Metacognition (Extension of CoI framework)
Pan et al. (2026)	High school Algebra students	Mathematics (K–12 Algebra)	Micro-blended learning (Teacher-led instruction + AI-facilitated online learning)	Co-facilitator (Adaptive scaffolding, feedback, socio-emotional support)	Teaching, Social, and Cognitive Presence (Distributed CoI across teacher and AI)
Xia et al. (2026)	70 University students	Science/Technology (specifically fMRI technology)	Virtual companion AI during online learning	Virtual companion (Identity-anthropomorphized AI)	Social Presence (Sequential mediator with positive emotion)
Xu et al. (2026)	132 Primary School Pupils	Mathematics (Probability Problems)	Instructional video	Animated Pedagogical Agent (PA)	Social Presence
Yang et al. (2026)	63 Writers	Creative and Argumentative Writing	Generative AI-assisted writing process	Writing assistant (CoAuthor tool)	Cognitive Presence

ONLINE LEARNING
Volume 30 · Issue SI2 · 2026 · e5893
https://doi.org/10.24059/olj.v30iSI2.5893

Systematic Review of Retention and Learning Engagement in Online Learning from 2015–2025

Curtis J. Bonk
Indiana University, Bloomington
Meina Zhu
Wayne State University
Min Young Doo
Kangwon National University

Abstract

This study aims to explore retention issues in online learning environments, with a primary focus on the role of learning engagement. Using the PRISMA approach, we analyzed 19 publications addressing engagement and retention in online learning environments from 2015 to 2025. In terms of publication productivity on this topic, 2019 was the most productive year, followed by 2020 and 2022. Regarding learning contexts, more than half of the studies were conducted in MOOC environments. Concerning influential factors of engagement and retention, instructor-related factors or instructional strategies were the most frequently examined, followed by learner-related individual factors and course characteristics. In terms of intervention strategies, games and gamification were the most frequently implemented in these studies, followed by asynchronous online discussions. Regarding types of engagement, multiple dimensions of engagement were most frequently examined, while behavioral engagement was the most studied as a single dimension. This study suggests that engagement is a critical factor in retention in online learning and also reveals an imbalance in the engagement literature with a need for more research on emotional engagement as well as more attention paid to engagement opportunities in various forms of blended learning. Future studies are suggested to explore underrepresented areas that might contribute to enhancing retention in online learning.

Keywords: Learning engagement, retention, online learning, systematic review

Introduction

Retention issues in online learning have persisted since the emergence of distance education (Bawa, 2016; Gunawardena & Zittle, 1997; Seery et al., 2021). de Freitas et al. (2015) argued that high dropout rates represent a significant drawback, despite the strength of online learning in promoting equitable access to educational opportunities. Estimates of dropout rates in online learning vary widely. For example, some studies of online learning report dropout rates that are 10% to 20% higher than those of traditional courses (Bawa, 2016). Even more troubling, Daniel (2012) reported that completion rates of massive open online courses (MOOCs) in the initial years of MOOCs often were under 10 percent (Daniel, 2012). Due to the high dropout rates prevalent in online learning, Clow (2013) coined the term "funnel of participation" to describe participation patterns in online courses, borrowing the concept from marketing.

One influential factor in retention is students' engagement (Martin & Borup, 2022), as it increases learning motivation and a sense of belonging, thereby enhancing learning achievement and performance and ultimately leading to higher levels of student satisfaction (Li et al., 2024; Martin & Bolliger, 2018). In response, the purpose of this study was to explore retention issues in online learning environments, with a primary focus on the role of learning engagement. Retention issues have gained continuous attention in online learning because dropout rates in

Corresponding author: Min Young Doo, mydoo@Kangwon.ac.kr
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

online courses are higher than those in traditional classroom settings (Lee & Choi, 2011). Many empirical studies have indicated that learning engagement is a predictor of learning outcomes and a key factor in reducing dropout rates (Bergdahl et al., 2020; Cortes, 2021; Fredricks et al., 2004; Vayre & Vonthron, 2017). However, "engagement" is a difficult construct to pin down making it extremely difficult to adequately measure. According to Newmann (1992), learning engagement is "students' psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or craft that academic work is intended to promote" (p. 12).

As delineated in common definitions of learning engagement and numerous empirical studies have reported (Doo & Kim, 2024; Lee et al., 2022; Martin & Borup, 2022), learning engagement is one of the primary influential factors of retention in online learning environments. Engagement is shaped by various factors, such as learner characteristics, learning environments, interactions with instructors and peers, and instructional design features. Instructional design and methods are particularly important because they can be designed and modified by instructors to support learning engagement. To provide practical implications for online learning instructors and administrators, the present study seeks to identify which instructional designs and technologies enhance learning engagement in ways to support higher retention in online learning, as well as further review the relationship between engagement and retention.

Literature Review

Given the significance of retention issues in online learning, numerous empirical studies, as well as systematic reviews and meta-analyses, have synthesized research findings on retention and learning engagement in online learning settings. In this review, we include representative studies addressing retention and engagement in both online learning environments and Massive Open Online Courses (MOOCs).

Learning Engagement Issues

As one of the most recent relevant studies, Doo and Kim (2024) conducted a meta-analysis synthesizing empirical studies that investigated the relationship between learning engagement and learning outcomes in online learning. The analysis included studies published between 2010 and March 2022. In total, Doo and Kim (2024) reviewed 34 empirical studies and found a small-to-medium overall effect size (g = .430). No significant differences were found among behavioral, cognitive, emotional, and general types of learning engagement. However, learning engagement showed a significantly stronger relationship with outcomes in the affective learning domain than in the cognitive and behavioral domains. Additionally, the relationship between learning engagement and learning outcomes was stronger among graduate students than among undergraduate students or adult learners.

Given that MOOCs have received substantial attention in engagement research in online learning, Wang et al. (2022) conducted a literature review on engagement in MOOCs, examining publications from 2015 to 2022. They reviewed 30 studies focusing on learning engagement in MOOCs. The types of data used as learning engagement indicators varied across studies. The most frequently used data collection methods were self-report measures (e.g., surveys) and automated analyses of log data obtained from online learning platforms.

Regarding influential factors, learning engagement was categorized into two broad types: (1) internal factors, and (2) external factors. Internal factors included learning satisfaction, academic motivation, achievement, competencies, autonomy, and self-regulated learning. External factors encompassed interaction with others, curriculum design, certificates, and technical support.

Retention Issues and Solutions

Regarding one of the primary variables of this study, Seery et al. (2021) reviewed 18 articles published between January 2015 and March 2019 and identified five themes related to retention issues in online learning. The most frequently mentioned factor was student affective and psychological factors (31.8%), followed by school-related factors, including institutional, faculty, and course factors (27.3%). Other factors included students' demographic and environmental factors (18.2%) and learning and academic factors (13.6%). Technological factors (9.1%) were the least frequently mentioned in relation to students' retention issues.

Using different types of learner groups, Paton et al. (2018) conducted a literature review on engagement and retention in vocational education and training (VET) online learning and MOOCs published between 2013 and 2017. Paton and her colleagues focused on VET programs and MOOCs because these contexts differ from traditional university education in terms of program needs, entry requirements, student age, and course duration. They analyzed 38 articles and identified 11 components related to engagement and retention. The most significant components included learners' sense of community, instructor commitment, instructional design, and assessment design.

Despite the equitable access and learning opportunities offered by online learning, as noted by de Freitas et al. (2015), there are concerns about access and engagement. Along these lines, Gardner and Leary (2023) conducted a systematic review focusing on first-generation and underrepresented minority students in online learning. They identified 15 reasons for low engagement among these students and collapsed them into three themes: (1) learner characteristics, such as language proficiency, sense of belonging, attitudes toward success, and readiness for online learning; (2) personal environments, including digital inequity, family obligations and support, and economic constraints; and (3) course-related factors, such as course design, interaction with peers and instructors, and course load. From this framework, it is clear that retention issues in online environments are multi-faceted and complex.

Strategies to Enhance Engagement and Retention in Online Learning

To achieve practical implications to enhance engagement and retention in online learning, we summarized each author's suggestions. At the end of their systematic review, Gardner and Leary (2023) emphasize that accurate recognition of where online learners are facing challenges in their student engagement framework or environment can help reveal or confirm what types of support are needed and allow for more targeted interventions. For instance, if lack of motivation is an issue, a mentoring program (Torres-Gracia, Zamudio-Orozco, & Garcia, 2026) could be considered as well as embedding mechanisms to acknowledge learning milestones. If a digital divide issue is apparent, learners might be made aware of campus technology or support resources or free and open educational resources from the world community (Bozkurt, et al., 2023; Canessa & Pisani, 2013; Naidu, 2023; UNESCO, 2022). If there are apparent issues with the course design features or structures, there might be a need for real world examples and

chances to apply course content. Flexible course design and delivery (Barua & Lockee, 2024) with opportunities for prompt feedback might also prove advantageous. In addition, such ideas can address cultural challenges and barriers.

Research from Seery et al. (2021) identified several areas for improvement to enhance retention. Course development (29.4%), student success support (26.5%), and faculty involvement (23.5%) together accounted for nearly 80% of the recommendations. Other, less frequently mentioned areas were related to engagement, including social engagement (11.8%) and emotional engagement (8.8%). Similarly, Paton et al. (2018) proposed several directions for future research to enhance engagement and retention, including instructionally sound course design, assessments consistent with learning objectives, opportunities for peer collaboration, sustained instructor communication, certification, and opportunities for further study.

In their study on student perceptions related to the importance of engagement in online learning environments, Martin and Bolliger (2018) surveyed 155 students to identify effective engagement strategies in online learning. They categorized these strategies into learner-to-learner, learner-to-instructor, and learner-to-content interactions, which align with Moore's (1993) three types of interaction in distance education. The findings indicated that learner-to-instructor strategies were perceived as the most important, including regular announcements or email reminders and the provision of grading rubrics for all assignments. Regarding learner-to-learner strategies, icebreaker or introductory discussions and collaborative work using online communication tools were identified as effective. For learner-to-content interaction, real-world projects and discussions guided by structured questions were perceived as beneficial for enhancing engagement in online learning.

Padilla Rodriguez et al. (2020) observed that participation in online learning tends to decrease as courses progress, contributing to low completion rates. Based on this finding, they compared different lengths of courses—one 6-week course and two 3-week courses. The results showed that students enrolled in the two 3-week courses were more active and engaged in learning.

As we have also shown, online learning engagement can vary drastically depending upon the timing, extent, and immediacy of the interaction as seen through the lens of synchronous versus asynchronous interaction (Bonk et al., 1998; Bonk, 2020; Park & Bonk, 2007). For instance, synchronous interactions (i.e., where participants learn at the same time) often result in briefer but more numerous interactions than asynchronous interactions (i.e., where participants learn at different times). On the other hand, asynchronous computer conferencing in online courses has been shown to foster more thoughtful and extended conversations and reflections than synchronous ones (Bonk et al., 1998).

In reviewing over a decade of fully online and blended courses at the University of Illinois Springfield, Boles et al. (2010) concluded that small class sizes of 25 students or less can lead to closer faculty and student relationships which can contribute to higher retention rates. Boles and her colleagues also found that student peer mentors can elevate retention by modeling best practices and facilitating online discussions, tracking participation, providing technology training when needed, lending key course insights, and answering student questions as they arise. Similarly, program coordinators can maintain close contact with students and keep track of student schedules and course progress, while supporting students who face particular barriers and challenges. Along these same lines, de Freitas et al. (2015) offered suggestions to make MOOCs

more engaging, such as incorporating gamification elements, using interactive learning content, providing immediate feedback on assessment results, and implementing personalized learning.

Among various engagement typologies, we adopted Fredricks et al. (2004) school engagement framework, a multidimensional model that encompasses behavioral, cognitive, and emotional engagement (Martin & Borup, 2022). In an online or blended course, behavioral engagement or involvement and effort might be seen in student participation in discussion forums, questions raised in an online presentation, making an online presentation, online sharing of one's projects, and suggesting relevant online resources. In contrast, cognitive engagement involves a learner's intellectual effort and investment in a class such as integrating seemingly disparate ideas in an online concept map or timeline, posting a weekly online blog or journal reflecting on what one has learned, and summarizing a discussion thread for key themes. Emotional or affective engagement might be displayed when sending positive emoticons during online group presentations, clapping or smiling at the end of an online guest speaker presentation, and social acknowledgements and thank you notes sent in an online discussion forum.

Accordingly, the research questions guiding this study are as follows:

1. What factors influence engagement and retention in online learning environments?
2. Which instructional methods and strategies are used to enhance learning engagement and retention in online learning environments?
3. Which types of online learning formats (e.g., fully online, blended, flipped, MOOCs, etc.) are most extensively researched in relation to engagement and retention?
4. Among behavioral, cognitive, and emotional engagement, which type is most frequently examined in relation to retention in online learning?

Methods

This study adopts a systematic review approach because retention is a broad concept that may not be fully explored through quantitative research such as meta-analyses (Martin, 2025). SPSS and Excel were utilized to help organize and analyze the data. To select eligible studies for this review, we applied the following inclusion criteria:

(1) studies that examined retention issues in online learning environments;

(2) peer-reviewed journal articles published within the last 10 years (since 2015);

(3) empirical research, including qualitative or quantitative studies employing a rigorous research design; and

(4) studies published in English.

We included all relevant research regardless of school levels, so school levels are neither inclusion nor exclusion criteria. Exclusion criteria included conceptual or theoretical papers and review articles such as meta-analyses or systematic reviews. Also excluded from this systematic review were non-peer-reviewed journals including dissertations and technical reports. Finally, we did not include publications prior to 2015, and studies published in languages other than English.

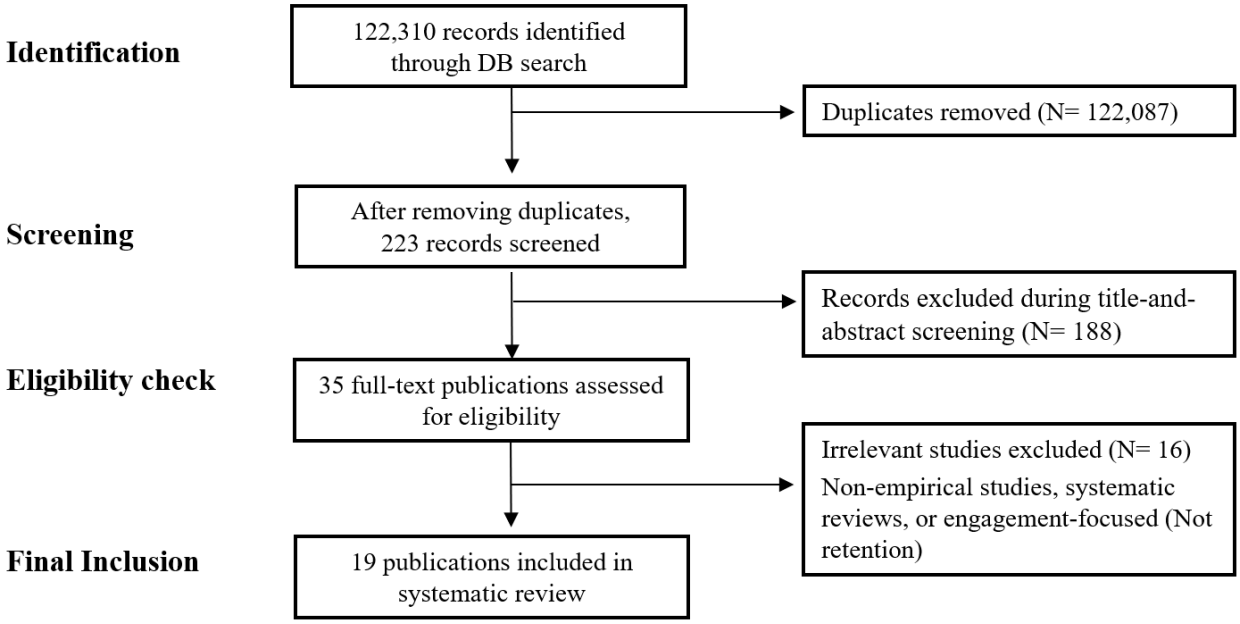
The keywords used for the search include “retention,” “dropouts,” “completion,” “learning engagement,” “cognitive engagement,” “emotional engagement,” “behavioral engagement,” “online learning,” “MOOCs,” “blended learning,” “flipped learning,” and their combinations. For the electronic database search, Google Scholar, Web of Science, ERIC (ProQuest), and Education Source (EBSCOhost) were used. The following Boolean search string was employed: (“retention” OR “dropouts” OR “completion”) AND (“learning engagement” OR “cognitive engagement” OR “emotional engagement” OR “behavioral engagement”) AND (“online learning” OR “MOOCs” OR “blended learning” OR “flipped learning”). Note that we purposefully relied on the above-mentioned four databases—Web of Science, Google Scholar, ERIC, and Education Source—due to Google Scholar's lack of reproducible indexing (Bramer, 2016; Tay, 2024). Hence, Google Scholar was useful as a supplement rather than as a primary database. In particular, Gusenbauer and Haddaway (2020) discovered extensive variations in the performance of different search engines. They point out that this variation requires substantive caveats when Google Scholar is used for systematic searches of the literature as the principal search engine.

To help ensure that no relevant studies are missed, an additional manual search was conducted in peer-reviewed journals highly relevant to the research topic, including *Online Learning Journal*, *Distance Education*, *American Journal of Distance Education*, and *The Internet and Higher Education*.

As a guideline for conducting this systematic review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework was used to identify studies eligible for analysis (Martin, 2025). The PRISMA framework is depicted in Figure 1. The quality of the studies identified through PRISMA were assessed using the Joanna Briggs Institute (JBI) checklist for systematic reviews (Aromataris et al., 2015; Aromataris et al., 2024). A recent study from Jia and Stern (2026) revealed that the JBI systematic reviews effectively utilize a variety of methods to determine studies for inclusion and exclusion based on the results of critical appraisal. In the present study, the uses of the JBI checklist found no major concerns in terms of methodology. Once the analysis scheme for this systematic review was established, two of the authors independently conducted full-text coding to determine study eligibility. The second author coded ten studies, and the third author independently coded nine. To assess inter-rater reliability, the third author randomly selected five of the ten studies coded by the second author, and the coding of these five overlapping studies was compared. Discrepancies were discussed and resolved by consensus, and coding definitions were refined when needed to clarify ambiguous cases. Notably, intercoder reliability was high (Cohen’s $k = .92$), and disagreements were resolved through discussion and additional review.

Figure 1

PRISMA Chart



Results

Descriptive data

We analyzed 19 publications that met our inclusion criteria. In terms of publication year, 2019 was the most productive year (four studies), followed by 2020 and 2022 (three studies each). Two studies were published in each of the following years: 2017, 2023, 2024, and 2025. Only one study was published in 2015, and no eligible publications were identified in 2021. The publications included in this study represented research conducted in 13 countries. China was the most productive country in research on engagement and retention in online learning ($N = 6$, 31.6%), followed by Australia and Taiwan ($N = 2$, 10.5% each). One publication each was contributed by researchers in Austria, Israel, Mexico, Colombia, Jamaica, Japan, Spain, the United Kingdom, the United States, and Vietnam. Notably, none of the 19 studies were specifically focused at the K-12 level.

Quantitative studies were the most prevalent research design (N = 8, 42.1%), including employing survey methods (e.g., Wang et al., 2019; You, 2022) and quasi-experimental designs (e.g., Huber et al., 2023; Rosser-Majors et al., 2022). The second most common approach involved the use of learning analytics (N = 5, 26.3%), such as analyzing participants' log data (e.g., Ortega-Arranz et al., 2019; Umer et al., 2017). Most of these studies were conducted in MOOC contexts. Three qualitative studies were identified (N = 3, 15.8%), including case studies (e.g., de Freitas et al., 2015) and in-depth interviews (e.g., Kahn et al., 2017). Other research designs included two (N = 2, 10.5%) mixed-methods studies (e.g., Hsu et al., 2025; Lan et al., 2020) and one action research study (e.g., Gay et al., 2020).

Across the 17 studies reporting participant data, the average sample size was 122,394.94 (*SD* = 502,182), ranging from eight (Kahn et al., 2017) to 2,071,147 (Wang et al., 2019). Two studies were excluded from this calculation: one study that used a "course" as the unit of analysis (Umer et al., 2017), and another study that employed a case study design (de Freitas et al., 2015). The substantial variation in sample sizes reflects the diversity of research methods, ranging from in-depth interviews to large-scale analyses of MOOC log data. Notably, Wang et al.'s (2019) study of over two million participants in a MOOC analyzed log data from a MOOC platform, resulting in their exceptionally large sample. Subtracting this outlier study, the average sample size decreased to 597.90 (*N* = 16).

The school levels of participants varied across studies, including high school students, undergraduates, graduate students, adult learners, and mixed-level samples. Undergraduate students comprised the largest proportion (42.1%), followed by studies including participants from all educational levels (31.6%). High school students, graduate students, and adult learners were each represented in one study.

Influential Factors of Engagement and Retention

Beyond identifying frequently studied variables, a synthesis of the reviewed studies revealed three distinct pathways through which engagement appeared to support retention. First, many studies conceptualized retention as a consequence of behavioral participation. Second, several studies suggested that retention is influenced by cognitive investment in learning rather than participation alone. Third, retention appeared connected to psychological and affective factors, even when emotional engagement was not explicitly measured. The primary dependent variables were course completion (N = 9, 47.4%), retention (N = 4, 21.1%), dropout rate (N = 2, 10.5%), and intention of retention (N = 1, 5.3%).

Regarding independent variables, instructor-related factors or instructional strategies were the most frequently examined (N = 5, 26.3%). These included game-based learning and gamification (Hsu et al., 2025; Huber et al., 2023; Ortega-Arranz et al., 2019), instructor training (Rosser-Majors et al., 2022), and high-touch student engagement strategies (Gay et al., 2020). Learner-related individual factors were the second most frequently investigated category (N = 4, 21.1%), encompassing demographic characteristics (Deng et al., 2017), psychological factors (Vo & Ho, 2024; Yang et al., 2020), and metacognitive reflexivity (Kahn, 2017). Course characteristics were also commonly examined (*N* = 4, 21.1%), particularly course length (Padilla Rodriguez et al., 2020) and assessment difficulty (de Freitas et al., 2015). Additionally, one study explored the effects of learning support on retention (Fan et al., 2024).

Instructional Methods and Strategies for Engagement and Retention

The second research question examined instructional methods and strategies designed to enhance learning engagement and retention in online learning environments. Because several studies included multiple courses or MOOCs, specific instructional methods or strategies were not reported in eight studies (*N* = 8, 42.1%). Among the reported approaches, games and gamification were the most frequently implemented strategies (*N* = 4) (e.g., Huber et al.; Ortega-Arranz et al.). Asynchronous online discussions were also commonly used (*N* = 3) (e.g., Gay et al.; Kahn et al.). Other instructional approaches included the integration of multiple strategies within a single course (e.g., Rosser-Majors et al.), active learning (e.g., Lan et al.), e-book–based

learning (e.g., Kuromiya et al.), and multimedia-supported teaching with instructor interaction (e.g., Fan et al.).

Types of Online Learning Contexts in Engagement and Retention Research

The third research question compared the types of online learning formats examined in studies of engagement and retention. The findings indicate that more than half of the studies were conducted in MOOC contexts ($N = 10$, 52.6%) (e.g., Deng et al., 2019; Hsu et al., 2025; Wang et al., 2019), outnumbering other forms of online learning. Fully online courses represented the second most frequently examined format ($N = 6$, 31.6%) (e.g., Huber et al., 2023; Kuromiya et al., 2022; Soffer et al., 2019). Additionally, two studies were conducted in online degree programs (Kahn et al., 2017; Vo & Ho, 2024). Only one study examined engagement and retention in the context of blended learning (Fan et al., 2024).

Types of Learning Engagement Associated with Retention

Using Fredricks et al.'s (2004) framework as an analytic lens revealed an important pattern across the reviewed literature. Although engagement is theoretically understood as a multidimensional construct encompassing behavioral, cognitive, and emotional dimensions, retention research has concentrated disproportionately on behavioral engagement. In effect, we analyzed and compared the types of learning engagement examined by the 19 studies included in this systematic review. Fredricks et al.'s (2004) school engagement framework was the most commonly used. Because each publication adopted different theoretical frameworks or typologies of engagement, direct comparisons using a single universal classification were not appropriate. As a single dimension, behavioral engagement was the most frequently examined ($N = 7$, 36.8%). In contrast, only one study focused exclusively on cognitive engagement in the context of engagement and retention. No studies examined social, affective, or emotional engagement as standalone dimensions.

An equivalent proportion of studies ($N = 7$, 36.8%) investigated multiple dimensions of engagement simultaneously. For example, Deng et al. examined behavioral, cognitive, emotional, and social engagement; Vo and Ho (2024) examined affective, cognitive, and behavioral engagement; and Soffer et al. examined cognitive, behavioral, and emotional engagement. Finally, four studies (21.1%) conceptualized engagement as a general, unidimensional construct rather than distinguishing among specific dimensions (e.g., She et al.; Yang et al., 2023).

Discussion

The purpose of this review was to synthesize recent empirical evidence on engagement and retention in online learning environments and to identify instructional and contextual factors that support sustained participation. Consistent with long-standing concerns about attrition in distance education (Bawa, 2016; Lee & Choi, 2011), the findings reaffirm that retention remains a central challenge across diverse online formats, particularly in MOOCs. At the same time, this review advances the field by systematically mapping instructional strategies, learning contexts, and engagement dimensions examined in relation to retention, revealing both convergences with prior research as well as critical gaps in that research.

Compared with previous systematic reviews, the present study offers several distinct contributions. Rather than reviewing engagement or retention independently, this review synthesizes empirical evidence examining the relationship between these two constructs across multiple online learning contexts. Furthermore, by organizing the literature using Fredricks et al.'s (2004) multidimensional engagement framework, the review identifies a clear imbalance in the existing evidence, with behavioral engagement receiving substantially greater attention than cognitive or emotional engagement. Finally, this review integrates instructional strategies, engagement dimensions, learning contexts, and retention outcomes into a single synthesis, providing a more comprehensive perspective on how instructional design can foster learner persistence in online learning environments. These contributions extend prior reviews by highlighting both theoretical gaps and practical opportunities for future research and instructional practice.

In addition, the present study includes more recent research than any of the previous studies mentioned above as it entails publications from 2015-2025. This is important as online learning has become more prominent and familiar to educators as a result of the COVID-19 pandemic which arose in 2020. The past decade has also witnessed an explosion of new online and blended learning format and delivery mechanisms such as MOOCs (Bonk et al., 2015; Zhu et al., 2020), HyFlex forms of learning (Beatty, 2019), and micro-credentials (Higher Education Commission in Mauritius, 2025; Martin & Mhicil, 2025).

Engagement as a Central Mechanism for Retention

Across the reviewed studies, we found that course completion and retention were the most frequently examined dependent variables, reinforcing the field's focus on persistence outcomes. The prominence of instructor-related and instructional design variables as independent predictors aligns with prior syntheses emphasizing the importance of course development, faculty involvement, and instructional quality (Paton et al., 2018; Seery et al., 2021). These findings extend earlier work suggesting that engagement serves as a key mechanism linking instructional design to learning outcomes (Bergdahl et al., 2020; Fredricks et al., 2004).

The present results are consistent with Doo and Kim's (2024) recent meta-analysis, which found a significant positive relationship between engagement and learning outcomes in online settings. Although Doo and Kim (2024) reported no significant differences among behavioral, cognitive, and emotional engagement in predicting outcomes, this review indicated that behavioral engagement is disproportionately emphasized in retention studies. This discrepancy suggests that, while multidimensional engagement frameworks are theoretically recognized (Fredricks et al., 2004), empirical investigations of retention tend to operationalize engagement through observable behaviors such as logins, assignment submissions, and discussion participation, particularly in learning analytics and MOOC-based research.

This behavioral emphasis may partially reflect the availability of large-scale log data (Wang et al., 2022). However, given that affective and psychological factors were identified as the most frequently mentioned retention-related themes in prior reviews (Seery et al., 2021), the limited standalone examination of emotional or cognitive engagement in the current corpus signals a theoretical and methodological imbalance. Emotional engagement, such as sense of belonging and enjoyment, has been previously linked to persistence and satisfaction (Martin & Bolliger, 2018); yet, emotional engagement remains underexplored as an independent predictor of

retention. Thus, this study highlights a critical need for future research to incorporate multidimensional and theory-driven measures of engagement beyond behavioral proxies.

Instructional Design and Gamification as Prominent Strategies

Consistent with earlier recommendations (de Freitas et al., 2015; Paton et al., 2018), gamification and game-based learning emerged as the most frequently implemented instructional strategies to enhance engagement and retention. This trend suggests growing interest in motivational design features to counteract the "funnel of participation" described by Clow (2013). Gamified elements, such as badges, leaderboards, and progress tracking, may foster behavioral persistence and short-term engagement (Meng, et al., 2024); however, whether they cultivate deeper cognitive or emotional investment remains an open question.

Asynchronous discussions were another commonly reported strategy, echoing Moore's (1993) interaction framework and Martin and Bolliger's (2018) findings that learner-instructor and learner-learner interactions are perceived as critical to engagement. These findings reinforce the notion that social presence and instructor communication are central to sustained participation. In contrast, relatively fewer studies explicitly examined structured student support systems, despite prior reviews emphasizing their importance for retention (Seery et al., 2021). This gap suggests that while course-level instructional strategies receive extensive or exclusive attention, institutional-level support mechanisms may be underrepresented in engagement-focused research.

Course characteristics, such as length and assessment difficulty, were also identified as influential. The finding that shorter or modularized courses may promote sustained participation aligns with Rodriguez et al.'s (2021) observation that, in longer courses, engagement declines over time. This pattern resonates with the broader literature on temporal attrition in MOOCs (Daniel, 2012), MOOC-like derivatives, and other online learning formats. It also suggests that instructional pacing and workload design are critical levers for improving retention.

MOOC Dominance and Implications for Generalizability

More than half of the reviewed studies were conducted in MOOC contexts, reflecting the availability of large-scale data and the visibility of retention challenges in open-access environments. While MOOCs provide valuable insights into participation patterns at scale, they differ substantially from credit-bearing online courses and degree programs in terms of learner motivations, entry requirements, and accountability structures (Paton et al., 2018). Consequently, the dominance of MOOC-based research may limit the generalizability of findings to other online learning formats.

Fully online courses constituted the second most frequently examined format, whereas blended learning was minimally represented. Given the increasing adoption of hybrid and blended models in K-12 (Tucker, 2020) and higher education (Beatty, 2019; Stein & Graham, 2014), this imbalance highlights an important direction for future inquiry. Perhaps the variety of blended learning formats, structures, and components make it more complex and difficult to study and coherently report about. Along these same lines, blended learning research might be more difficult to understand and interpret due to the multiple components and easily changeable ingredients. At the same time, it is plausible that blended formats are transformed and replaced

more quickly than fully online ones. Comparative studies across formats could clarify whether engagement–retention dynamics differ based on course structure, learner population, or institutional context.

Learner Characteristics and Equity Considerations

Although instructor-related variables were most frequently studied, learner-related individual factors, including psychological variables, metacognitive reflexivity, and demographic characteristics, also emerged as influential. These findings echo Gardner and Leary's (2023) review, which emphasized learner characteristics, personal environments, and course-related factors as contributors to low engagement among first-generation and underrepresented students.

Notably, technological factors were rarely foregrounded in prior reviews (Seery et al., 2021), and our findings similarly suggest that retention challenges are less about access to technology tools and resources per se and more about different psychological, pedagogical, and design factors. This finding underscores the importance of viewing retention through a socio-ecological lens that integrates individual, instructional, and contextual dimensions rather than attributing attrition solely to technical barriers. Refined adjustments to that lens will require concerted efforts of researchers in this space both to conduct research on retention and learning engagement in online environments as well as to translate their findings so that educators and instructional designers can implement them for increasingly effective and higher quality courses and experiences.

Practical and Policy Implications

The findings of this review have important implications for instructional designers, online instructors, institutional administrators, and educational leaders seeking to improve learner engagement and retention in online learning environments.

Implications for instructional designers and instructors.

The findings suggest that engagement should be intentionally designed rather than assumed to emerge naturally during online learning. Because instructor-related factors and instructional strategies were the most frequently identified influences on engagement and retention, instructional designers should incorporate evidence-based engagement strategies throughout course development. These include designing regular instructor communication, providing timely and personalized feedback, facilitating structured asynchronous discussions, integrating collaborative learning opportunities, and using gamification or other motivational features when aligned with learning objectives. Importantly, instructional designers should move beyond promoting behavioral engagement alone by designing activities that also foster cognitive engagement through reflection and problem-solving as well as emotional engagement by cultivating learners' sense of belonging, motivation, and connection to the learning community.

Implications for institutional administrators.

Improving student retention requires coordinated institutional support in addition to effective course design. Institutions should invest in professional development that equips faculty with strategies for promoting online engagement and sustaining instructor presence. Learning

management system analytics can be leveraged to identify students exhibiting early signs of disengagement, enabling timely outreach and academic support before withdrawal occurs. Institutions should also establish integrated student success services, including academic advising, tutoring, technical support, and mental health resources, that address the multiple factors influencing learner persistence identified across the reviewed studies.

Policy implications.

At the institutional level, learner engagement and retention should be recognized as shared organizational responsibilities rather than solely the responsibility of individual instructors. Institutions should develop policies and quality standards that encourage evidence-informed online course design, regular monitoring of engagement indicators, and continuous evaluation of student retention initiatives. Resource allocation should prioritize faculty development, instructional design support, learning analytics infrastructure, and student support services that have the potential to improve persistence in online learning. As online and blended learning continue to expand across higher education, institutional policies that systematically support learner engagement may contribute to improved student success, equity, and completion rates.

Limitations

This study has several limitations that need to be mentioned. First, none of the studies included in this systematic review of the research was focused squarely on the K-12 level. As K-12 schools and school districts increasingly provide online and blended learning course options (Stein & Graham, 2014; Tucker, 2020), it is vital to better understand how learner retention and engagement are impacted when youth engage in highly digital forms of learning. Another limitation was that only peer-reviewed journal articles published in English within the past 10 years were included, which may have excluded relevant studies published in other languages and contributed to the relatively small sample size ($N = 19$). Given that online learning emerged during the 1990s, there may be informative research on online learning retention, motivation, and engagement prior to 2015. In addition, we might have excluded important discussions of online learner engagement and retention found in book chapters, conference proceedings, institutional reports, newspapers, and dissertation theses. Consequently, we were unable to obtain a complete picture of the country of origin of MOOC research due to our restriction to publications in English journals.

We also are aware that by focusing specifically on engagement and retention together in online learning environments, studies addressing only engagement or only retention may not have been fully captured. It is conceivable that many individual studies used different terminology—such as retention, dropout, or completion—which may have limited the comprehensiveness of the search and selection process. Similarly, as mentioned earlier, learner engagement is a multi-dimensional or multi-faceted construct that could refer to behavioral, cognitive, or emotional engagement or some mix between these three aspects of engagement. Such complexity makes it extraordinarily hard to review the existing literature and summarize key data patterns found in it.

Finally, there are inherent limitations in the use of systematic reviews of the research instead of primary research data studies. For instance, systematic reviews are compiled summaries of the research data in studies meeting set criteria. As a result, the analyzed data sets are often stripped

of unique characteristics or specific findings or suggestions for the field. In addition, such reviews pool individual studies that may each suffer from assorted biases, small sample sizes, and data selection and presentations issues. The studies meeting all the criteria may also be constrained by statistical heterogeneity in terms of research protocols and procedures, sampling, and outcome measures. Researchers should be aware of these various limitations when conducting follow-up studies as they seek to make important contributions to the online and blended learning fields.

Future Directions

Among the future research possibilities, future advances in generative AI technology might enable novel forms of learner feedback, mentoring, resource allocation, and content accessibility that could result in enhanced insights into effective learner engagement and retention strategies and practices (Rad et al., 2024) whether in solitary learning pursuits or in team competitions (Bonk et al., 2025; Kim, 2024). It is plausible that AI assistants like chatbots might transform learning environments in ways that elevate learner retention across educational sectors and learner age brackets. Clearly, there are myriad opportunities for researchers to explore various forms of AI technology in education as well as generative AI pedagogy for learning engagement and retention.

One thing is certain, new waves of technology tools will emerge during the coming decade or two for educators to experiment with as they attempt to enhance learning engagement, and, in turn, elevate learner retention. As they do, new forms of behavioral, cognitive, and emotional or affective engagement will undoubtedly arise across every sector of education from young learners to adults. At the same time, online and blended learning derivatives, expansions, enhancements, and evolvements will transform the entire educational ecosystem (Kim & Chung, 2015; Kim & Lee, 2020) and create learning delivery formats and opportunities that few, if any, are contemplating or planning for today. As that happens, learning engagement and its steadfast companion, learning retention, will continue to be critical variables for educational researchers to explore and for instructional designers, institutional administrators, and government policy makers to support financially, conceptually, technologically, and pragmatically.

Declarations

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest Statement

The authors have no competing interests to declare that are relevant to the content of this article.

References

Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., & Tungpunkom, P. (2015). Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-based Healthcare*, *13*(3), 132–140. https://doi.org/10.1097/XEB.0000000000000055

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (Eds.). (2024). JBI Manual for evidence synthesis. JBI. Available from: https://synthesismanual.jbi.global. https://doi.org/10.46658/JBIMES-24-01 https://synthesismanual.jbi.globalhttps://doi.org/10.46658/JBIMES-24-01

Barua, L., & Lockee, B. B. (2024). A review of strategies to incorporate flexibility in higher education course designs. *Discover Education*, *3*, 127 (2024). https://doi.org/10.1007/s44217-024-00213-8

Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—A literature review. *SAGE Open, 6*(1), 1–11. https://doi.org/10.1177/2158244015621777

Beatty, B. J. (2019). *Hybrid-flexible course design (1st ed.)*. EdTech Books. https://edtechbooks.org/hyflex

Bergdahl, N., Nouri, J., Fors, U., & Knutsson, O. (2020). Engagement, disengagement and performance when learning with technologies in upper secondary school. *Computers & Education*, *149*, Article 103783. https://doi.org/10.1016/j.compedu.2019.103783

Boles, E., Cass, B., Levin, C., Schroeder, R., & Smith, S. (2010). Sustaining students: Retention strategies in an online program. *EDUCAUSE.* https://er.educause.edu/articles/2010/12/sustaining-students-retention-strategies-in-an-online-program

Bonk, C. J., Hansen, E. J., Grabner, M. M., Lazar, S., & Mirabelli, C. (1998). Time to "Connect": Synchronous and asynchronous case-based dialogue among preservice teachers. In C. J. Bonk, & K. S. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse* (pp. 289-314). Mahwah, NJ: Erlbaum.

Bonk, C. J., Lee, M. M., Reeves, T. C., & Reynolds, T. H. (Eds.). (2015). *MOOCs and open education around the world.* NY: Routledge. https://doi.org/10.4324/9781315751108

Bonk, C. J. (2020). Pandemic ponderings, 30 years to today: Synchronous signals, saviors, or survivors? *Distance Education*, *41*(4), 589-599. https://doi.org/10.1080/01587919.2020.1821610

Borup, J., Graham, C. R., West, R. E., Archambault, L., & Spring, K. J. (2020). Academic communities of engagement: An expansive lens for examining support structures in blended and online learning. *Educational Technology Research and Development*, *68*(2), 807–832. https://doi.org/10.1007/s11423-020-09744-x

Bozkurt, A., et al. (2023). Openness in education as a praxis: From individual testimonials to collective voices. *Open Praxis*, *15*(2), pp. 76–112. https://doi.org/10.55982/openpraxis.15.2.574

Bramer, W. M. (2016). Variation in number of hits for complex searches in Google Scholar. *Journal of the Medical Library Association*, *104*(2), 143-145. https://doi.org/10.3163/1536-5050.104.2.009

Canessa, E., & Pisani, A. (2013). High school open on-line courses (HOOC): A case study from Italy. *European Journal of Open, Distance and E-Learning*, *16*(1), 131-140. http://www.eurodl.org/materials/contrib/2013/Canessa_Pisani.pdf

Clow, D. (2013). MOOCs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)* (pp. 185–189). ACM. https://doi.org/10.1145/2460296.2460332

Cortes, K. (2021). *Students' perceptions of engagement in online courses and its effect on academic performance and retention rates* [Unpublished doctoral dissertation]. Arizona State University.

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education, 2012*(3), Article 18. https://doi.org/10.5334/2012-18

*de Freitas, S. I., Morgan, J., & Gibson, D. (2015). Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *British Journal of Educational Technology, 46*(3), 455–471. https://doi.org/10.1111/bjet.12268

*Deng, R., Benckendorff, P., & Gannaway D (2020). Linking learner factors, teaching context, And engagement patterns with MOOC learning outcomes. *Journal of Computer Assisted Learning, 36*(5), 688–708. https://doi.org/10.1111/jcal.12437

Doo, M. Y., & Kim, J. (2024). Leaning engagement and learning achievement in online learning in higher education: A meta-analysis. *Distance Education, 45*(1), 60-82. https://doi.org/10.1080/01587919.2024.2303484

*Fan, S., Trimble, A., Kember, D., Muir, T., Douglas, T., Wang, Y., Masters, J., & Mainsbridge, C. (2024). Supporting engagement and retention of online and blended-learning students: A qualitative study from an Australian University. *The Australian Educational Researcher, 51*, 403–421. https://doi.org/10.1007/s13384-022-00605-5

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gardner, K., & Leary, H. (2023). Online learning for first-generation and underrepresented minoritized students: A literature review using a model of student engagement. *Online Learning, 27*(1), 263–291. https://doi.org/10.24059/olj.v27i1.3392

*Gay, G. H. E., & Betts, K. (2020). From discussion forums to eMeetings: Integrating high touch strategies to increase student engagement, academic performance, and retention in large online courses. *Online Learning, 24*(1), 92–119. https://doi.org/10.24059/olj.v24i1.1984

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education, 11*(3), 8–26. https://doi.org/10.1080/08923649709526970

Gusenbauer, M., & Haddaway, N. R. (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*, 181–217. https://doi.org/10.1002/jrsm.1378

Higher Education Commission in Mauritius (2025, August). A blueprint on micro-credentials in higher education: From concept to execution. https://www.hec.mu/pdf_downloads/NCVTS_Blueprint/MC%20BROCHURE_08102025_Final01.pdf

*Hsu, W.-S. (2025). Analyzing student engagement and learning outcomes in a gamified blended cybersecurity course using LMS-based behavioral data. *International Journal of Engineering Pedagocy (iJEP), 15*(6), 111-122. https://doi.org/10.3991/ijep.v15i6.57015

*Huber, S. E., Cortez, R., Kiili, K., Lindstedt, A., & Ninaus, M. (2023). Game elements enhance engagement and mitigate attrition in online learning tasks. *Computers in Human Behavior, 149*, Article 107948. https://doi.org/10.1016/j.chb.2023.107948

Jia, R. M., & Stern, C. (2026). The inclusion or exclusion of studies based on critical appraisal results in JBI qualitative systematic reviews: An analysis of practices. *Research Synthesis Methods, 17*(2), 277-292. https://doi.org/10.1017/rsm.2025.10042

*Kahn, P., Everington, L., Kelm, K., Reid, I., & Watkins, F. (2017). Understanding student engagement in online learning environments: The role of reflexivity. *Educational Technology Research and Development, 65*(1), 203–218. https://doi.org/10.1007/s11423-016-9484-z

Kim, P. (2024). Trends of AI in education and new competencies for the future at exploring the ethical use of AI TIP conference. YouTube. [Video]. https://youtube/DsUEhxlNku8?si=h2vXfm-P2zNfdg5o

Kim, P., & Chung, C. (2015). Creating a temporary spontaneous mini-ecosystem through a MOOC. In C. J. Bonk, Lee, M. M., T. C. Reeves, & T. H. Reynolds (Eds.), *MOOCs and open education around the world* (pp. 155-168). Routledge.

Kim, P., & Lee, J. (2020). Evolution of online learning environments and the emergence of intelligent MOOCs. In K. Zhang, C. J. Bonk, T. C. Reeves, & T. H. Reynolds (Eds.), *MOOCs and open education in the Global South: Challenges, successes, and opportunities* (pp. 329-341). Routledge.

*Kuromiya, H., Majumdar, R., Miyabe, G., & Ogata, H. (2022). E-book-based learning activity during COVID-19: Engagement behaviors and perceptions of Japanese junior-high

school students. *International Journal of Educational Technology in Higher Education, 17*(1), Article 12. https://doi.org/10.1186/s41039-022-00184-0

*Lan, M., & Hew, K. F. (2020). Examining learning engagement in MOOCs: A self-determination theoretical perspective using mixed method. *International Journal of Educational Technology in Higher Education, 17*, Article 7. https://doi.org/10.1186/s41239-020-0179-5

Lee, J., Park, T., & Davis, R. O. (2022). What affects learner engagement in flipped learning and what predicts its outcomes? *British Journal of Educational Technology, 53*(2), 211-228. https://doi.org/10.1111/bjet.12717

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development, 59*(5), 593–618. https://doi.org/10.1007/s11423-011-9207-0

Li, Z., Bonk, C. J., & Zhu, M. (2024). Community in the open: Supports, challenges, and impacts of local learning communities of K-12 adolescent MOOC learners from Nepal. *Online Learning*, *28*(3), 497-523. https://doi.org/10.24059/olj.v28i3.3463

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

Martin, F., & Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning, 22*(1), 205–222. https://doi.org/10.24059/olj.v22i1.1092

Martin, F., & Borup, J. (2022). Online learner engagement: Conceptual definitions, research themes, and supportive practices. *Educational Psychologist*, *57*(3), 162–177. https://doi.org/10.1080/00461520.2022.2089147

Martin, M., & Mhicil, M. N. G. (2025). Micro-credentials in Mauritius: towards a national framework for TVET and higher education UNESCO. https://doi.org/10.54673/dplj5721 or https://www.iiep.unesco.org/en/publication/micro-credentials-mauritius-towards-national-framework-tvet-and-higher-education

Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22-29). Routledge.

Naidu, S. (2023). From open access to open science—open education in transition. *Distance Education* (2023), 44(4), 599–603. https://doi.org/10.1080/01587919.2023.2278218

Newmann, F. M. (Ed.) (1992). *Student engagement and achievement in American secondary schools.* Teachers College Press.

*Ortega-Arranz, A., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Martínez-Monés, A., Gómez-Sánchez, E., & Dimitriadis, Y. (2019). To reward and beyond: Analyzing the effect of reward-based strategies in a MOOC. *Computers & Education, 142*, Article 103639. https://doi.org/10.1016/j.compedu.2019.103639

*Padilla Rodriguez, B. C., Armellini, A., & Rodriguez Nieto, M. C. (2020). Learner engagement, retention and success: Why size matters in massive open online courses (MOOCs). *Open*

Learning: The Journal of Open, Distance and e-Learning, 35(1), 46–62. https://doi.org/10.1080/02680513.2019.1665503

Paton, R. M., Fluck, A. E., & Scanlan, J. D. (2018). Engagement and retention in VET MOOCs and online courses: A systematic review of literature from 2013 to 2017. *Computers & Education, 125*, 338–346. https://doi.org/10.1016/j.compedu.2018.06.013

Park, Y. J., & Bonk, C. J. (2007, Winter). Synchronous learning experiences: Distance and residential learners' perspectives in a blended graduate course. *Journal of Interactive Online Learning*, *6*(3) 245-264.

Rad, H. S., Alipour, R., & Jafarpour, A. (2024). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application. *Interactive Learning Environments*, *32*(9), 5020–5040. https://doi.org/10.1080/10494820.2023.2208170

*Rosser-Majors, M. L., Rebeor, S., McMahon, C., Wilson, A., Stubbs, S. L., Harper, Y., & Sliwinski, L. (2022). Improving retention factors and student success online utilizing the Community of Inquiry framework's instructor presence Model. *Online Learning*, *26*(2), 6-33. https://doi.org/10.24059/olj.v26i2.2731

Seery, K., Barreda, A. A., Hein, S. G., & Hiller, J. L. (2021). Retention strategies for online students: A systematic literature review. *Journal of Global Education and Research*, *5*(1), 72-84. https://doi.org/10.5038/2577-509X.5.1.1105

*She, M., Xu, Y., Li, Z., & Hu, D. (2025). Exploring the determinants on continuance participation of college students toward blended learning: The stimulating role of course characteristics and instructor support. *Frontiers in Psychology, 16,* 1522810. https://doi.org/10.3389/fpsyg.2025.1522810

*Soffer, T., & Cohen, A. (2019). Students' engagement characteristics predict success and completion of online courses. *Journal of Computer Assisted Learning,* 35, 378–389. https://doi.org/10.1111/jcal.12340

Stein, J., & Graham, C. R. (2014). Essentials for blended learning: A standards-based guide. Routledge.

Tay, A. (2024, May 12). Why EXACTLY is Google Scholar bad for evidence synthesis, systematic reviews? *Aaron Tay Substack*. Available: https://aarontay.substack.com/p/why-exactly-is-google-scholar-bad-for

Torres-Gracia, C., Zamudio-Orozco, L., & Garcia, M. M. (2026). A qualitative examination of mentoring practices for first-generation, low income STEM students through a critical intersectionality lens. *SAGE Open*, 1-16. https://doi.org/10.1177/21582440261431861

Tucker, C. (2020). *Balance with blended learning: Partner with your students to reimagine learning and reclaim your life*. Corwin.

*Umer, R., Susnajac, A., Mathrani, A., & Suriadi, S. (2017). Prediction of students' dropout in MOOC environment. *International Journal of Knowledge Engineering, 3*(2), 19–27.

UNESCO (2022). The 2019 UNESCO recommendation on open educational resources (OER): Supporting universal access to information through quality open learning materials. UNESCO Digital Library, Paris. Available: https://unesdoc.unesco.org/ark:/48223/pf0000383205.locale=en

Vayre, E., & Vonthron, A.-M. (2017). Psychological engagement of students in distance and online learning: Effects of self-efficacy and psychosocial processes. *Journal of Educational Computing Research, 55*(2), 197–218. https://doi.org/10.1177/0735633116656849

*Vo, H., & Ho, H. (2024). Online learning environment and student engagement: The mediating role of expectancy and task value beliefs. *The Australian Educational Researcher, 51*(6), 2183–2207. https://doi.org/10.1007/s13384-024-00689-1

Wang, R., Cao, J., Xu, Y., & Li, Y. (2022). Learning engagement in massive open online courses. *Frontiers in Education, 7*, 1074435. https://doi.org/10.3389/feduc.2022.1074435

*Wang, W., Guo, L., He, L., & Wu, Y. J. (2019). Effects of social-interactive engagement on the dropout ratio in online learning: insights from MOOC. *Behaviour & Information Technology*, *38*(6), 621–636. https://doi.org/10.1080/0144929X.2018.1549595

*Yang, H-H. & Lin, J-Y. (2023). Students' persistence intention in MOOCs in the psychomotor domain: An extended 3P model of the teaching and learning perspective. *Frontiers in Psychology,* 14, 1094138. https://doi.org/10.3389/fpsyg.2023.1094138

*You, W. (2022). Research on the relationship between learning engagement and learning completion of online learning students. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(01), 102–117. https://doi.org/10.3991/ijet.v17i01.28545

Zhu, M., Sari, A. R., & Lee, M. M. (2020). A comprehensive systematic review of MOOC research: Research techniques, topics, and trends from 2009 to 2019. *Educational Technology Research and Development, 68*(4), 1685-1710. https://doi.org/10.1007/s11423-020-09798-x

*References marked with an asterisk indicate studies included in the systematic review

Appendix

Table 1

Coding Scheme for Study Characteristics

Category	Code	Definition
Engagement types	Behavioral	Engagement reflected in learning behaviors or participation
	Cognitive	Engagement involving mental effort, thinking, or investment in learning.
	Emotional	Affective status, reactions, feelings, or attitudes toward learning
	Multiple	Two or more engagement types were examined in the study
	Overall	Engagement was examined without specifying a particular type
Learning environment	Fully online	Entirely online learning environments, excluding MOOCs.
	MOOCs	Massive Open Online Courses
	Blended	Combination of online and face-to-face instruction.
	Flipped	A pedagogical approach in which direct instruction is moved from the group learning space to the individual learning space.
School levels	Elementary, middle, high, undergraduate, graduate, adult	Learner population identified in the study
	All	Study included multiple educational levels.
	Not applicable	Educational level was not reported
Research methods	Quantitative, qualitative, mixed methods, learning analytics, action research	Research design or analytical approach explicitly reported by the study
		Qualitative / Constructivist: Phenomenology, Case Study, Narrative etc.
		Quantitative / Positivist: Experimental, Quasi-experimental, Survey etc.
		Mixed Methods / Pragmatic: Explanatory Sequential, Convergent
		Critical / Transformative: Critical Ethnography
		Design-Based: DBR Cycles learning analytics, etc.
		Others: Learning analytics
Independent variables	Individual factors	Learner characteristics or individual-level factors
	Instructor/instructional strategies	Instructor characteristics or instructional approaches

Category	Code	Definition
	Course characteristics	Characteristics of the course or learning context
	Support for learning	Resources or support provided to facilitate learning
	Various	Multiple categories of factors were examined
	Not applicable	No independent variable was identified
Dependent variables	Retention	Continuation/persistence in the course or program
	Course completion	Successful completion of a course or learning activity
	Dropout rate	Withdrawal/non-completion rate
	Not reported	Relevant outcome was not reported
Instructional methods	Game/gamification	Game-based or gamified instructional approaches
	Active learning	Instructional approaches requiring active learner participation
	Course resources	Use of learning resources or materials as part of the instructional approach
	Lecture	Lecture-based instruction
	Asynchronous learning	Asynchronous instructional activities such as narrated videos and online assessments
	Online discussion	Discussion conducted through online platforms
	E-meeting	Synchronous online meetings
	Team project/group work	Collaborative project or group-based activities
	Various	Multiple instructional methods were reported without a single dominant method
	Not reported	Instructional methods were not reported
Country		Country was coded according to the location reported for the study

Table 2

Characteristics of the Included Studies

#	Authors	Years	Country	Research Design	School level	Sample size	Independent Variable	Engagement type	Dependent Variable	Instructional methods/strategies	Learning environment
1	de Freitas et al. (2015)	2015	Australia	Qualitative	Undergraduates	Case study	Course characteristics, instructional strategies	Behavioral engagement	Not reported	Games and play	MOOCs
2	Deng et al. (2020)	2019	China	Quantitative	All	1,452	Individual factors, Course characteristics	Multiple	Course completion	NA	MOOCs
3	Fan et al. (2024)	2024	Australia	Qualitative	Undergraduates	41	Support for learning	Overall engagement	Retention	Weekly narrated videos and assessments with ongoing interaction via online platforms	Blended
4	Gay & Betts (2020)	2020	Jamaica	Action research	Undergraduates	3,386	Instructor/Instructional strategies	Behavioral engagement	Course completion	e-meeting, discussion forums, group works	Fully online
5	Hsu (2025)	2025	Taiwan	Mixed methods	Undergraduates	99	Instructor/Instructional strategies	Overall engagement	Course completion	Gamification	MOOCs
6	Huber et al. (2023)	2023	Austria	Quantitative	NA	296	Instructor/Instructional strategies	Behavioral engagement	Dropout rate	Game	Fully Online
7	Kahn et al. (2017)	2017	UK	Qualitative	Graduate	8	Individual factors	Cognitive engagement	Not reported	Asynchronous online discussions	Online degree program
8	Kuromiya et al. (2022)	2022	Japan	Learning analytics	High school	358	NA	Behavioral engagement	Course completion	Ebook	Fully Online
9	Lan & Hew (2020)	2020	China	Mixed methods	All	693	Various	Multiple	Course completion	Active learning, course resources, instructor attributes;	MOOCs
10	Ortega-Arranz (2019)	2019	Spain	Learning analytics	ALL	648	Instructor/Instructional strategies	Behavioral engagement	Course completion	gamification	MOOCs
11	Padilla Rodriguez	2020	Latin American	Learning analytics	Adult learners	617	Course characteristics	Behavioral engagement	Retention	Not reported	MOOCs

#	Authors	Years	Country	Research Design	School level	Sample size	Independent Variable	Engagement type	Dependent Variable	Instructional methods/strategies	Learning environment
	et al. (2020)		(Mexico, Columbia)								
12	Rosser-Majors et al. (2022)	2022	US	Quantitative	Undergraduates	20 instructors	Instructor/Instructional strategies	Multiple	Course completion	Various	Fully online
13	She et al. (2025)	2025	China	Quantitative	ALL	466	Course characteristics, individual factors	Overall engagement	Retention	NA	MOOCs
14	Soffer, T., & Cohen, A. (2019).	2019	Israel	Learning analytics	Undergraduates	646	Various	Multiple	Course completion	Not reported	Fully online
15	Umer et al. (2017)	2017	China	Learning analytics	NA	40 courses	Individual factors	Behavioral engagement	Retention	Not reported	MOOCs
16	Vo & Ho (2024)	2024	Vietnam	Quantitative	Undergraduates	351	Individual factors	Multiple	Not reported	Not reported	Online degree program
17	Wang et al. (2019)	2019	China	Quantitative	All	2,071,147	Various	Multiple	Dropout rate	Lecture, discussion	MOOCs
18	Yang & Lin (2023)	2023	Taiwan	Quantitative	ALL	300	Individual factors	Overall engagement	Retention intention	NA	MOOCs
19	You (2022)	2022	China	Quantitative	Undergraduates	186	NA	Multiple	Course completion	Not reported	Fully online

Note: "Multiple engagement" indicates that two or more engagement types were discussed together. "Overall engagement" refers to cases in which no specific type of engagement was indicated

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5895
https://doi.org/10.24059/olj.5895

Managing Overlapping Review Studies in a Second-Order Meta-Analysis: Describing and Exploring Existing Strategies

Carl D. Westine and Chang Tong
University of North Carolina at Charlotte
Ting Sun
University of Utah
Tong Wu
Riverside InsightsFt

Abstract

This study explores different techniques for managing overlaps in second-order systematic reviews to highlight promising practices for open, distance, and digital education (ODDE) research. Second-order reviews, also known as overviews, provide unique evidence to better describe mean effects and assess their variability, but they also introduce new methodological challenges for quality control due to overlapping primary studies. Recent efforts to categorize and evaluate best practices for managing overlaps in other disciplines highlight significant inconsistencies in practice. As the frequency of ODDE systematic reviews and meta-analyses grows, there is a need to produce high-quality second-order studies to ensure policy makers can properly translate evidence into practice. Several leading overlap management methods are examined, including visualization tools such as citation matrices, bipartite citation networks, stacked bar charts, and intersection heatmaps; quantification tools such as corrected covered area (CCA), adjusted CCA and weighted CCA; decision-based and researcher-based approaches for resolving overlap. These approaches, largely developed in healthcare contexts, are illustrated using an existing dataset from a recent overview (Sun et al., 2025). CCA quantifies the degree of overlap among primary studies, while adjusted CCA further accounts for structural missingness arising when particular primary studies could not have been included in certain reviews because of differences in eligibility or timing. Although wCCA incorporates primary-study sample size into the assessment of overlap, it may be difficult to apply when information is unavailable. Existing thresholds for managing overlaps are largely based on practical guidance, lacking empirical evidence supporting specific cut-offs as decision thresholds.

Keywords: second-order meta-analysis, umbrella review, overlap, corrected covered area, citation matrix

Introduction

Research syntheses are a critical form of evidence used to shape research agendas. As this body of literature continues to expand, second-order reviews have emerged as an important approach for integrating findings across multiple systematic reviews. Various names have been used to describe this form of research: reviews of reviews, meta-syntheses, overview studies, umbrella-reviews or second-order meta-analyses (Cooper & Koenka, 2012; Polanin et al., 2017). Second-order reviews help researchers assess related problems with different interventions or examine studies with the same intervention but different outcomes (Becker & Oxman, 2008). While these reviews are not required to be purely quantitative, one form that is, a second-order meta-analysis, provides several distinct benefits like improving the estimation of true mean

Corresponding author: Carl Westine, cwestine@charlotte.edu
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

effects and enabling assessment of the extent to which any true variance across the mean effect sizes exists (Schmidt & Oh, 2013).

Open, Distance, and Digital Education (ODDE) broadly encompasses teaching and learning processes in which educational technologies, digital media, and digital tools are used to deliver educational content and to support communication, interaction, collaboration, assessment, and evaluation (Zawacki-Richter & Jung, 2023). As research in this field expands across educational contexts, populations, delivery modes, and technologies, quantitative synthesis has become increasingly important for building cumulative evidence. For example, in distance education, Shachar (2008) noted that the abundant comparative literature remained inconclusive and often contradictory because studies varied in treatments, settings, measurement instruments, and research methods, and therefore advocated the use of meta-analysis. As meta-analytic evidence accumulates, returning to large bodies of primary studies becomes resource-intensive. Furthermore, within the fast-developing field of ODDE, interventions and contextual use can change quickly with new innovations and evolving uses across a variety of settings contributing to a literature base with imprecise definitions and descriptions of practice as well as contradictory evidence. This makes ODDE a natural context for growing second-order studies.

Second-order meta-analyses offer a more efficient means of synthesizing existing meta-analytic evidence and deriving cumulative conclusions (Young, 2017), as well as examining factors that lead to conflicting findings. Thus far, only a small but growing number of second-order meta-analyses have emerged across specific ODDE research areas (e.g., Sun et al., 2025; Borokhovski et al., 2022; Martin et al., 2022; Tamim et al., 2011; Yang & Xiang, 2024). However, as the volume and quality of systematic reviews and meta-analyses grows (Gnanasanjevi et al., 2025; Zawacki-Richter et al., 2025), more will be produced. Empirical trends documenting the emergence and importance of second-order reviews in other disciplines (Steenbergen-Hu et al., 2016) reinforce the important role that second-order studies will play in dictating practice within ODDE.

Although many second-order methods have been adapted from first-order studies, second-order studies do present unique challenges given the nested structure of the data across multiple levels. One unique challenge is overlapping primary studies, resulting in double counting of effects. In anticipation of a growing demand for second-order studies within ODDE, our efforts aim to describe and demonstrate several existing methods for identifying, examining, and removing overlaps. In doing this we hope to promote the efficient creation of high-quality second-order reviews and start a dialog to help researchers construct best practices within the ODDE context to ensure these comprehensive syntheses are most effective.

The Challenge of Overlaps

Given the nature of review studies as a form of synthesis, one challenge of any review study is addressing the influence of overlapping observations (Borenstein et al., 2009). Overlap occurs when the same participant effects appear in more than one included study. In a second order study, this can occur in a variety of ways and across more than one level (Polanin et al., 2017; Schmidt & Oh, 2013). This underscores the importance for researchers to clearly describe any identified overlaps and manage them in a transparent way.

Overlaps produce dependency, which in a statistical context can inflate Type 1 error and produce biased estimates. Dependency also manifests in several other related ways. For example,

some studies may report multiple outcomes on the same participants. Additionally, specialization can result in several similar studies being published by the same researcher. These other types of dependencies do not reflect the same double-counting of studies that occurs with overlap, and as a result can often be accounted for with statistical modeling. Various analysis techniques have been proposed and developed to model and account for dependency (e.g., robust variance meta-regression, and multivariate meta-regression, and multi-level modeling). These options offer valuable contributions to account for dependency when sample sizes are sufficiently large, but must be integrated with strategies to remove overlap stemming from double-counting.

Moreover, the amount of overlap that is present in a review study can be viewed as an evaluative criterion that has different levels of value depending on the study aims and researcher's epistemology. For example, as noted above, in a meta-analysis, where quantitative effects are weighted by study size, overlaps indicate double counting, a violation of statistical independence that is interpreted as undesirable. However, in qualitative reviews, overlaps may not be viewed as problematic because findings are not weighted by the study participants. Instead, overlaps may constitute an interesting study opportunity to characterize and explore.

We examine overlaps as they pertain to second-order meta-analyses where statistical independence is viewed as important and therefore minimizing the influence of overlap is paramount. We highlight and examine recent methodological advances from other fields to help streamline the process for ODDE researchers. By describing and demonstrating methods from this literature-base and reflecting on the process of our own experiences conducting second-order reviews we seek to provide greater transparency to both the practice and the root causes of the challenges associated with conceptualizing and resolving overlaps in second-order reviews within ODDE.

Overlaps in Second-Order Studies

As noted above, overlaps in second-order studies can occur in several ways. Just like in a meta-analysis, study participants may be the same in more than one primary study. This can occur in medical fields, where a specific diagnosis may be a driver of study inclusion and therefore increase the likelihood of the same participants appearing in multiple primary studies. Additionally, these patients may seek treatment for long periods of time and therefore be included in multiple research studies. Similarly, in education, overlaps might occur in the field of Special Education, where the presence of a specific (or uncommon) disability may induce greater likelihood of study inclusion. In ODDE, this type of overlap is less plausible. This is because in education the leveled and time-bound learning structure limits the opportunity and value of repeated interventions involving the same outcomes from the same participants. While cases certainly exist, they tend to be the exception. For example, researchers could conduct and publish multiple related (e.g., stacked treatment) studies that involve a variety of comparisons involving a common set of participants. However, in our own experience, these types of studies are uncommon.

More commonly, overlaps in second-order studies occur because there are primary studies that are included across multiple meta-analyses, and this should be expected in ODDE given the need for updating meta-analyses in response to innovations. This type of overlap can be wholly encompassing of study (such as what can occur longitudinally when one meta-analysis study successively builds off another) it can also vary depending on the inclusion criteria and search

strategies (e.g., one meta-analysis might include a few primary studies that are also included in only meta-analysis).

Theoretically, a single primary study might appear in one or up to all included meta-analyses. In a second-order meta-analysis because the primary studies are not the unit of analysis, they cannot easily be individually removed through deduplication. To reduce the amount of overlap to none or at least to an acceptably minimal level, whole meta-analyses must be discarded, potentially limiting the coverage of the second-order study.

Over the years, several strategies for managing overlapping primary studies have been employed. However, recent efforts to categorize practices for managing overlaps in other fields highlight significant reporting inconsistencies (Lunny et al., 2020; Pieper et al., 2014) and inaccuracies in the labeling of pairwise versus total dependencies (Bracchiglione et al., 2023). Even when overlaps are identified, techniques for resolving dependencies vary (Gutierrez-Arias et al., 2024), and remain largely untested. Additionally, nearly all strategies are based on reviews within the healthcare fields where managing overlaps necessarily differs given the size of reviews and the vast number that adhere to quality standards (e.g., Cochrane). The limited empirical evidence comparing the different strategies leaves researchers with unclear guidance (Lunny et al., 2021)

Different conceptualizations of overlaps can theoretically impact decisions on which studies to remove. Researchers must carefully examine and then communicate the nature of overlaps through visualization and calculation techniques to inform and defend any decisions on how best to resolve any issues with overlaps (Gutierrez-Arias et al., 2024).

To better position ODDE researchers to meet these needs, in the next section we will examine existing and emerging overlap management (quantification, visualization, and removal) approaches for second-order systematic reviews. We will also draw on our own experiences conducting second-order studies in ODDE to illustrate valuable process and provide methodological guidance to support the creation of second-order reviews in ODDE research.

Existing Examples of Managing Overlap from ODDE

Second-order meta-analyses in ODDE constitute an emerging research trend with only a few recent examples. We have identified only a few examples within the ODDE literature. As a result, limited guidance is available on how to handle overlaps in second order reviews within the field. To address this, we examine several approaches found in healthcare to the ODDE context. In this study, we describe and illustrate three categories of overlap management methods using Sun et al. (2025) as a running case. In the Sun et al. study, 15 relevant meta-analyses on multimedia were identified, but three studies did not provide enough details about the included primary studies to code for overlaps. Overlaps among the remaining 12 articles were examined, and one article was removed due to significant overlaps. By illustrating several leading methods for managing overlaps we aim to reflect on existing practices from healthcare and address the applicability of certain methods for use in ODDE. In doing so, we aim to help ODDE researchers more efficiently, effectively, and transparently manage overlaps in their second-order meta-analyses.

Table 1 highlights several examples and illustrates the various strategies for how overlap has been handled in existing second-order studies. Across these studies, approaches to primary-study

overlap varied considerably. Some studies reduced overlap through explicit selection rules, such as the 25% criterion (Demir et al., 2024; Tamim et al., 2011), whereas others combined overlap reduction with validation or robustness analyses to examine whether the overall findings changed after overlap was reduced (Yang & Xiang, 2024; Young, 2017). Other studies retained overlapping evidence in the main synthesis but conducted a de-duplicated validation analysis, chose not to formally quantify overlap based on prior validation evidence, or explicitly decided not to apply an overlap-reduction criterion (Bernard et al., 2014; Borokhovski et al., 2022; Hew et al., 2021). Ünal et al., (2026) quantified and visualized overlap by using GROOVE and CCA. Meanwhile, some studies considered statistical independence or acknowledged dependence as a limitation without formally quantifying primary-study overlap across meta-analyses (Sailer et al., 2024; Yi et al., 2025). Overall, these studies suggest that approaches to identifying, quantifying, and managing overlap remain highly study-specific, with no consistent strategy yet established across ODDE research.

Table 1

Second-Order Meta-Analytic Studies Identified in ODDE field.

Study	Context	Handling of overlap
Tamim et al. (2011)	Educational technology and learning	Assessed primary-study overlap and applied a ≤25% overlap criterion to select meta-analyses for inclusion.
Bernard et al. (2014)	Technology integration in higher education	Acknowledged primary-study overlap and noted Tamim et al.'s (2011) ≤25% overlap criterion but chose not to apply it, considering dependence from overlapping studies less severe in second-order meta-analysis.
Young (2017)	Technology-enhanced mathematics instruction	Assessed primary-study overlap and removed highly overlapping meta-analyses (>50%), reducing overall overlap from 20.8% to 13.8% and below the recommended 25% threshold (Cooper & Koenka's; 2012); a fully de-duplicated validation analysis produced a nearly identical effect (g = 0.38 vs. 0.39).
Hew et al. (2021)	Flipped classroom	Retained overlapping meta-analyses in the main second-order synthesis and conducted a de-duplicated primary-study-level analysis as validation.
Borokhovski et al. (2022)	Technology application in teaching and learning	Did not formally quantify primary-study overlap; excluded obvious duplicate/subset meta-analyses and relied on Tamim et al.'s (2011) validation indicating little added precision from further overlap assessment.
Demir et al. (2024)	Technology-based mathematics instruction	Retained overlap in the main synthesis and conducted a de-duplicated primary-study validation; the effect sizes were similar (g = 0.45 vs. 0.41), suggesting little impact of overlap on the overall estimate.
Sailer et al. (2024)	Technology-enhanced learning in higher education	Acknowledged overlap as a limitation, but did not formally quantify or adjust for primary-study overlap.

Study	Context	Handling of overlap
Yang & Xiang (2024)	Mobile learning and academic achievement	Assessed primary-study overlap and recursively removed the most overlapping meta-analyses to achieve <50% and <25% overlap thresholds (Sipe & Curlette, 1997; Tamim et al., 2011; Wilson & Lipsey, 2001); effect estimates remained similar across the full sample and reduced-overlap analyses (g = .652, .673, and .580).
Yi et al. (2025)	Technology-assisted foreign language learning	Considered data independence within individual meta-analyses but did not formally assess primary-study overlap across meta-analyses.
Ünal et al. (2026)	Artificial intelligence applications in education	Used GROOVE and CCA to quantify overlap; when overlap exceeded 10% (Bracchiglione et al., 2022), the narrower or less current meta-analysis was excluded.

Strategies for Managing (Quantifying, Visualizing, and Resolving) Overlaps

The potential problem of having overlaps typically arises early in the research process when screening for eligibility but is ultimately examined more fully as the final sample of studies to be used becomes more crystalized. To begin to address the overlaps, specific data preparation actions need to be taken to enable quantifying and visualizing the overlaps, before ultimately resolving the overlaps by removing select studies. As a first step, researchers must understand the nature of any overlaps that exist within their study. This requires deciding on an appropriate method for quantifying overlaps as well as describing through visualization, any overlaps that do exist.

Step 1: Understanding the Overlaps

Quantifying

Various strategies for quantifying overlaps exist. A recent systematic review of healthcare literature highlighted several prevailing approaches (Gutierrez-Arias et al., 2024). At a most basic level, researchers will sometimes calculate the percentage of overlapping primary studies among the systematic reviews. This can be reported for pairs of studies or for the total set of studies. Alternatively, a range of all pairwise overlaps may be reported among the studies included. Although conceptually easy to compute, one problem with simply reporting percentages is the amount of double-counting that can occur. For this reason, in the past decade, researchers have largely moved beyond simple percentages.

Calculating the Corrected Coverage Area. In 2014, Pieper et al. characterized the quantity of overlap using a corrected coverage area (CCA) to better account for primary studies that overlap with different sets of systematic reviews. The CCA removes the double counting of overlapping primary studies by considering each primary study once. While a bit more laborious to compute, this method is particularly relevant for comparing the amount of statistical independence in two different second-order studies. It provides a more consistent characterization of the aggregate amount of overlap that exists to allow for better comparison of overlaps between several second-order studies of different sizes. Recently, Ying et al. (2025)

presented a weighted CCA (wCCA), which expands the CCA calculation by weighting each systematic review by the sample sizes of the primary studies in each review. This gives the second-order researcher even more standardization in the amount of overlap and further improves comparability.

Missing Data. An important concern with quantifying overlaps is missing data. Missing data can occur in a variety of ways. For example, if a meta-analysis does not provide complete details on each of the primary studies included, then the extent of overlaps cannot be fully evaluated. This can be a major limitation when trying to assess statistical independence. For wCCA, the problem is compounded, as sample sizes for each individual primary study need to be known to properly weight each study.

Another type of missing data is structural missing data, for example due to the search ranges for systematic reviews not being the same. Primary studies in one systematic review may not be included in the timeframe of another study. For this reason, an adjusted CCA which removes these out-of-range studies from the calculation can be computed (Bougioukas et al, 2021).

Recent empirical research has pointed out the varying quality of systematic reviews in ODDE. According to Zawacki-Richter et al. (2025), one-third of articles in ODDE do not strictly follow PRISMA. Based on our own experience extracting information from ODDE studies for second-order reviews there is plenty of room for improvement in reporting the necessary information in primary studies. Complete information for assessing overlaps is not always provided, it can be inefficient to track down, or very difficult to extract consistently for each review due to a lack of enforcement of standardized reporting. With greater standardization in reporting of both primary studies and systematic reviews, the ODDE field will greatly enhance the assessment of the impact of overlaps in second-order studies.

Visualizing

Visualizing the overlaps between studies goes together with quantifying the amount of overlap. Multiple studies have introduced visual techniques to help describe overlaps that exist. Pieper et al. (2014) popularized the process of visualizing overlaps using a matrix. Gutierrez-Arias et al. (2024) found this was the most presented visualization. However, other techniques have been demonstrated as well. For example, Bougioukas et al. (2021) presented several other examples using traditional visualizations like Venn and Euler diagrams and network plots. Later, Bougioukas et al (2023) generated a tool to facilitate the efficient creation of a citation matrix and CCA heat maps, and Bracchiglione et al (2023) combined several visualization techniques to innovate how overlaps are presented using bar graphs.

Instructional Application Using Examples from ODDE - Resolving Overlaps

In this section we utilize data from a previous study by Sun et al. (2025) to demonstrate methods for understanding and describing the overlaps that are present in a study. We illustrate the process and reflect on several approaches for quantifying and visualizing the data.

The research team must first extract the necessary data from the articles. This means compiling a list of every primary study used in each systematic review to be included in the second-order systematic review. If there are overlaps, the same citations will be on the list two or

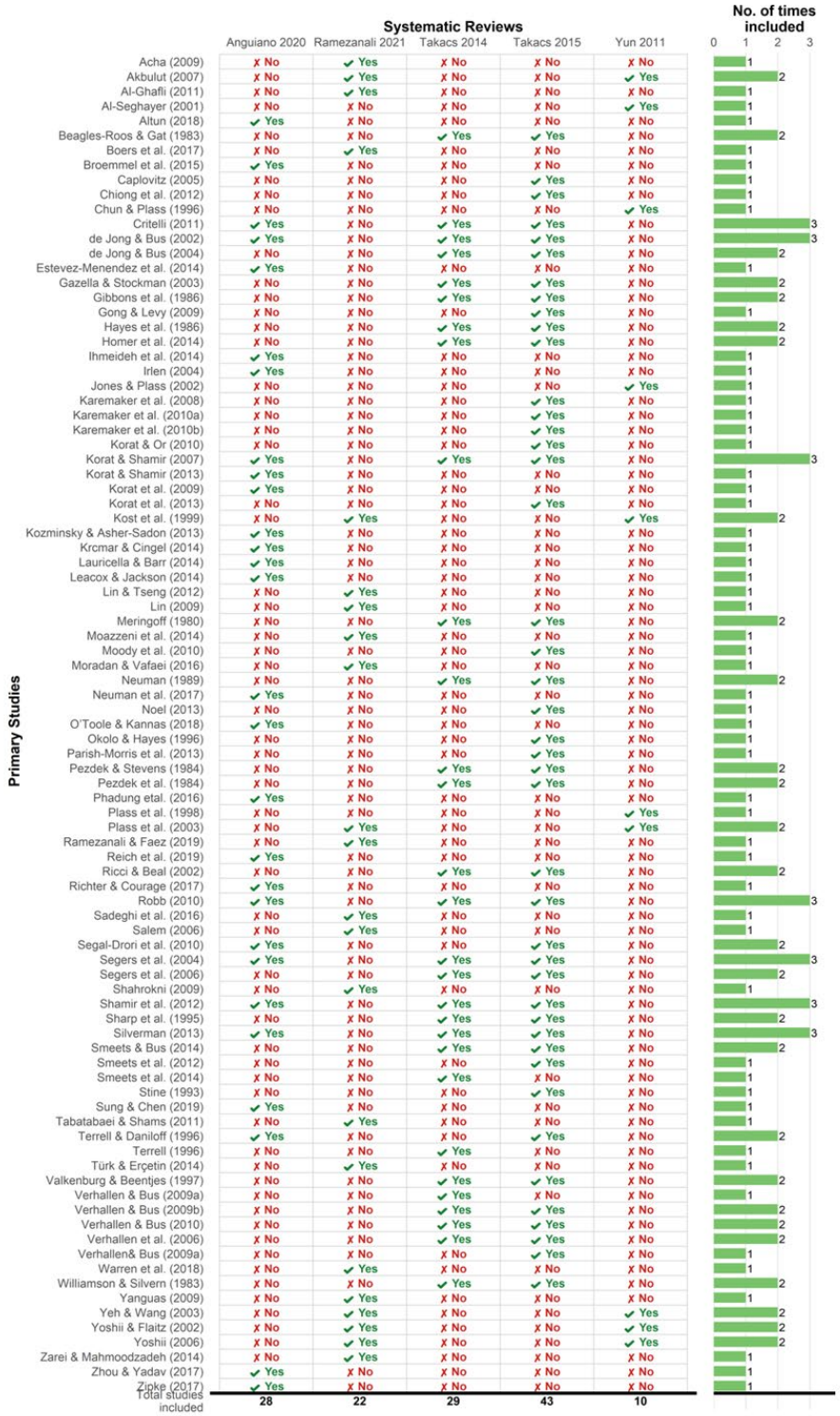
more times. Ideally all systematic reviews will include a full list of articles. If not, the researcher will need to proceed with only the subset of available articles, noting the limitation. For Sun et al. (2025), three studies did not have usable primary study lists, and needed to be removed from the overlaps analysis.

Citation Matrix Construction

The Citation Matrix is the foundation for calculating the CCA. It is a binary study-by-review table that documents the inclusion of primary studies across multiple systematic reviews. The leftmost column lists unique primary studies, and the subsequent columns represent individual systematic reviews and cells are coded to indicate inclusion or non-inclusion (See Figure 1). To create a citation matrix, researchers need to extract all primary studies from each review, identify unique studies after removing duplicates, and code their presence (1) or absence (0) across reviews. This structure enables identification and quantification of citation overlap across reviews.

Figure 1

Citation Matrix



Corrected Coverage Area

The Corrected Coverage Area (CCA) is a quantitative measure used to assess the degree of overlap among primary studies included in the systematic reviews (Pieper et al., 2014; Bougioukas et al., 2021). The CCA can be calculated using the following formula:

$$CCA = \frac{(N - r)}{(r * c - r)}$$

where *N* represents the total number of primary study occurrences across the systematic reviews (i.e., the sum of primary studies included within each review, including duplicates), *r* represents the number of unique primary studies included, and c indicates the number of systematic reviews. Pieper et al (2014) also provide guidance on interpreting the CCA percentages: 0-5% - slight, 6-10% - moderate, 11-15% - high, and >15% - very high. In Figure 1, *N* is identified as the sum of the values in the rightmost column (No. of times included, which has been augmented with visual bars for clarity), *r* is the number of individual study rows, and *c* is the number of study columns. For illustration purposes, Figure 1 only includes five of the 12 studies included in the overlap analysis. The CCA value can also be calculated using the ccaR() function in R once the full citation matrix has been prepared. Below, in Table 2, is the R output using the full set of data from Sun et al. (2025):

Table 2

R Output for CCA Calculation

reviews	N	r	c	CCA_Proportion	CCA_Percentage
12	326	285	12	0.01307815	1.3

This output indicates 326 total study occurrences (*N*) and 285 unique primary studies (r) across 12 reviews (c). The calculated CCA proportion is 0.013 (1.3%), indicating a very low level of overlap among reviews.

Adjusted Corrected Coverage Area

The adjusted CCA accounts for structural missingness, which occurs when primary studies were published after a given systematic review and therefore could not have been included (Bougioukas et al., 2021). In the overlap matrix, these structurally unavailable studies were coded as NA rather than as absent (i.e., coded as 0). This approach prevents underestimation of overlap due to temporal sequencing and provides a more accurate representation of redundancy across reviews (Bougioukas et al., 2021). The adjusted CCA is calculated as:

$$Adjusted\ CCA = \frac{(N - r)}{(r * c - r - X)}$$

where *N*, *r*, and *c* are the same as in the CCA calculation, and *X* is the total number of structural missingness in the citation matrix. The adjusted CCA can also be calculated in R using the ccaR() function after updating the citation matrix to account for structural missingness. Below, in Table 3, is the R output for the full set of data from Sun et al. (2025):

Table 3

R Output for Adjusted CCA Calculation

reviews	N	r	c	Structural_missingness	CCA_Proportion	CCA_Percentage
12	326	285	12	674	0.01665989	1.7

The ccaR() output for the adjusted CCA suggests that there are 674 structurally missing cells in the citation matrix. After accounting for the structural missingness, the adjusted CCA proportion is 0.017 (1.7%), which is slightly higher than unadjusted CCA. The results indicate that overlap across the reviews remains low even after adjusting for sparsity in the citation matrix.

Weighted Corrected Coverage Areas

The wCCA is a newly proposed extension of the traditional CCA, which incorporates sample-size–based weights (Ying et al., 2025). The wCCA can be calculated as:

$$wCCA = \frac{(N_w - r_w)}{(r_w * c - r_w)}$$

where N_w is the weighted sample size which consists of the sum of the square roots of the sample sizes of all primary studies, aggregated across all SRs and r_w represents the sum of the square roots of sample sizes for unique primary studies, where each primary study is counted only once. To manually compute the wCCA, the citation matrix needs to be appended to include an additional column which includes the sample size for each primary study listed in the first column. Compared with the CCA, wCCA provides a more precise assessment of the overlap by accounting for differential contributions across reviews. A similar extension to account for structural missingness could also be applied to create an adjusted wCCA.

Some practical challenges emerged when calculating the wCCA using ODDE articles from Sun et al. (2025). As Zawacki-Richter et al. (2025) pointed out, many systematic reviews lack transparent reporting, which can hinder the retrieval of key primary characteristics, including sample size. Pollock et al. (2017) also found similar data could be presented differently across reviews. Both were evident in our experience.

To illustrate the wCCA, we consider the same subset of five systematic reviews as above. However, only three systematic reviews reported the sample size for each primary study. Moreover, even when these sample sizes were reported, the identification of multiple related within-subjects comparisons required decision-making and interpretation of reported sample sizes to address the double-counting participants across conditions, which also affects statistical independence. Among the 87 primary studies included, this occurred in five articles. In addition, one primary study could not be retrieved.

To explore how different decisions might affect the wCCA, we applied three calculation strategies although differences in study focus and design could introduce bias. First, we used the sample sizes reported in the primary studies. Second, we relied on the sample sizes reported in the systematic reviews and supplemented missing values with those from primary studies. Third, we again used the review-reported sample sizes, but for missing cases we estimated values based

on the total reported sample and the average of the unknown portion. The resulting wCCA values for just these three studies were 0.1289, 0.1140, and 0.1146, respectively.

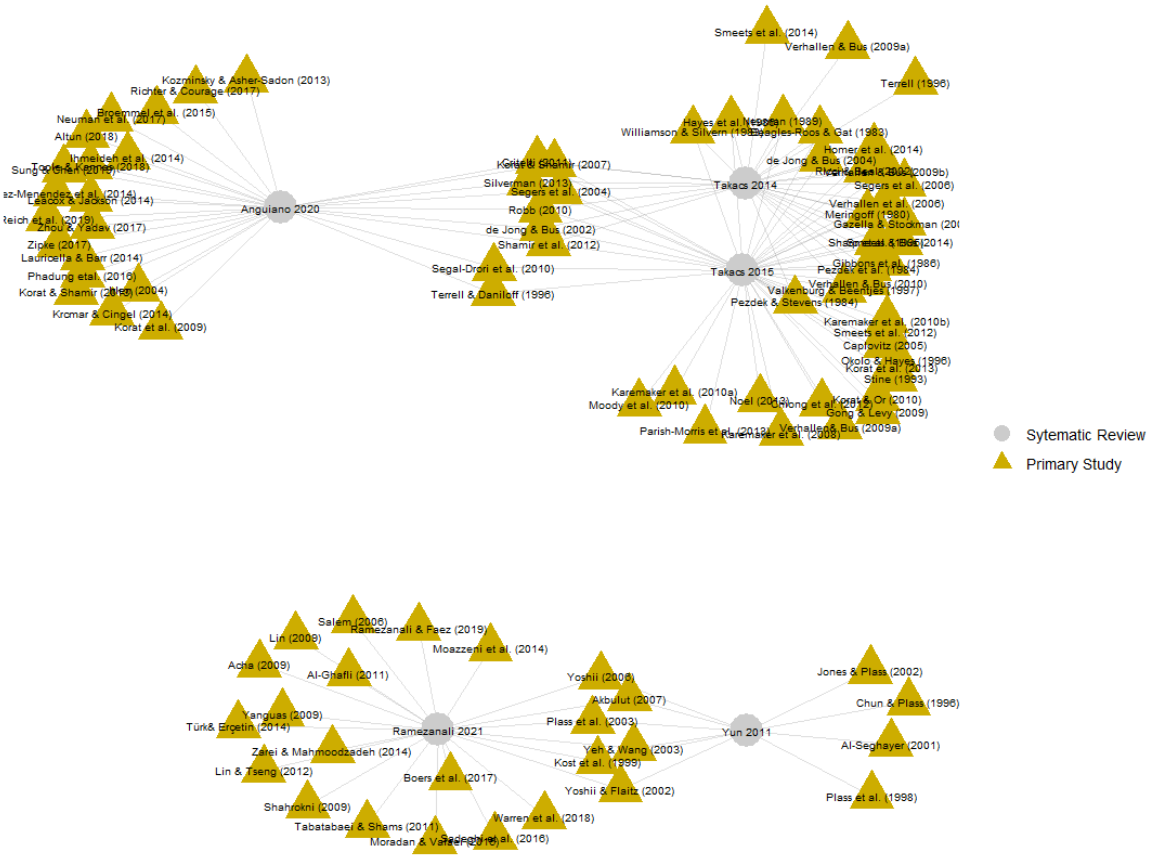
Supplemental Visualization Techniques to Gain Additional Clarity

While CCA-type values provide an overall summary measure, they do not indicate which specific reviews overlap most or where the overlap occurs. This can be gleaned from the citation matrix with careful inspection, though additional visualization strategies could be informative. Several visualization methods and illustrative examples were reviewed by Bougioukas et al. (2021). Figures and tables can be useful for depicting overlaps in terms of density and spectrum; however, most visualization techniques become difficult to interpret when the number of studies exceeds approximately five. For example, Venn diagrams, Euler diagrams, UpSet diagrams, and network-based approaches such as bipartite graphs have been used to illustrate study overlap. Although these methods can serve as helpful intermediate tools for exploring patterns of overlap, they are less suitable for reporting purposes because the visualizations often become overly crowded and difficult to interpret.

This limitation is illustrated in Figure 2, which presents a bipartite "cloud-like" citation network based on five first-order meta-analyses selected from Sun et al. (2025) as a demonstration. In the figure, grey nodes represent first-order meta-analyses, and yellow nodes represent primary studies. The visualization shows that Takacs (2014) and Takacs (2015) share a substantial number of primary studies, which is unsurprising given that they were authored by the same research team. These two meta-analyses also share numerous primary studies with Anguiano (2020). A separate cluster on the right side of the figure indicates that Ramezanili (2021) and Yun (2011) share a smaller number of primary studies. However, even with only five meta-analyses included, the primary study nodes appear densely clustered, limiting the clarity and interpretability of the visualization.

Figure 2

Bipartite citation network



Stacked Bar Chart

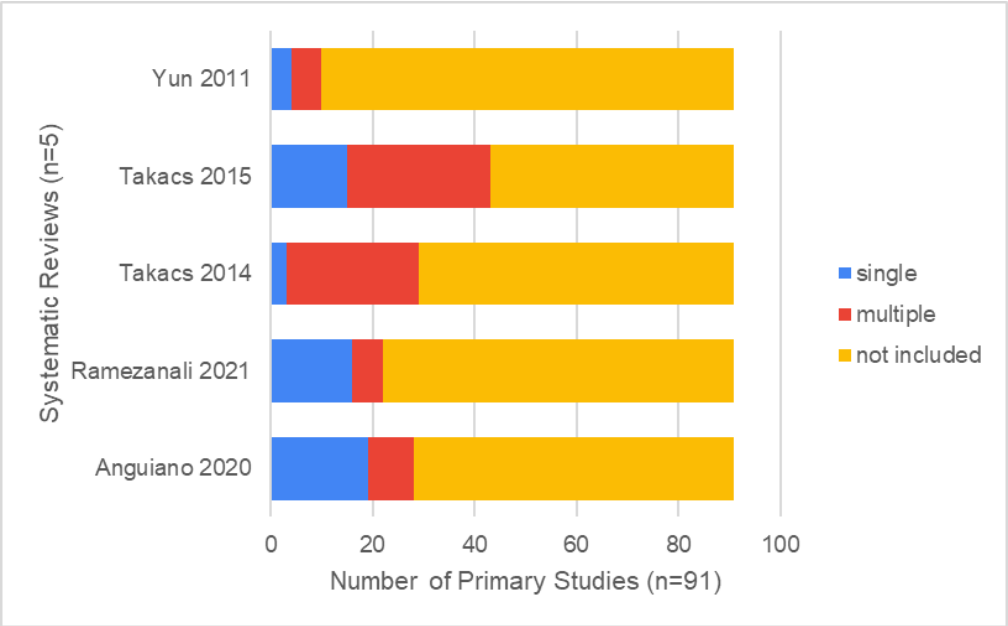
It is not uncommon for second-order studies to include many first-order meta-analyses, each of which may synthesize multiple primary studies (Lunny et al., 2020). Contrasted with the citation matrix (shown in Figure 1), which summarizes the number of times each primary study appears in the systematic reviews, it is necessary to also convey the percentage and nature of overlaps occurring within each systematic review. In this context, stacked bar plots are a useful visualization method for providing a relatively straightforward, high-level summary of overlap patterns, particularly in terms of the frequency of shared versus unique primary studies. However, bar plots cannot represent the specific structure of overlap (i.e., which primary studies are shared by first-order meta-analyses).

As shown in Figure 3 using the subset of articles from Sun et al. (2025), we labeled each bar segment using three categories, single, multiple, and not included, following Bougioukas et al. (2021). The blue segments (single) indicate primary studies that are exclusively included in each first-order meta-analysis. The red segments (multiple) represent primary studies that are included in the corresponding first-order meta-analysis as well as in one or more other first-order meta-

analyses. The yellow segments (not included) indicate primary studies that are not included in the corresponding first-order meta-analysis.

Figure 3

Stacked Bar Chart of Primary Study Inclusion Status



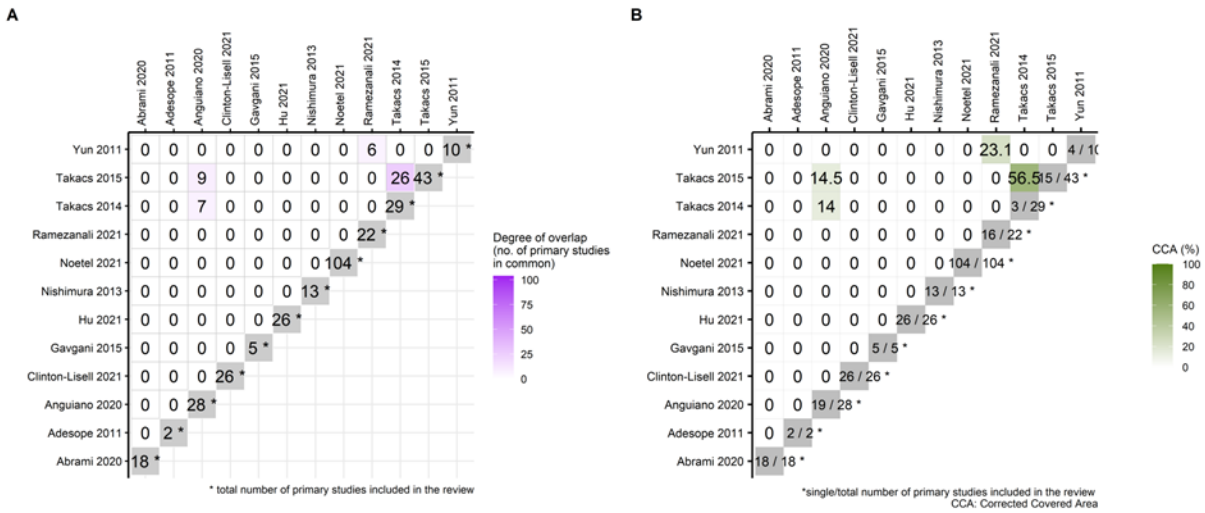
Pairwise Intersection Heatmap

The pairwise intersection heatmap illustrates both the absolute number of overlapping primary studies between each pair of systematic reviews and the corresponding pairwise CCA percentages. These visualizations can be produced using the ccaR() package (Bougioukas et al., 2023) or with an Excel based GROOVE tool from Bracchiglione et al (2022). The map provides not only an overall summary of overlap but also a detailed pairwise comparison.

In Figure 4, Panel A displays the number of shared primary studies between each pair of systematic reviews and Panel B presents the percentage CCA for each pair of reviews. The most notable overlap occurs between Takacs (2015) and Takacs (2014), which share 26 primary studies, corresponding to a high pairwise CCA of 56.5% in Panel B. Additional overlap is observed between Yun (2011) and Ramezanali (2021) (6 shared studies; 23.1%), Takacs (2015) and Anguiano (2020) (9 shared studies; 14.5%), and Takacs (2014) and Anguiano (2020) (7 shared studies; 14.0%). However, most cells in the matrix are zero, indicating no shared primary studies between most review pairs and reinforcing the overall finding of small redundancy.

Figure 4

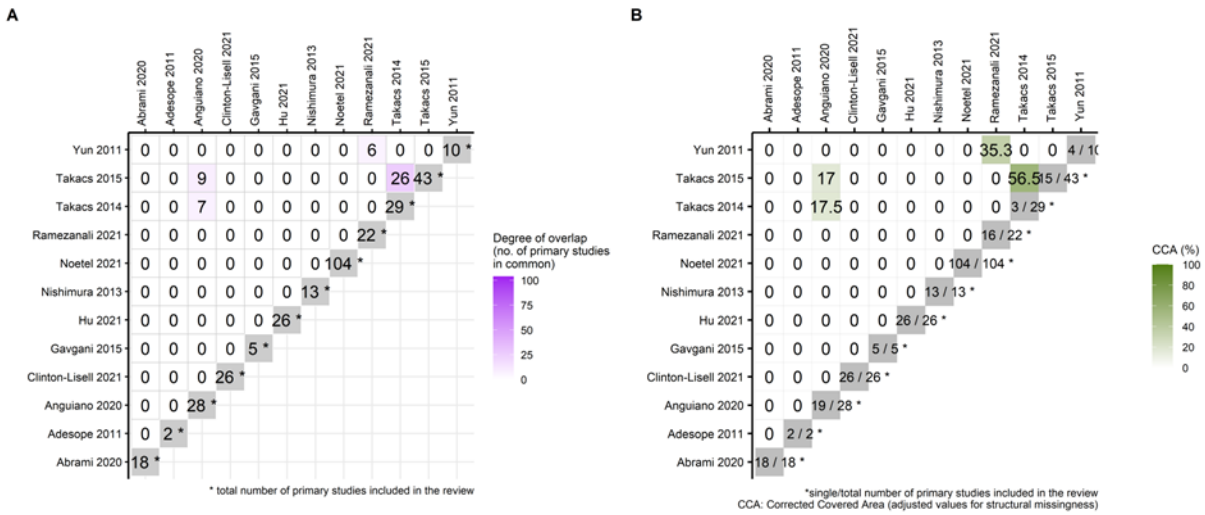
Pairwise Intersection Heatmap



In the adjusted pairwise intersection heatmap (Figure 5), Panel A remains unchanged, as it presents the absolute number of overlapping primary studies between each pair of systematic reviews. Because the percentage CCA values are recalculated to account for structural missingness, Panel B changes after adjustment. The adjusted values are slightly higher and provide a more accurate estimate of overlap.

Figure 5

Adjusted Pairwise Intersection Heatmap



Bougioukas et al. (2021) provide a comprehensive overview of several available visualization methods, including detailed descriptions as well as a discussion of their respective advantages and limitations. Readers interested in a more in-depth examination of overlap visualization techniques are referred to in that paper.

Step 2: Resolving the Overlaps

Once a clear understanding of the extent of overlaps is developed, a strategy for resolving/removing overlaps needs to be deployed to minimize the impact of the validity threat. While there have been recent strides in developing strategies and tools to measure and visualize overlaps, there is far less understanding of how to resolve overlaps. Recent research suggests that the lack of attention to overlaps is problematic and persistent. It is often reported that at least half of published overview studies fail to address overlap, and many other studies only acknowledge it as a threat or quantify it (Gutierrez-Arias et al., 2024; Lunny et al. 2020; Pieper et al., 2014b; Polanin et al., 2017; Pollock et al., 2017;). In this section we describe several prevailing methods for resolving overlaps.

Strategies for Resolution

In a recent review of overviews in healthcare three dominant approaches for resolving overlaps prevailed: computing a new meta-analysis, following a (typically quality-based) algorithm, or author-derived (Gutierrez-Arias et al., 2024). The first of these is extremely inefficient, requiring additional data collection and coding to repeat and combine systematic reviews. Depending on how extensive the revised meta-analysis is, it also could answer a different research question, one that primarily focuses on the summary effect and is therefore not

the same as performing a second-order systematic review. Therefore, we explore the other two options in more depth.

Decision-Based Rules and Algorithms. Various forms of decision algorithms have emerged over time, drawing on early work of assessing statistical independence violations from systematic reviews and meta-analyses. Many early efforts focused on readily available quantitative measures and relied on just one or two criteria for judgement, generally involving a percentage overlap calculation. For example, Sipe and Curlette (1997) used decision rules based on the size of the overlapping studies as well as recency, if sufficient overlap existed, to remove a study (and therefore the overlap). Wilson and Lipsey (2001) proposed a recursive elimination process that removed the largest overlapping studies one-at-a-time until a sufficiently small amount of overlap remained.

However, several authors have also pointed to the value of qualitative information, such as quality indicators, particularly if a systematic review is not exclusively quantitative in focus. The Jadad algorithm (Jadad et al., 1997) was designed to help researchers perform systematic reviews to remove discordant pairs of related research, by considering several qualitative factors including relevance, quality, and selection criteria. Cooper and Koenka (2012) also highlighted other qualitative indicators that could be used to help prioritize removal of studies (e.g., completeness, recency, size, rigor, and source.) Later, Hennesey and Johhson (2020) proposed a decision-tree that incorporates the use of CCA to instigate additional investigation decisions about removal of overlaps.

A few important decisions need to be made when defining a recursive algorithm. The first action is to identify the appropriate comparison source. A researcher could compare the extent that each individual article contains primary studies that overlap with the aggregate field of systematic reviews in the second-order study or individually consider the overlap for all specific pairs of systematic reviews. The next action pertains to the chosen comparison, a specific measure (e.g., percentage or CCA) or measures (e.g., percentage, size, quality, recency) needs to be decided. For comparisons involving just one measure, the decision is determinant. When multiple measures are considered, a framework for valuing each criteria needs to be established. This is often done with a decision-tree (see Jadad et al [1997] or Hennesey & Johnson [2020] for example) or by weighting. Using the chosen measure, an individual study is removed from the analysis, and the process is repeated (e.g., the study with the largest percentage of primary studies overlapping the aggregate field, or one of the two studies with the largest pairwise overlap is removed.) Finally, the last action is to define an appropriate cut-off for the recursive process to end. For example, using percentages, Wilson and Lipsey (2001) suggested a 25% overlap. Newer measures (e.g., using CCA or Adjusted CCA thresholds) could also be used.

Researcher-Based Decisions. Despite all the advances to achieve systematic decision-based algorithms to guide the removal of overlaps, the fact remains that there is a lack of consensus about the proper technique(s) to employ to address overlaps to achieve acceptable levels of statistical independence. Furthermore, there is limited empirical guidance for researchers on how to navigate removal of overlaps efficiently. Consequently, many researchers still present their own methods for resolving overlaps, essentially imposing their own value-judgements into the process (Strech & Tilburt, 2008). While it seems reasonable that researchers would be informed of important considerations within the literature they are reviewing, inconsistency of methods ultimately only limits comparability.

Even though some decision-based algorithms exist, there is still plenty of opportunity for a researcher to insert their own decision into the process, and many do. Some algorithms rely only on an established measure or framework of measures, or defined cut-offs. However, given the relative novelty of conducting second-order reviews, guidelines and standards are not well established or known.

Instructional Application Using Examples from ODDE - Understanding Overlaps

Within ODDE we identified three overview studies that employed a combination of both decision-based algorithms and researcher-based decisions. Each is rooted in the Wilson and Lipsey (2001) method noting a suggested cut-off level of 25% (Sun et al., 2025; Martin et al., 2022; Tamim et al., 2011). Martin et al. (2022) describe a sensitivity analysis to compare the final analysis with less restrictive cut-off levels (e.g., 50%). Sun et al. (2025) used a 50% cut-off given a limited sample size.

As illustrated in Figure 3, some studies exceed the 50% cut-off level (i.e., there are more primary studies appearing in multiple papers than in just one). In this case Takacs et al. (2014) has the highest percentage of primary studies overlapping with other reviews, so it is removed. As Tamim et al. (2001) points out, the recursive process may take several iterations. Further exploration of Figure 1 shows Takacs et al. (2014) and Takacs et al. (2015) are highly aligned. The removal of just one of these studies limits the percentage of overlaps in the remaining reviews, and no other removal was required. As a result, all estimations of effects would need to be recalculated using the smaller dataset.

The process described above utilized percentage of overlap, but one could easily conceptualize other rules including the use of pairwise CCAs and Adjusted CCAs; however, Pieper et al (2014) and others (e.g., Lunny [2020]) cautioned against using CCA as an absolute cutoff and instead have advocated for a more qualitative-based decision. Ultimately, further research into the use of CCA-type measures is needed to assess the sensitivity of the measure to different overlap scenarios.

In the Sun et al. (2025) example, the results are the same. Pairwise comparisons would highlight the significant overlap between Takacs (2014) and Takacs (2015), and measures of recency and size would dictate removal of the 2014 piece. Alternatively compared to the field, the most overlap occurs with the 2014 article, and it is therefore first in line for removal. It's easy to hypothesize a scenario where one article overlaps significantly with the field but doesn't have a large pairwise overlap with any other article.

Discussion

Second-order meta-analyses are an important tool for combining first-order meta-analyses to examine various related interventions, contextual factors, and study populations. They are useful for synthesizing and updating evidence as innovations change how interventions are conceptualized and used. Yet, they also present new challenges, like managing overlaps, for researchers to navigate. Standardizing practice is a necessary step for ensuring findings from these studies are useful. In our experience, addressing overlaps in second-order studies is a daunting challenge, with much of the research residing in unfamiliar fields.

For practitioners working in ODDE, several implications for practice and future research can be drawn from this synthesis. The ODDE field evolves quickly with new technologies generating widespread use within a short period of time. The recent example of generative artificial intelligence is a good example that continues to transform education in all contexts: a small finite time frame, but many meta-analyses. Thus, within ODDE many second-order studies are likely to encounter considerable overlap given narrow time frames. Encouraging consistent practices is important for promoting the great value of these second-order studies. Within ODDE, significant challenges of transparency are already apparent (Zawacki-Richter et al, 2025). In Sun et al. (2025), a study of only 15 meta-analyses, 20% of the included meta-analyses did not have adequate data to include in the overlaps assessment.

To adequately address overlaps, researchers need to first thoroughly understand the overlaps through calculation and visualization. The researcher's first goal should be to generate a citation matrix and produce a pairwise heat map and stacked bar chart. With these outputs the researcher can adequately describe the range, frequency, and distribution of pairwise CCA comparisons as well as the overall CCA. However, some important ODDE questions consider impacts across generations (Moore & Kearsley, 2011). For ODDE studies that consider questions spanning different generations (e.g., Martin et al., 2022), use of adjusted CCA in the overlap quantification is an important consideration. However, extending calculations beyond CCA to consider more refined calculations including wCCA appears to be difficult within ODDE, given inconsistencies in reporting of meta-studies. Promoting greater consistency of reporting practices for meta-analyses with an aim to also utilize these meta-analyses for second-order studies will enhance transparency and enable more refined research.

The PRIOR guidelines (Gates, 2022) have sought to align researchers in how to characterize and report on overlaps (Pollock et al., 2019). However, little prescription is offered to manage overlaps beyond "Describe the extent of primary study overlap across the included systematic reviews." (Gates, 2022, data supplement). Given recent advances to efficiently visualize and quantify overlaps, the empirical derivation of best practices presents an emerging research need (Bracchiglione et al., 2026), with empirical comparisons of resolution strategies remain a particularly important avenue for future research. Overall, ODDE researchers should clearly assess and report overlaps and transparently justify the decision to retain the overlaps without adjustment or removal.

The present work represents a methodological commentary, helping to import tools and strategies from other fields in an effort to advance the ODDE field into better practice and encourage more empirical research in this area. Broadly, there has been limited empirical research on quantifying, visualizing, and resolving overlaps (Gutierrez-Arias et al., 2024; Lunny, 2020). Even in healthcare, where much of the research resides, there is a dearth of empirical evidence depicting the impact of strategies used to limit violations of statistical independence in second-order studies. Such a dearth of guidance limits not only the guidance to move forward, but also negatively impacts the validity inferences based upon existing overview studies, as it is unclear the extent that failing to address overlap can have on robustness of reported findings. As Schmidt and Oh (2013) suggest, more extensive simulation work (akin to Tracz, Elmore, and Pohlmann [1992]) is needed to inform decision-making around overlaps in second-order studies. In our judgement, the ODDE context presents an important context for providing empirical research to help justify and dictate best practices. . As the number of ODDE second-order reviews grows, systematic reviews should be conducted to describe and examine how these

strategies are applied in wider practice. Additionally, future research should use simulation studies to empirically evaluate the performance of different overlap management strategies across varying overlap scenarios.

Limitations

The illustrative nature of this study introduces several limitations. Our application draws on a single ODDE overview in which the overall level of overlap is relatively low, limiting the opportunity to examine different approaches under severe overlap conditions. The application of wCCA also revealed a practical constraint, as inconsistent reporting of primary-study sample sizes made the measure difficult to implement fully. This paper is intended to illustrate and discuss available approaches rather than to systematically and empirically compare their performance, which is a necessary next step for formulating best-practices. Thus, the running example is intended to be demonstrative and should not be interpreted as evidence that one resolution strategy or decision rule is better than another.

References

Anguiano, C. J. (2020). *Effects of multimedia-enhanced storybooks on young childrens' vocabulary and comprehension knowledge: A meta-analysis*. (Publication No. 27955486) [Doctoral dissertation, Washington State University]. *ProQuest Dissertations & Theses Global*.

Becker, L. A., & Oxman, A. D. (2008). Overviews of reviews. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions: Cochrane book series.* John Wiley。 https://doi.org/10.1002/9780470712184.ch22

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* John Wiley & Sons. https://doi.org/10.1002/9780470743386

Borokhovski, E., Tamim, R., & Pickup, D. (2022). *Technology application in teaching and learning: Second-order review of meta-analyses.* Commonwealth of Learning. https://doi.org/10.56059/11599/4069

Bougioukas, K. I., Diakonidis, T., Mavromanoli, A. C., Haidich, A.-B. (2023). ccaR: a package for assessing primary study overlap across systematic reviews in overviews. *Research Synthesis Methods, 14*, 443-454. https://doi.org/10.1002/jrsm.1610

Bougioukas, K. I., Vounzoulaki, E., Mantsiou, C. D., Savvides, E. D., Karakosta, C., Diakonidis, T., Tsapas, A., & Haidich, A.-B. (2021). Methods for depicting overlap in overviews of systematic reviews: An introduction to static tabular and graphical displays. *Journal of Clinical Epidemiology*, *132*, 34–45. https://doi.org/10.1016/j.jclinepi.2020.12.004

Bracchiglione J., Meza, N., Bangdiwala, S. I., Niño de Guzmán, E., Urrútia, G., Bonfill, X., & Madrid, E. (2022). Graphical Representation of Overlap for OVErviews: GROOVE tool. *Research Synthesis Methods,13*(3): 381–388. https://doi.org/10.1002/jrsm.1557

Bracchiglione, J., Meza, N., Pérez-Carrasco, I., Vergara-Merino, L., Madrid, E., Urrútia, G., & Bonfill Cosp, X. (2023). A methodological review finds mismatch between overall and pairwise overlap analysis in a sample of overviews. *Journal of Clinical Epidemiology*, *159*, 31–39. https://doi.org/10.1016/j.jclinepi.2023.05.006

Bracchiglione J., Meza, N., Pieper, D., Lunny, C., Vargas-Peirano, M. Vicuña, J., Briceño, F., Garnham Parra, R., Pérez Carrasco, I., Urrútia, G., Bonfill, X., & Madrid, E. (2026). Impact of matrix-construction assumptions on quantitative overlap assessment in overviews: A meta-research study. *Research Synthesis Methods, 17*, 348-364. https://doi.org/10.1017/rsm.2025.10056

Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist, 67*(6), 446–462. https://doi.org/10.1037/a0027119

Gates, M., Gates, A., Pieper, D., Fernandes, R. M., Tricco, A. C., Moher, D., Brennan, S. E., Li, T., Pollock, M., Lunny, C., Sepúlveda, D., McKenzie, J. E., Scott, S. D., Robinson, K. A., Matthias, K., Bougioukas, K. I., Fusar-Poli, P., Whiting, P., Moss, S. J., & Hartling, L. (2022). Reporting guideline for overviews of reviews of healthcare interventions:

Development of the PRIOR statement. *BMJ, 378*, e070849. https://doi.org/10.1136/bmj-2022-070849

Gnanasanjevi, G., Balasubramaniam, P., Sriram, N., Lakshmi, S., Gangai Selvi, R., & Suresh, J. (2025). Impact of open and distance learning on agricultural education: A systematic review and bibliometric analysis. *Plant Science Today, 12*(2). https://doi.org/10.14719/pst.7505

Gutierrez-Arias, R., Pieper, D., Lunny, C., Torres-Castro, R., Aguilera-Eguıa, R., Oliveros, M.-J., & Seron, P. (2024). Only half of the authors of overviews of exercise-related interventions use some strategy to manage overlapping primary studies–a metaresearch study. *Journal of Clinical Epidemiology, 170*, 111328. https://doi.org/10.1016/j.jclinepi.2024.111328

Hennesey, E. A., & Johnson, B. T. (2020). Examining overlap of included studies in meta-reviews: guidance for using the corrected covered area index. *Research Synthesis Methods, 11*,134-145. https://doi.org/10.1002/jrsm.1390

Jadad, A. R., Cook, D. J., & Browman, G. P. (1997). A guide to interpreting discordant systematic reviews. *Canadian Medical Association Journal, 156*(10), 1411-1416.

Lunny, C., Brennan, S. E., McDonald S., & McKenzie, J. E. (2017). Toward a comprehensive evidence map of overview of systematic review methods: paper 1—purpose, eligibility, search and data extraction. *Systematic Reviews*, 231. https://doi.org/10.1186/s13643-017-0617-1

Lunny, C., Brennan, S. E., McDonald S., & McKenzie, J. E. (2018). Toward a comprehensive evidence map of overview of systematic review methods: paper 2—risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Systematic Reviews*, 159. https://doi.org/10.1186/s13643-018-0784-8

Lunny, C., Brennan, S. E., Reid, J., McDonald S., & McKenzie, J. E. (2020). Overviews of reviews incompletely report methods for handling overlapping, discordant, and problematic data. *Journal of Clinical Epidemiology, 118*, 69-85. https://doi.org/10.1016/j.jclinepi.2019.09.025.

Lunny, C., Pieper, D., Thabet, P., & Kanji, S. (2021). Managing overlap of primary study results across systematic reviews: practical considerations for authors of overviews of reviews. *BMC Medical Research Methodology, 21*(1), 140. https://doi.org/10.1186/s12874-021-01269-y

Martin, F. (2025). *Systematic reviews in educational research: A practical guide for research synthesis.* Thousand Oaks, CA: Sage Publications, Inc.

Martin, F., Sun, T., Westine, C. D., Ritzhaupt, A. D. (2022). Examining research on the impact of distance and online learning: A second-order meta-analysis study. *Educational Research Review, 26*, 100438.

Moore, M. & Kearsley, G. (2011). *Distance education: A systems view* (3rd edition). Cengage Learning.

Pieper, D., Antoine, S. L., Mathes, T., Neugebauer, E. A., & Eikermann, M. (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology, 67*(4), 368-375. https://doi.org/10.1016/j.jclinepi.2013.11.007

Pollock, A., Campbell, P., Brunton, G., Hunt, H., & Estcourt, L. (2017). Selecting and implementing overview methods: implications from five exemplar overviews. *Systematic reviews, 6*(1), 145. https://doi.org/10.1186/s13643-017-0534-3

Pollock, M., Fernandes, R. M., Pieper, D., Tricco, A. C., Gates, M., Gates, A., & Hartling, L. (2019). Preferred Reporting Items for Overviews of Reviews (PRIOR): A protocol for development of a reporting guideline for overviews of reviews of healthcare interventions. *Systematic reviews, 8*(1), 335. https://doi.org/10.1186/s13643-019-1252-9

Polanin, J. R., Maynard, B. R., & Dell, N. A. (2017). Overviews in Education Research: A Systematic Review and Analysis: A Systematic Review and Analysis. *Review of Educational Research*, *87*(1), 172-203. https://doi.org/10.3102/0034654316631117

Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly*, *55*(1), 105-133. https://doi.org/10.1002/tesq.579

Schmidt, F. L., & Oh, I. S. (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes, 121*(2), 204–218. https://doi.org/10.1016/j.obhdp.2013.03.002

Sipe, T. A., & Curlette, W. L. (1997). A meta-synthesis of factors related to educational achievement: A methodological approach to summarizing and synthesizing metaanalysis. *International Journal of Educational Research, 25*, 583–698. https://doi.org/10.1016/S0883-0355(96)80001-2

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K-12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses. *Review of Educational Research*, *86*(4), 849–899. http://www.jstor.org/stable/44668238

Strech, D. & Tilburt, J. (2008). Value judgements in the analysis and synthesis of evidence. *Journal of Clinical Epidemiology, 61*(6), 521-524. https://doi.org/10.1016/j.jclinepi.2008.01.001

Sun, T., Wu, T., Martin, F., & Westine, C. D. (2025). Effects of multimedia-based education on learning outcomes: An evidence gap map analysis and second order meta-analysis. *Journal of Educational Multimedia and Hypermedia, 32*(1), 57-95. https://doi.org/10.70725/798356qwfmej

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2014). Can the computer replace the adult for storybook reading? A meta-analysis on the effects of multimedia stories as compared to sharing print stories with an adult. *Frontiers in psychology*, *5*, 1366.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A metaanalysis. *Review of Educational Research*, *85*(4), 698-739.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (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

Tracz, S. M., Elmore, P. B., & Pohlmann, J. T. (1992). Correlational meta-analysis: Independent and nonindependent cases. *Educational and Psychological Measurement, 52*, 879-888. https://doi.org/10.1177/0013164492052004007

Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*, 413–429. https://doi.org/10.1037/1082-989x.6.4.413

Yang, H. L. & Xiang, J. J. (2024). Mobile learning and students' academic achievement: A second-order meta-analysis study. *Journal of Educational Computing Research, 62*(4), 741-763. https://doi.org/10.1177/07356331241226589

Ying, X., Bougioukas, K. I., Pieper, D., & Mayo-Wilson, E. (2025). Weighted corrected covered area (wCCA): A measure of informational overlap among reviews. *Research Synthesis Methods*, 702-708. https://doi.org/10.1017/rsm.2025.19

Yun, J. (2011) The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning, 24*(1), 39-58. https://doi.org/10.1080/09588221.2010.523285

Zawacki-Richter, O., Cefa, B., & Bai, J. Y. (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