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

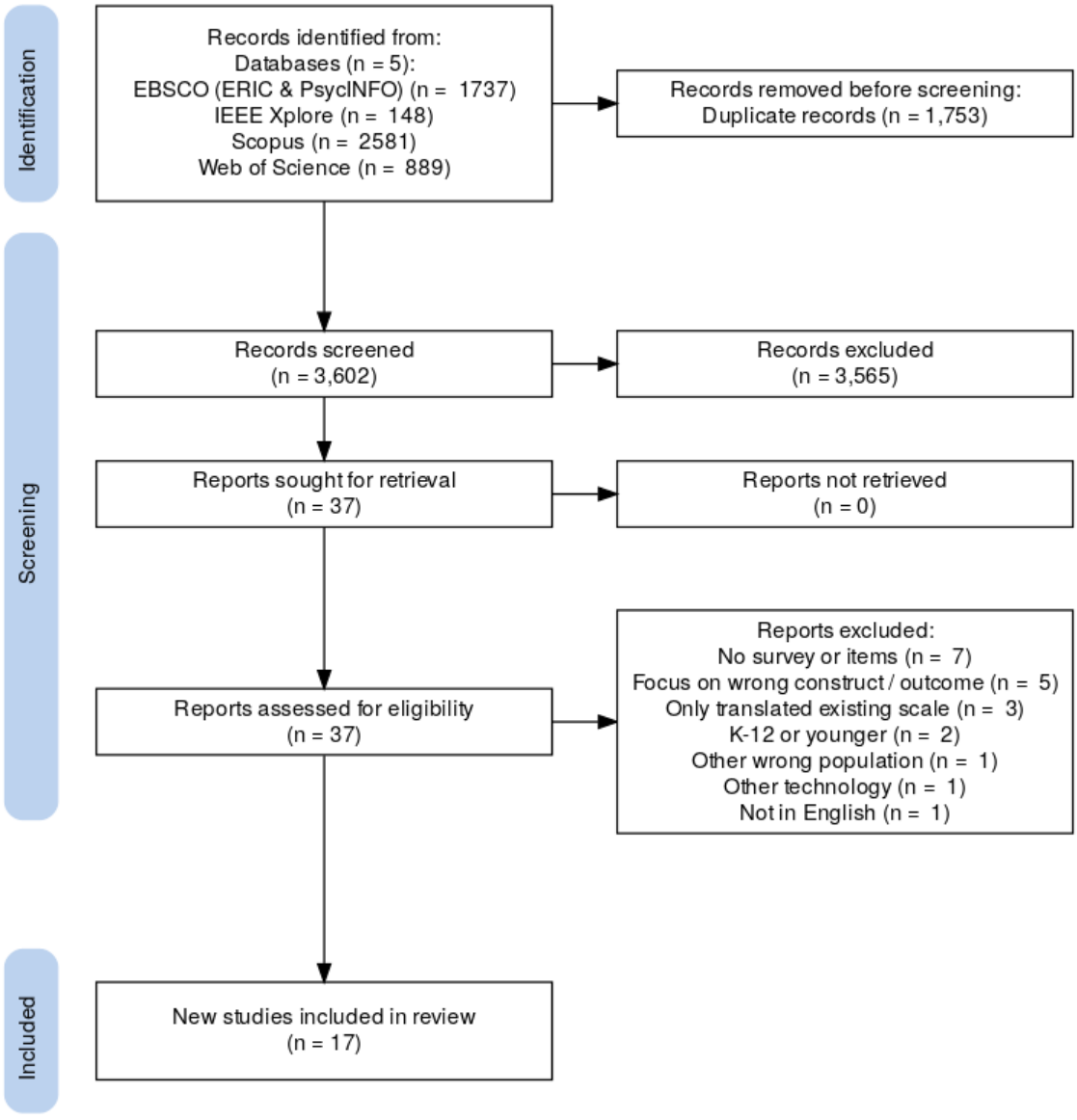
A total of 5,355 records were identified from five academic databases. Retrieved records were equally distributed among three authors, who independently conducted title and abstract screening based on predefined inclusion and exclusion criteria. Inter-rater reliability at the screening stage was assessed 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*)