INTERNATIONAL PREMIER JOURNAL OF LANGUAGES & LITERATURE (IPJLL)
VOLUME 3, ISSUE 2, 2025 p-ISSN: 3007-2336 e-ISSN: 3007-2344

INVESTIGATE HOW AI CAN ANALYZE PROJECT-BASED OR PERFORMANCE-BASED ASSESSMENTS MORE HOLISTICALLY

Moneeza Aslam moneezyaslam@gamil.com	MPhil Scholar, Department of Education, University of Gujrat, Gujrat, Punjab, Pakistan.
Dr. Saira drsaira.ijaz@uog.edu.pk	Lecturer, Department of Education, University of Gujrat, Punjab, Pakistan.
Irsa Tayyab irsatayyab019@gmail.com	MPhil Scholar, Department of Education, University of Gujrat, Gujrat, Punjab, Pakistan.

Abstract

This research investigates the potential for using artificial intelligence (AI) to examine project-based and performance-based evaluations holistically. Conventional assessment methods tend to concentrate on isolated task outcomes (e.g., end product or grade) and do not consider such factors as process, collaboration, creativity, iteration, and reflection on the part of the learner. Through the use of AI-powered analytics on rich data from project-based assessments (PBAs), this study investigates if AI can present worthwhile insights in various dimensions of students' performance—content knowledge, skills, process engagement, and metacognitive reflection. The research followed a mixed-methods study using 200 undergraduate teacher-education students with AI tools processing assessment artefacts (e.g., project outputs, records of student collaboration, process artefacts) and producing analytic reports. Results show that AI can accurately detect patterns of process engagement and collaboration, and that students whose AI-reports of process scores were also higher received more robust performance. Inference for design of assessment, teacher practice, and AI tool design are explored.

Keywords: *Artificial Intelligence, Project-Based Assessment, Performance-Based Assessment, Holistic Assessment, Educational Analytics, Process Data, Collaboration Analytics, Metacognitive Reflection*

Corresponding Author: Moneeza Aslam (MPhil Scholar, Department of Education, University of Gujrat, Gujrat, Punjab, Pakistan).
Email: moneezyaslam@gamil.com

1. Introduction

Over the past few years, the incorporation of artificial intelligence (AI) into educational assessment has become a revolutionary force with the potential to redefine how we quantify learning. Conventional assessment practices in higher education and teacher-education programs have typically focused on discrete, frequently pen-and-paper, high-stakes, one-time exams giving numeric answers in lieu of nuanced qualitative information on learning processes. However, the push towards more authentic assessment modalities—such as project-based assessment (PBA) and performance-based assessment (PeBA)—has created new opportunities and challenges in how educators evaluate student learning (Project-Based Learning, 2024). As noted by The Association of State Colleges & Universities (ASCCC) in their discussion of AI-powered education, authentic assessments such as projects, portfolios, presentations and simulations aim to evaluate higher-order thinking and skills rather than rote recall. Project-based evaluations provide deeper, more integrated understanding of student learning since they tend to involve real-world application of knowledge, collaboration, iterative design and reflection (Thomas, 2000; see also IEEE Education article on PBL).

But assessing PBAs continues to be cumbersome and prone to inconsistency: teachers need to work on process artefacts, group dynamics, creativity, iteration logs, reflection journals—involving uneven scoring and minimal feedback. Growing access to AI technologies makes available new means of processing large amounts of rich assessment data (text, multimedia, interactions) and inferring insights on learner processes, collaboration, metacognition, and outcomes.

As defined in the conceptual model by Kadel, Shailendra & Saxena (2025) for PBA in the generative AI era, the traditional measures have to change: the model stresses process-based evaluation, multi-modal and multifaceted data collection, and tailored AI-guided feedback.

In addition, Miller (2024) suggests that assessment professionals need to redefine their roles in the age of AI, spending less time on hand scoring and more on crafting assessments and scaffolding feedback based on insights from AI analytics.

Against this background, this research examines the ways in which AI can analyze project-based or performance-based evaluations holistically—that is, not only scoring the final product but assessing aspects like process engagement, collaboration quality, metacognitive reflection, iteration, and outcome. We apply our attention specifically to pre-service teachers undertaking project-based tasks and question whether AI analytics can offer robust and worthwhile measures of student process and performance, and how these match up with traditional outcome measures.

By exploring this in a teacher education program, we seek to provide empirical evidence on the value of AI-based holistic assessment for multifaceted authentic tasks, and provide implications for assessment design, teacher feedback and the use of AI tools in teacher preparation programs.

1.1.Rationale of the Study

Real assessments like project-based and performance-based activities are strongly promoted in teacher preparation because they enable pre-service teachers to exhibit actual classroom teaching skills (lesson planning, classroom management, reflecting on practice) instead of giving close-ended test answers.

However, it is challenging to implement and administer these assessments: teachers' time requirements for scoring, feedback and analysis of rich artefacts are heavy; consistency in scoring across students and scorers is tricky; and the majority of process dimensions (collaboration, iteration, reflection) are underresearched. AI has potential to relieve these burdens by processing large amounts of data (collaboration records, reflection texts, iteration history), identifying patterns (e.g., who did what in team work, how many revision cycles were there, quality of reflective discussion), and generating dashboards or feedback to be consumed by teachers and students.

Despite its potential, little empirical work so far exists on how AI can conduct such holistic evaluation in practice, particularly in teacher-education settings. Examining this deficit is opportune since as universities implement more sophisticated evaluations, the use of AI analytics to complement them can become a necessity to ensure scalability, quality of feedback, and more nuanced understanding of learning. Statement of the Problem

Although project-based and performance-based evaluations are cherished for their realism and capacity to measure higher-order abilities, their feasibility is limited by resource-prohibitive scoring, limited feedback, and inability to capture process and collaborative dimensions.

Conventional scoring approaches tend to concentrate on end results and overlook process variables (e.g., number of revision rounds, peer collaboration process, depth of reflection). Meanwhile, AI applications so far have largely concerned discrete assessment tasks (e.g., automated essay grading, multiple-choice analysis) instead of comprehensive analytics of intricate assessment artefacts. As such, teacher-education programmes currently do not have standards-approved means to use AI for integrated evaluation of PBAs—raising questions regarding feasibility, validity, reliability, and utility of AI-based analyses in this regard. This paper responds to the following general problem: Can AI effectively be applied to analyze project-based or performance-based evaluations in an integrated way (i.e., process, collaboration, iteration, reflection, outcome), and how do the derived analytics connect with traditional outcome measures?

1.2.Research Objectives

- To develop and deploy an AI analytics system that aggregates data from project-based tests (artefacts, logs, reflections, collaboration metrics) and generates holistic performance measures.
- To investigate correlations between AI-derived holistic measures (process engagement, quality of collaboration, reflection at the metacognitive level, number of iterations) and traditional outcome scores of project-based/performance-based testing.
- To assess teacher and student attitudes towards the AI analytics reports on the dimensions of usefulness, fairness, transparency and feedback value.
- To make suggestions about how teacher-education courses may incorporate AI-powered analytics into the evaluation of real-world tasks.

1.3.Research Questions

- What holistic performance measures may be extracted by AI from project-based test data (e.g., process engagement, collaboration, reflection, iteration) and how may these measures be operationalised?
- How statistically correlate with traditional outcome scores of project-based and performance-based assessments are AI-produced holistic performance measures
- How are the AI-analystic reports viewed by teacher educators and students as useful, fair, transparent, and giving feedback?
- What are design and implementation considerations for effective incorporation of AI analytics into project-based assessment processes in teacher-education contexts

1.4.Definitions of Key Variables

- **AI-Generated Holistic Performance Indicators:** Quantitative or qualitative measures emerging from the AI analytics system from data related to project-based/performance-based activities. These are:
- **Process Engagement:** number of cycles of revision, task time, artefact version sequence.

- **Collaboration Quality:** assessed through contribution logs, peer to peer interaction, social network analysis for team work.
- **Metacognitive Reflection:** reflection text quality of students, insight depth, frequency of self-regulation words, evaluated through natural-language processing (NLP).
- **Iteration Count:** student-submitted number of redesign or revision cycles.
- **Traditional Outcome Score:** The score assigned by teacher-educators on the end product or performance of the project/assessment, according to predefined rubrics (e.g., content knowledge, creativity, technical performance).
- **Perceptions of AI Reports:** Students' and teachers' subjective ratings of AI-produced reports regarding their usefulness, perceived fairness, transparency, trust, and value of feedback.
- **Project-Based/Performance-Based Assessment (PBA/PeBA):** An assessment mode where students produce prolonged tasks or performances (e.g., planning a teaching unit, conducting a micro-teaching session, designing a multimedia project) that entail planning, iteration, collaboration and reflection.

2.Literature Review

Research on the application of AI in testing—and particularly in real-world, project-based or performance-based settings—is increasing but still in its infancy. A systematic review by Luo, Zheng, Yin et al. (2025) of AI-based learning tools concluded that while the applications of AI to assessment and evaluation were large, their effects on skill-based outcomes (which tend to map onto project-based activities) were less robust.

The authors contended that design elements (e.g., transparency, multimodal input, student agency) and assessment frameworks are underdeveloped.

With respect to project-based learning, the role of AI has been investigated in generative and co-design. Zheng, Yuan, Guo et al. (2024) led workshops with college students investigating the potential of AI in supporting PBL and examined student-AI interaction data as new assessment material.

Their work highlighted the importance of student attitude and preference within AI impacting the way data were to be interpreted, highlighting the difficulty of designing AI analytics for PBA environments.

Du Plessis (2025) explored how lecturers in open-distance and e-learning environments applied ChatGPT to facilitate project-based assessment, prioritising critical thinking, creativity and problem-solving as advantages of PBA—but also raising questions regarding fairness, authenticity and student preparedness.

Miller's (2024) work on assessment professionals reported that AI displaces the traditional role of human assessors towards defining assessment architecture and interpreting AI output, but not just manual scoring.

From the PBL literature more generally, the best practices of good project-based assessment are clearly established: authenticity (real-world connectedness), student agency, teamwork, inquiry and reflection.

These aspects mean that an overall evaluation model should encompass not merely end-product quality but the process, iteration and peer collaboration.

Collectively, the literature indicates that although AI technologies offer promise for deep analytic ability in testing, their use to comprehensive evaluation of PBAs is limited. Some of the main areas of limitation are: process metric operationalisation, process indicator-outcome linkage, human judgement validation of AI metrics, fairness/transparency of AI analytics, and teacher/student perceptions of AI reports. This research attempts to address some of these areas of limitation.

The explosive growth of AI methods—most notably natural language processing (NLP), machine learning, and multimodal analytics—has provided new opportunities for measuring intricate, real-world learning assignments such as project-based and performance-based assessments (PBAs). Current systematic research contends that AI applications to testing are wide-ranging (e.g., machine scoring, process mining, multimodal feature extraction), but a large number of studies are still exploratory and focus on tool capability rather than stringent validation of learning effectiveness (Luo et al., 2025). Researchers thus urge thoughtful design of AI evaluation systems that address input variety (text, speech, video, interaction logs), transparent operational definitions of constructs, and strong validity evidence prior to such systems' application in high-stakes decisions (Luo et al., 2025)

Work directly investigating AI in PBL environments emphasizes both potential and interpretive richness. Zheng et al. (2024) employed co-design workshops to unveil how interactions between students and AI create novel assessment data streams (e.g., draft histories, prompts, AI suggestions) and illustrated that student attitudes to AI actually influence those data substantively. Their results indicate that analytics constructed from AI-facilitated interactions cannot be taken at face value without considering learner behavior, intentionality, and affordances of the given AI instrument—a vital proviso for anyone trying to take AI output as neutral indices of student ability (Zheng et al., 2024).

Contrasting studies of teachers' practices identify a transformation in roles of the assessor with the presence of AI in the ecology of assessment. Miller (2024) and Du Plessis (2025) contend that AI replaces some conventional manual scoring work but raises the level of assessor work around specifying assessment architecture, interpreting indicators from AI, and ensuring fairness and authenticity. Du Plessis's application of open-distance contexts established practical advantage—shorter feedback loops, scaffolding of thinking—but also set out new challenges for lecturers, such as interpreting process metrics and resolving AI judgments against contextual awareness of student work (Du Plessis, 2025; Miller, 2024).

Methodological innovations from learning analytics and process-mining literatures are particularly applicable to comprehensive PBA analysis. Process mining and sequence analysis methods can elicit revision patterns, dialog turns, and contribution networks from collaboration logs; NLP models can label reflection depth or argument quality; and multimodal fusion techniques can aggregate these signals into composite indices of engagement or teamwork quality (Purwono et al., 2024; Kadel et al., 2025). But the literature always underscores that the construct validity and reliability of such AI-generated indicators should be empirically substantiated—preferably by triangulation with human raters, rubric-based assessment, and outcome measures—prior to assuming that they capture meaningful dimensions of learning (Purwono et al., 2024; Luo et al., 2025).

Fairness, transparency, and explainability are perpetual issues. Multiple authors point out that black-box AI metrics have the potential to obscure biases (e.g., rewarding verbosity instead of depth, penalizing students who contribute less but more substantial contributions) and can harm specific groups of people if models are trained on culturally limited corpora (Zheng et al., 2024; Du Plessis, 2025). Accordingly, the literature suggests human-in-the-loop designs where AI outputs appear as interpretable signs along with provenance data (how the score was calculated), and instructors still have discretion to modify or contextualize AI inferences (Miller, 2024; Luo et al., 2025).

A second key strand is concerned with transfer and formative utility. Even where AI metrics correspond to rubric ratings or ultimate product quality, there are questions whether AI-driven feedback enhances learning processes or subsequent performance. A number of empirical studies establish encouraging short-term correlations (e.g., process-measures predicting outcome variance), but longitudinal support for transfer to later authentic tasks is limited (Luo et al., 2025). The research therefore demands experimental and longitudinal designs that examine if AI-aided formative loops (AI analytics → teacher mediation → student revision) yield lasting learning gains over what traditional feedback does.

Design principles that unfold from the reviewed work steer towards multimodal, transparent, and learner-centred systems. Good AI for PBA must (a) draw from several data streams (artifacts, logs, reflections), (b) clearly operationalise process constructs and relate them to rubrics, (c) offer understandable feedback for action by teachers and students, and (d) be co-designed with end-users to produce context validity (Zheng et al., 2024; Kadel et al., 2025; Purwono et al., 2024). These guidelines serve to cross the gap of technical potentiality and pedagogical utility that most papers recognize.

Lastly, a persistent empirical lacuna spurs further investigation: although numerous studies state the promise of AI-mediated holistic evaluation and document early correlation or feasibility results, there is little validated evidence on (1) how to operationalize process measures in a robust manner, (2) how AI indicators correspond to high-quality human judgment across contexts, (3) the implications of fairness for diverse learners, and (4) the pedagogical effects of AI-facilitated formative cycles (Luo et al., 2025; Du Plessis, 2025). Filling these gaps will call for mixed-methods research blending algorithm development, psychometric validation, human-AI adjudication studies, and field trials assessing learning transfer—exactly the agenda that the current study seeks to fill.

3.Research Methodology

This research utilized a mixed-methods design consisting of quantitative analysis of AI-derived indicators and traditional outcome scores, in addition to qualitative survey/interview data regarding teacher and student perceptions.

3.1. Design and Procedure

A sample of 200 undergraduate pre-service teachers in a teacher-education program in a large public university in Punjab, Pakistan, took part in a semester-long PBA. Students worked on a teaching-unit design project (planning, collaboration, iteration and reflection) backed by an AI analytics system. Data gathered included: version logs, chat logs of collaboration, reflection journals, artefact submission, and the resultant final teacher-scored product. The AI analytics system analyzed these data and generated indicator scores for each student/team (process engagement, collaboration quality, metacognitive reflection, iteration count). Finally, surveys and focus-group interviews collected 30 students' and 10 teacher educators' perceptions of the AI reports.

3.2.Population and Sampling

The population includes all pre-service teachers undertaking the teacher-education programme in the semester (N ≈ 300). A purposive sample of 200 students was recruited (complete cohort) and 10 teacher-educators overseeing the projects. To assess the perception component, a stratified sample of 30 students (at high, medium, low process engagement levels) was recruited to participate in focus groups. Teacher sampling was purposive to ensure inclusion of those with different experience levels

3.3.Data Collection

Data sources were:

- Log records of project artefact versioning (timestamps, file names, content length
- Collaboration/chat records (team discussion, no. of messages, contributor numbers
- Student reflection diaries (text submitted at end of every revision cycle)
- Last project artefact and teacher-assessed result based on rubric (knowledge, pedagogy, creativity, execution)
- AI analytic reports generated by the system per student/team
- Surveys (Likert scale + open-ended) assessing perceptions of AI reports (usefulness, fairness, transparency, trust)
- Focus-group interview transcripts and interviews with teacher educators.

3.4.Data Analysis

- **Quantitative:** Descriptive statistics of AI indicator scores and standard outcome scores; Pearson correlation coefficients among process indicators and outcome scores; multiple regression of outcome score predicted from AI indicators while controlling for pre-existing GPA. Reliability analysis of AI measures (inter-rater reliability between AI scores and teacher ratings on selected process measures).
- **Qualitative:** Thematic analysis of open-ended survey responses and interview transcripts on usefulness, fairness, transparency, trust, and implementation challenges of the AI analytics.

3.5.Research Instruments

- **AI Analytics Framework:** Custom-developed system that analyzes log/artefact/reflection data and calculates four indicator scores (process engagement, collaboration quality, metacognitive reflection, iteration count).
- **Teacher Outcome Rubric:** A teacher-educator standard rubric for grading final projects on dimensions of content knowledge, pedagogy design, creativity, iteration quality, and reflection (0-100 score range).
- **Perception Survey:** 20-item Likert scale survey (1 = strongly disagree to 5 = strongly agree) assessing four domains of perception (usefulness, fairness, transparency, trust) with two open-ended items.
- **Focus-Group Interview Protocol:** Semi-structured guide probing experiences of receiving AI reports, how reports impacted reflection/iteration, and impressions of fairness and transparency.

4.Results and Analysis

4.1.Quantitative Results

Table 1: Correlations between AI Indicators and Conventional Outcome Score

Indicator	Pearson	p-value
Process Engagement	.52	< .001
Collaboration Quality	.47	< .001
Metacognitive Reflection	.55	< .001
Iteration Count	.40	< .001

4.1.1. Interpretation

Table 1 presents moderate to strong positive relationships between the AI-based indicators and the traditional outcome scores. More specifically, metacognitive reflection had the highest coefficient ($r = .55$), which meant that students whose reflection journals were more highly rated by the AI tended to produce higher quality project outcomes. The relationships are significant statistically ($p < .001$), which implies significant relationships between process measures and performance.

Table 2**: **Regression Predicting Outcome Score from AI Indicators (N = 200)

Predictor	B	SE	β	p- value
Process Engagement	0.30	0.07	.28	< .001

Collaboration Quality	0.24 0.08	.22	.003	
Metacognitive Reflection	0.35	0.08	.30	<.001
Iteration Count	0.18	0.09	.15	<.045
Prior GPA (control)	0.20	0.05	.18	<0.001

Model: $R^2 = .46$, Adjusted $R^2 = .44$, $F(5,194) = 33.4$, $p < .001$

4.1.2. Interpretation

Table 2 shows that, after controlling for previous GPA, the four AI-based indicators together account for roughly 46 % of the variance in outcome scores—an impressive share. Of the predictors, metacognitive reflection possessed the largest standardized coefficient ($\beta = .30$), followed by engagement in process ($\beta = .28$). The quality of collaboration and number of iterations were also significant predictors but with smaller effect sizes. These findings imply that the AI-generated holistic indicators significantly associate with student performance on PBA tasks.

4.2.Qualitative Results

From thematic analysis of perception data, there were four dominant themes:

1. **Usefulness for feedback and reflection** – Students told us that they found receiving AI reports useful to visualize their process, see how many revisions they made, and think about their pattern of working together. One student wrote: "Seeing the chart of my messages and revisions made me realize I wasn't contributing much at the beginning."
2. **Fairness and transparency issues** – Some teacher-educators questioned the computation of the AI metrics and whether or not they fully accounted for creativity or context. One teacher remarked: "The AI might tally up number of messages, but quality of ideas is much more difficult to measure."
3. **Trust in AI analysis** – While students appreciated the value of AI reports, some students were cautious: "I felt the AI score didn't reflect my reflection depth, just length of journal." This refers to user validity and trust issues
4. **Implementation and workflow challenges** – Time, training and infrastructure limitations were highlighted by students and teachers alike: "We needed to learn how to interpret the AI dashboard; initial confusion delayed progress."

4.3.Summary

This research designed and implemented an AI analytics platform to evaluate comprehensively project-based and performance-based assignments in an educator training environment. The AI-derived measures of process engagement, collaboration quality, metacognitive reflection and iteration count were highly correlated with traditional outcome measures, and collectively accounted for almost half the variance in performance. Students and teachers also reportedly used the AI reports for feedback and reflection purposes, although there are still issues with fairness, transparency and trust.

5.Discussion

The results are consistent with the wider literature on AI-based assessment that advocates for more integrated analytics of real-world tasks (Luo et al., 2025).The emergence of metacognitive reflection as the strongest predictor is also consistent with the historical PBL literature that underscored reflection as a primary force for learning (Thomas, 2000). The $R^2 \approx .46$ of the regression model is also highly encouraging and points to AI metrics accounting for significant variance in students' outcomes over prior GPA.

But the qualitative results indicate that AI analytics integration is not entirely technical—trust, metric design clarity, training, and infrastructure are involved. Teacher educators' concerns regarding quality of contributions and creativity processes resonate with previous research on AI fairness and assessment transparency (Miller, 2024).

These findings indicate that AI analytics can potentially support teacher scoring by offering process data and feedback loops, thus facilitating scalability and richness of assessment. However, teachers need to design and validate AI metrics carefully, offer transparency to users, and integrate dashboards into reflective feedback cycles.

Limitations are the one-institution, one-semester setting, AI report novelty effects, and limited validation against independent human raters of AI scoring. Future studies should investigate longitudinal transfer of AI-measured process metrics into actual classroom teaching performance, diverse contexts, and subject domains.

5.1.Recommendations

- Teacher-education courses should pilot AI analytics dashboards for project-based/performance-based measurement to offer process feedback, but pair them with teacher-moderated reflection sessions to interpret measures.

- AI testing tool developers have to put the highest emphasis on metric transparency (i.e., how collaboration quality is calculated), enable human override or comment, and involve teachers in designing metrics.
- Professional development for teacher educators must include interpretation of AI analytics, leveraging them within feedback cycles, and explaining metric meaning to students.
- Institutions should fund infrastructure (log capture, collaboration tools, reflection journals) to provide robust data for AI systems and to secure students' data privacy and ethics.
- Future studies should cross-check AI measures with human ratings, investigate disciplinary difference (STEM vs arts), and investigate whether process-analytics feedback enhances long-term teaching competence.

6.Conclusion

This research offers empirical proof that analytics provided by AI have the capability of meaningfully analyzing project-based and performance-based measures holistically, encapsulating process, collaboration, reflection and iteration alongside outcome. The moderate to high correlations and explanatory power of AI indicators imply that such analytics can augment conventional scoring and add value to feedback and insight. Concurrently, qualitative findings highlight the value of open design, educator training and incorporation of AI reports within pedagogic workflows. With teacher-education programs increasingly embracing genuine assessment, AI analytics present a scalable and insightful complement—given implementation is considerate, ethical and aligned with educator and student requirements.

References

"With AI – Authentic Assessment and Learning". (n.d.). ASCCC. https://asccc.org/content/ai-powered-education-authentic-assessments-and-learning

Du Plessis, E. (2025). Embracing project-based assessments in the age of AI in open distance e-learning. International Journal of Information and Education Technology, 15(2), 374–380.

Kadel, R., Shailendra, S., & Saxena, U. R. (2025). Navigating the new landscape: A conceptual model for project-based assessment (PBA) in the age of GenAI. arXiv. https://arxiv.org/abs/2508.11709

Luo, J., Zheng, C., Yin, J., & Teo, H. H. (2025). Design and assessment of AI-based learning tools in higher education: A systematic review. International Journal of Educational Technology in Higher Education, 22, 42. https://doi.org/10.1186/s41239-025-00540-2

Miller, W. (2024). Adapting to AI: Reimagining the role of assessment professionals. Assessment & Evaluation in Higher Education. https://files.eric.ed.gov/fulltext/EJ1461832.pdf

Project-Based Learning: A catalyst for student engagement and holistic learning. (2024). IEEE Teaching & Learning. https://teaching HYPERLINK "https://teaching.ieee.org/project-based-learning-a-catalyst-for-student-engagement-and-holistic-learning/".ieee.org/project-based-learning-a-catalyst-for-student-engagement-and-holistic-learning/

Purwono, U., Ma'arif, M., Marhoon, M., Abougarair, S., & Sharkawy, A. (2024). Generative AI as a pedagogical co-pilot in higher education: A systematic review. Journal of Technology and Pedagogy in Education, 4(1), 21–45.

Zheng, C., Yuan, K., Guo, B., Mogavi, R. H., Peng, Z., & Ma, X. (2024). Charting the future of AI in project-based learning: A co-design exploration with students. arXiv. https://arxiv.org/abs/2401.14915