

AICoFe: Implementation and Deployment of an AI-Based Collaborative Feedback System for Higher Education

Alvaro Becerra^{1,*}, Alejandra Palma¹ and Ruth Cobos¹

¹GHIA Group, School of Engineering, Universidad Autónoma de Madrid, Spain

Abstract

Effective peer feedback is essential for developing critical reflection in higher education, yet its impact is often limited by the inconsistent quality of student-generated comments. This paper presents the implementation and deployment of AICoFe (AI-based Collaborative Feedback), a system designed to bridge this gap through a human-centered AI approach. We describe a modular architecture that orchestrates a multi-LLM pipeline—utilizing GPT-4.1-mini, Gemini 2.5 Flash, and Llama 3.1—to synthesize quantitative rubric data and qualitative observations into coherent, actionable feedback. Key to the system is a “teacher-in-the-loop” mediation workflow, where educators use specialized Learning Analytics dashboards to curate and refine AI-generated drafts before delivery. Furthermore, we detail the underlying data infrastructure, which employs a hybrid SQL and MongoDB strategy to ensure traceability and manage semi-structured feedback versions.

Keywords

Learning Analytics, Peer Feedback, Generative AI, Human-AI Collaboration, Large Language Models, Higher Education, Dashboards, Collaborative Assessment

1. Introduction

Among the pedagogical strategies aimed at supporting the development of skills, peer feedback and peer assessment have proven to be particularly effective. By engaging students in the evaluation of their peers’ work, peer feedback promotes active learning, critical reflection, and deeper understanding of quality criteria [1, 2]. Moreover, acting as evaluators allows students not only to receive feedback but also to learn from the assessment process itself, strengthening their self-regulation and evaluative judgment skills [3].

Despite its educational potential, the implementation of peer feedback presents several well-documented challenges. Students often struggle to provide constructive, specific, and actionable comments, especially when they lack feedback literacy or confidence in their evaluative abilities [4]. As a result, peer feedback may become superficial, inconsistent, or difficult for recipients to interpret and apply [5]. These limitations can reduce student engagement and diminish the impact of peer assessment on learning outcomes.

Recent advances in Generative Artificial Intelligence (GenAI) open new opportunities to address these challenges and enhance peer feedback processes. Prior approaches to automatic feedback relied mainly on predefined rules, templates, or numerical indicators [6, 7]. More recently, large language models have demonstrated the ability to generate coherent, context-aware, and personalized feedback based on both quantitative data and qualitative inputs [8]. In educational settings, GenAI has been successfully applied to improve the clarity, usefulness, and actionability of feedback, supporting students in understanding their performance and identifying concrete steps for improvement [9, 10].

Building on this perspective, this paper presents AICoFe (*Artificial Intelligence-based Collaborative Feedback* system), a system designed to enhance peer and self-evaluation processes in educational settings. From a human-computer interaction perspective, AICoFe is conceived as a human-centered, interactive system that supports human-AI collaboration through role-specific dashboards and teacher-in-the-loop feedback mediation. The system enables both teachers and students to conduct evaluations

LASI Spain 26: Learning Analytics Summer Institute Spain 2026, Zamora, May 07–08, 2026

*Corresponding author.

✉ alvaro.becerra@uam.es (A. Becerra); alejandra.palma@uam.es (A. Palma); ruth.cobos@uam.es (R. Cobos)

🆔 0009-0003-7793-2682 (A. Becerra); 0009-0005-6555-7250 (A. Palma); 0000-0002-3411-3009 (R. Cobos)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

using customized rubrics that integrate quantitative scores and qualitative observations. Powered by generative AI, AICoFe generates personalized feedback for each student by combining the outputs of multiple large language models, providing actionable insights to improve their skills. The system’s Learning Analytics (LA) dashboards offer dynamic features, including personalized summaries, comparative graphs, tools for self-assessment, and a feedback history interface that tracks LLM contributions and supports teacher draft management. Additionally, video recordings of student performances support reflective learning and a more thorough evaluation process.

Extending our previous work [11], this paper focuses on the implementation details and deployment experience of the system. The remainder is structured as follows. Section 2 briefly reviews related work. Section 3 describes the system architecture and its main technical components in detail. Section 4 presents conclusions and future work.

2. Related Work

2.1. Automatic Feedback Systems in Education

Before the emergence of GenAI, automatic feedback systems in education primarily relied on rule-based mechanisms, comparing student responses to predefined answers or visualising performance indicators through dashboards [6, 7]. Providing automated feedback for oral presentations has been a notable application of this line of work, with multimodal systems demonstrating measurable improvements in presentation skills in real learning settings [12] and recent open-source tools enabling broader adoption and contextual adaptation [13]. Some approaches incorporated machine learning techniques such as embedding models to assess idea novelty in collaborative settings [14], or used fixed templates to deliver automatic feedback in online learning environments [15].

2.2. Learning Analytics Dashboards

Learning Analytics dashboards (LADs) have emerged as a mechanism to support reflection and self-regulation in educational settings, offering visualisations of performance data that help students and teachers make sense of assessment outcomes [16]. Research on LAD design has highlighted the importance of aligning dashboard features with pedagogical goals and the specific needs of different user roles, including students and instructors [6]. Co-design approaches have proven particularly effective in this regard, as illustrated by dashboards developed in collaboration with instructors to deliver contextualised, human-centred feedback in MOOC environments [15]. Related work has further expanded this perspective by exploring human-centered visual attention analytics based on eye-tracking and AI [17, 18], as well as multimodal dashboard systems for MOOCs [19] that combine biometric, behavioral, and log data to support a more holistic interpretation of learner activity [20].

2.3. GenAI and LLMs for Feedback Generation

The emergence of large language models has substantially expanded the possibilities for automatic feedback generation in education. Unlike earlier rule-based approaches, LLMs can produce coherent, context-aware, and personalised feedback from both quantitative and qualitative inputs [8]. Empirical studies report positive student perceptions of GenAI-generated feedback, particularly valuing its clarity, level of detail, and usefulness [9, 21, 22, 10].

GenAI has also begun to support peer feedback specifically, helping to make comments more constructive and actionable while reducing unhelpful or inappropriate contributions [23, 24, 25]. Peer assessment research has long highlighted both the potential and the limitations of student-generated feedback: while peer feedback promotes active learning and evaluative judgment [1, 2], students frequently struggle to produce constructive and specific comments, particularly when they lack feedback literacy [4], resulting in feedback that may be superficial, inconsistent, or difficult to act upon [5]. GenAI-assisted approaches offer a promising path to address these limitations at scale. Unlike prior systems that rely

on a single model or fixed templates [15, 14], AICoFe combines three independently fine-tuned LLMs under teacher mediation, enabling richer perspective diversity and preserving pedagogical authority over the final feedback.

3. System Description

AICoFe is organised into four main modules (see Figure 1): (i) the **Visualization Module**, which provides role-specific dashboards; (ii) the **Management Module**, which handles data storage and connectivity; (iii) the **Feedback Generation Module**, which orchestrates multi-LLM feedback production; and (iv) the **Recording Module**, which captures student performances in video and audio format.

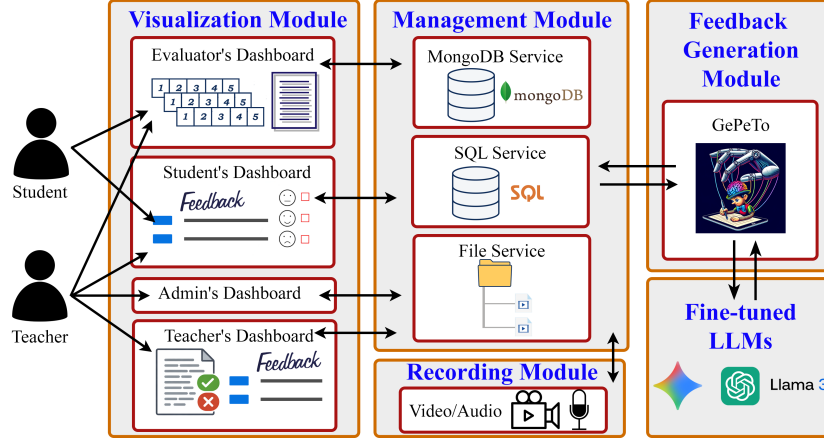


Figure 1: Architecture of the AICoFe system, showcasing its modular design with four main modules.

3.1. Management Module

The Management Module provides the data infrastructure underpinning the entire system. It integrates two complementary database technologies: a relational SQL database and a MongoDB document store.

The **SQL database** manages structured academic data, including courses, student groups, rubric definitions, and user accounts. It also stores the evaluators' quantitative scores and qualitative comments associated with each assessment. It supports the organisation of evaluation activities across multiple courses and cohorts, enabling the system to be deployed simultaneously in different educational contexts.

The **MongoDB database** is used to store semi-structured data that varies in shape across evaluation instances, such as the multiple feedback versions generated by each LLM, teacher drafts and curations, and the metadata associated with each feedback sentence (including its source LLM). It also stores the ratings of coherence and usefulness assigned to the generated feedback. This flexible document model is particularly suited to the multi-LLM feedback pipeline, where each evaluation instance may generate a variable number of feedback versions with different internal structures.

The **File Service** manages the storage and retrieval of audiovisual recordings associated with student presentations, linking them to the corresponding evaluation instances in the relational database. Prior to participation, students are provided with an informed consent form covering data usage and recording authorisation; participation in video recording is strictly opt-in. All audiovisual data are stored securely in compliance with European data protection regulations (GDPR).

3.2. Visualization Module

The Visualization Module provides a set of interactive Learning Analytics dashboards developed with the Dash framework. Each dashboard is tailored to the role of the user interacting with the system.

The screenshot shows a dashboard titled "DESIGN" with two evaluation sections. The first section, "Text Readability", has a Likert scale where "Excellent" is selected, and a comment box containing "A good font size on all slides, with good background contrast." The second section, "Relevance of Visual Elements", has a Likert scale where "Very Acceptable" is selected, and a comment box containing "Good charts and diagrams that added value to the presentation."

Figure 2: Screenshot of the AICoFe evaluator dashboard.

3.2.1. Evaluator Dashboard

The Evaluator Dashboard (Figure 2) supports both teachers and students during the assessment process. After selecting a rubric and a target student, evaluators assign scores to each rubric item using a Likert scale and complement these with written qualitative comments. Detailed descriptions of the rubric levels remain accessible throughout the evaluation to support alignment in the interpretation of assessment criteria. The system also captures interaction metadata — such as the timing of score selections — which can be examined later to explore evaluation dynamics and detect potential biases.

3.2.2. Student Dashboard

The Student Dashboard (Figure 3) is designed to foster reflective learning through self-assessment and feedback exploration. Students can access a video recording of their own presentation and complete a self-evaluation using the same rubric applied by peers and teachers. Once submitted, the dashboard visualises differences and alignments between self-assessment scores and the aggregated evaluations received. Students also receive AI-supported collaborative feedback composed by the teacher from the LLM-generated outputs. A short questionnaire allows students to express their agreement with and perceived usefulness of the feedback, providing additional qualitative input for system evaluation.

3.2.3. Teacher Dashboard

The Teacher Dashboard is the central interface for feedback review and curation. It allows teachers to inspect quantitative scores and qualitative comments provided by both peers and teachers (Figure 4a), access student performance videos, and review the feedback generated by the different LLMs supported by the system. Teachers can also compose the final feedback by selecting sentences from the LLM outputs (Figure 4b).

Teachers can select individual sentences or entire paragraphs from the AI-generated feedback of each LLM and combine them to compose the final feedback that will be presented to the student. This selective composition reinforces the human-centered nature of the system by positioning teachers as active mediators of AI-generated content. Additionally, the dashboard integrates a text-to-speech functionality (using Google Text-to-Speech) that enables teachers to listen to the detailed evaluation comments, facilitating review without the need to read every comment individually.

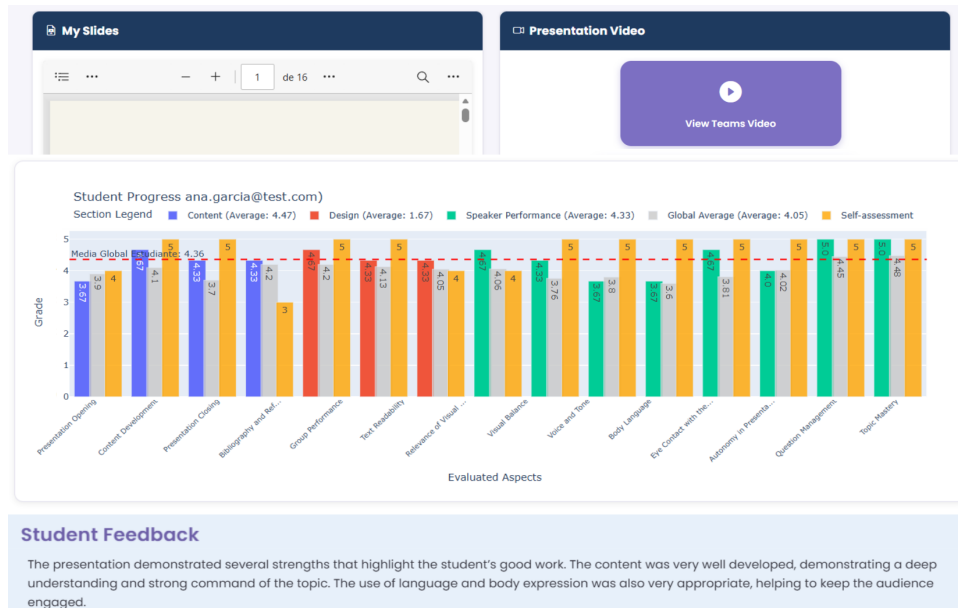


Figure 3: Screenshot of the AICoFe student dashboard.

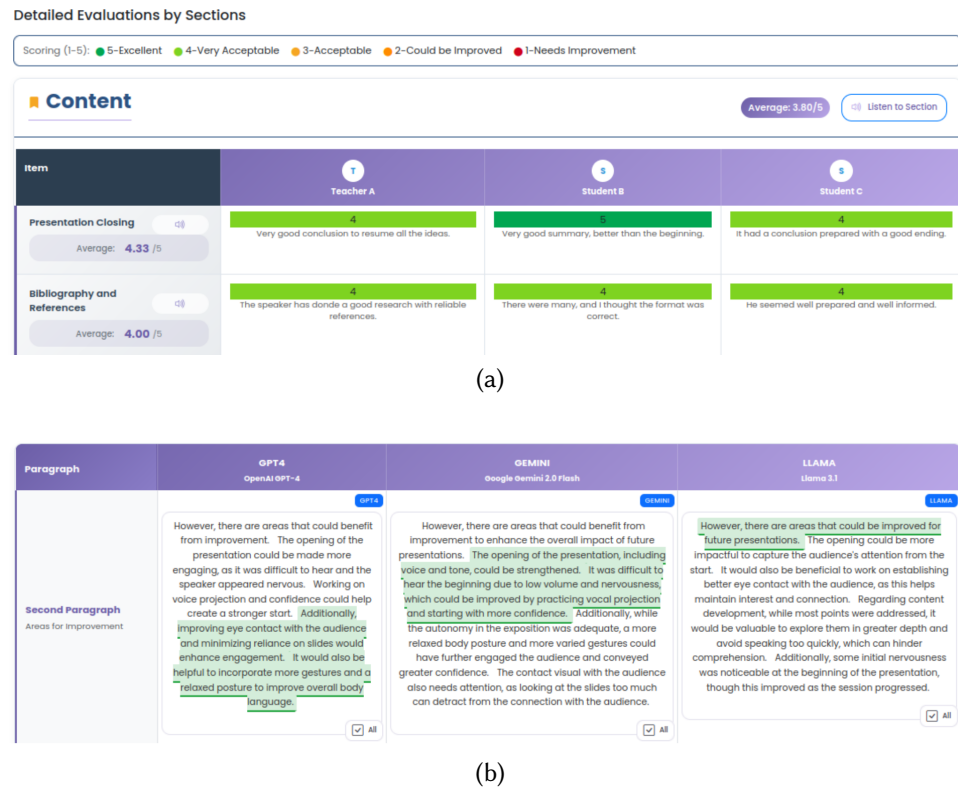


Figure 4: Teacher Dashboard views: (a) granular feedback from the three LLMs with sentence-level selection, and (b) detailed evaluations by rubric item and evaluator.

3.2.4. Feedback History and LLM Contribution Tracking

A key feature of the Teacher Dashboard is the **Feedback History** interface (Figure 5), which provides a comprehensive log of all feedback sent to each student. For each feedback entry, the interface displays a visual legend indicating the proportion of content contributed by each LLM (GPT-4.1-mini, Gemini 2.5 Flash, and Llama 3.1), as well as the extent to which the teacher modified or curated the AI-generated text before sending it to the student.

This transparency mechanism serves two purposes. First, it allows teachers to reflect on their own curation patterns across students and sessions, potentially identifying tendencies to favour certain LLMs or to apply greater or lesser editorial effort. Second, it provides a data source for analysing the role of teacher mediation in the feedback process at scale.

In addition to the history log, the interface supports draft management: teachers can save intermediate versions of composed feedback as drafts before finalising and sending them to students. This is particularly useful in settings where feedback curation is done across multiple sessions or where the teacher wishes to revisit a draft after reviewing additional student materials.

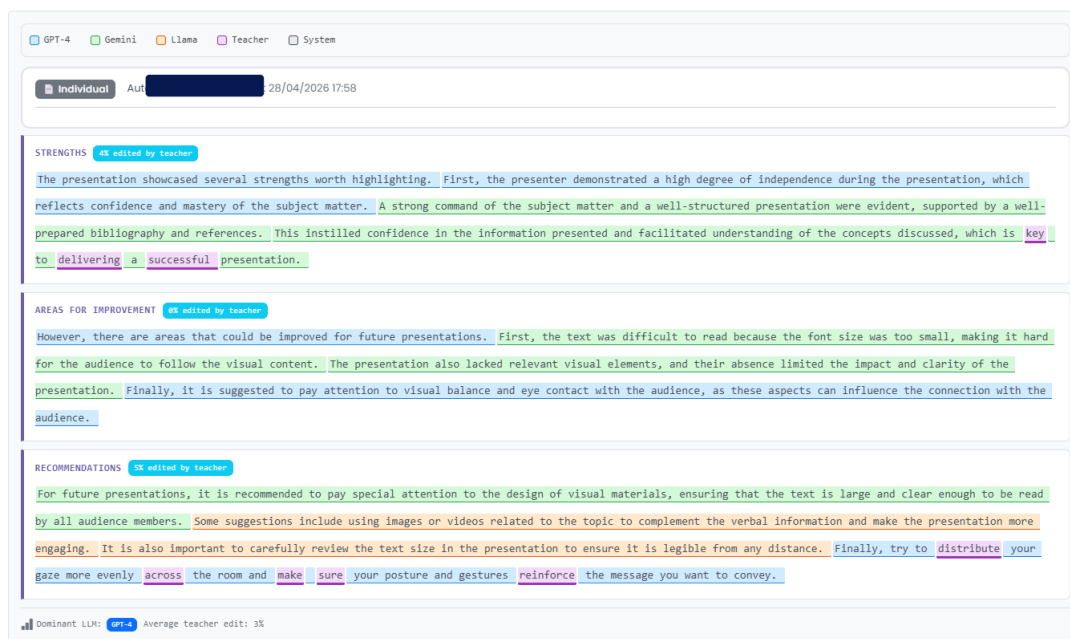


Figure 5: Feedback History interface in the Teacher Dashboard, showing the LLM contribution legend.

3.2.5. Admin Dashboard

System configuration and evaluation orchestration are managed through the Admin Dashboard, which allows administrators to create courses, define student groups, assign rubrics, and change the language of the system and the generated feedback. This dashboard is intended for system administrators and is not visible to students or teachers during normal operation.

3.3. Feedback Generation Module

The Feedback Generation Module is responsible for producing personalised feedback by combining quantitative assessment results from teachers and peers with the qualitative observations collected during the evaluation process. The module extends prior work [26] to a multi-model setting in which three independent feedback instances are generated, one per LLM.

3.3.1. Pre-processing and Prompt Construction

As an initial step, the module pre-processes the qualitative input collected from evaluators. Textual comments are normalised to remove invalid characters and are automatically screened to assess their relevance with respect to the evaluated rubric item using a keyword-based filtering heuristic that matches comments against rubric-specific term lists. Once the data have been filtered, the system constructs structured prompts based on course-specific templates. These prompts integrate average scores per rubric item with validated qualitative observations provided by multiple evaluators. To

improve contextual accuracy, the prompts also incorporate rubric level descriptions and selected instructional materials associated with the assessed skill. All data sent to the generative models are anonymised prior to transmission, ensuring that no student-identifying information leaves the institutional environment in identifiable form before reaching commercial LLM providers.

3.3.2. Multi-LLM Generation

The module supports three LLMs: **OpenAI GPT-4.1-mini**, **Google Gemini 2.5 Flash**, and **Meta Llama 3.1**. Using three distinct models enriches pedagogical feedback with diverse perspectives while reducing the risk of single-model biases or systematic errors reaching students.

Each model receives the same structured prompt and independently generates a feedback instance. All three models are fine-tuned using data from the SOPHIAS dataset [27], which comprises 50 AICoFe evaluation instances for which teachers generated feedback using the LLMs and subsequently reviewed and refined it. These teacher-validated feedback instances are used as training data to improve the models' alignment with the pedagogical objectives of the course.

Each feedback instance follows a predefined three-paragraph structure: (i) *strengths*, highlighting well-performed aspects; (ii) *areas for improvement*, identifying specific weaknesses; and (iii) an *action plan*, providing concrete recommendations for future improvement.

Each model was fine-tuned using supervised learning on input–output pairs structured according to the three-paragraph feedback schema. **GPT-4.1-mini** was adapted via the OpenAI fine-tuning API for 3 epochs with a learning rate multiplier of 2, consuming 205,239 tokens and reaching a final training loss of 0.934. **Gemini 2.5 Flash** was fine-tuned through Google Vertex AI for 40 cycles, with the optimal checkpoint automatically selected at step 45, achieving a final loss of 0.759 and accuracy of 0.798. **Llama 3.1** was fine-tuned locally via LoRA ($r = 16$, $\alpha = 32$) with 4-bit NF4 quantisation for 3 epochs, reaching a final training loss of 1.210. All three models were qualitatively validated through teacher review prior to deployment.

3.3.3. Post-processing and Validation

After generation, a validation stage ensures quality. The system verifies constraints defined by teachers, such as minimum and maximum feedback length, overall coherence, and the exclusion of restricted terms (e.g., evaluator names or other identifying information). If the generated feedback does not meet these requirements, the module automatically triggers a regeneration process.

3.4. Recording Module

The Recording Module supports reflective learning by capturing students' oral presentations in video and audio format. Recordings are obtained using a standard laptop with an integrated webcam and microphone, relying on Microsoft Teams as the recording platform due to its widespread availability within the institution's infrastructure. Recordings are linked to the corresponding evaluation instances in the database and are made accessible to both students (for self-assessment) and teachers.

4. Conclusions and Future Work

This paper has described the implementation and deployment of AICoFe, an AI-based collaborative feedback system designed to support peer, self-, and teacher assessment in higher education. We have detailed the system's modular architecture, including its multi-LLM feedback generation pipeline, role-specific Learning Analytics dashboards, and feedback history interface with LLM contribution tracking and draft management. Evaluation results confirm that students perceive the AI-assisted feedback as highly coherent and useful, while system usability scores indicate excellent acceptance across both undergraduate and master's levels.

AICoFe is currently being deployed at a larger scale during the ongoing academic year, involving approximately 80 students and 5 teachers across one undergraduate course and one master's course at Universidad Autónoma de Madrid (UAM). This extended deployment will allow a more comprehensive evaluation of the system's impact on peer feedback quality and reflective learning. Future analyses will examine students' perceptions of feedback utility, coherence, and actionability, compare AI-mediated feedback with traditional manual feedback practices where possible, and investigate whether the use of AICoFe is associated with improvements in feedback quality, reflective learning, and presentation performance. Future developments will also explore the integration of multimodal data, including non-verbal cues captured during presentations into the feedback generation process. [28, 20].

Acknowledgements

Support by: Cátedra ENIA UAM-VERIDAS en IA Responsable (NextGenerationEU PRTR TSI-100927-2023-2), M2RAI (PID2024-160053OB-I00, MICIU/FEDER) and SNOLA (RED2022-134284-T). A. Becerra is funded by a predoctoral contract (FPI) from the Comunidad de Madrid (PIPF-2024/COM-34288).

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Gemini in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] N.-F. Liu, D. Carless, Peer feedback: the learning element of peer assessment, *Teaching in Higher Education* 11 (2006) 279–290.
- [2] K. J. Topping, Peer assessment, *Theory into Practice* 48 (2009) 20–27.
- [3] E. V. Popta, M. Kral, G. Camp, R. L. Martens, P. R.-J. Simons, Exploring the value of peer feedback in online learning for the provider, *Educational Research Review* 20 (2017) 24–34.
- [4] Y. Wei, D. Liu, Incorporating peer feedback in academic writing: a systematic review of benefits and challenges, *Frontiers in Psychology* 15 (2024) 1506725.
- [5] O. Viberg, M. Baars, R. F. Mello, N. Weerheim, D. Spikol, C. Bogdan, D. Gasevic, F. Paas, Exploring the nature of peer feedback: An epistemic network analysis approach, *Journal of Computer Assisted Learning* 40 (2024) 2809–2821.
- [6] R. Bodily, T. K. Ikahihifo, B. Mackley, C. R. Graham, The design, development, and implementation of student-facing learning analytics dashboards, *Journal of Computing in Higher Education* 30 (2018) 572–598.
- [7] A. P. Cavalcanti, A. Barbosa, R. Carvalho, F. Freitas, Y.-S. Tsai, D. Gašević, R. F. Mello, Automatic feedback in online learning environments: A systematic literature review, *Computers and Education: Artificial Intelligence* 2 (2021) 100027.
- [8] M. Giannakos, R. Azevedo, P. Brusilovsky, M. Cukurova, Y. Dimitriadis, D. Hernandez-Leo, S. Järvelä, M. Mavrikis, B. Rienties, The promise and challenges of generative AI in education, *Behaviour & Information Technology* (2024) 1–27.
- [9] C. D. Kloos, C. Alario-Hoyos, I. Estévez-Ayres, P. Callejo-Pinardo, M. A. Hombrados-Herrera, P. J. Muñoz-Merino, P. M. Moreno-Marcos, M. Muñoz-Organero, M. B. Ibáñez, How can generative AI support education?, in: *2024 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, 2024, pp. 1–7.
- [10] Z. Zhang, Z. Dong, Y. Shi, T. Price, N. Matsuda, D. Xu, Students' perceptions and preferences of generative artificial intelligence feedback for programming, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 2024, pp. 23250–23258.

- [11] A. Becerra, R. Cobos, Enhancing the professional development of engineering students through an ai-based collaborative feedback system, in: 2025 IEEE Global Engineering Education Conference (EDUCON), IEEE, 2025, pp. 1–9.
- [12] X. Ochoa, F. Dominguez, Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting, *British Journal of Educational Technology* 51 (2020) 1615–1630.
- [13] X. Ochoa, H. Zhao, Openopaf: An open-source multimodal system for automated feedback for oral presentations., *Journal of Learning Analytics* 11 (2024) 224–248.
- [14] I. U. Haq, M. Pifarré, E. Fraca, Novelty evaluation using sentence embedding models in open-ended cocreative problem-solving, *International Journal of Artificial Intelligence in Education* (2024) 1–28.
- [15] P. Topali, R. Cobos, U. Agirre-Uribarren, A. Martínez-Monés, S. Villagrà-Sobrino, 'instructor in action': Co-design and evaluation of human-centred LA-informed feedback in MOOCs, *Journal of Computer Assisted Learning* 40 (2024) 3149–3166.
- [16] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, J. L. Santos, Learning analytics dashboard applications, *American Behavioral Scientist* 57 (2013) 1500–1509.
- [17] M. Navarro, A. Becerra, R. Daza, R. Cobos, A. Morales, J. Fierrez, Vaad: Visual attention analysis dashboard applied to e-learning, in: 2024 International Symposium on Computers in Education (SIIE), IEEE, 2024, pp. 1–6.
- [18] A. Becerra, R. Cobos, Integrating eye-tracking and artificial intelligence for human-centered visual attention analytics in online learning, *IE Comunicaciones: Revista Iberoamericana de Informática Educativa* (2025) 21–32.
- [19] A. Becerra, R. Daza, R. Cobos, A. Morales, M. Cukurova, J. Fierrez, M2lads: A system for generating multimodal learning analytics dashboards, in: 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), IEEE, 2023, pp. 1564–1569.
- [20] A. Becerra, R. Cobos, C. Lang, Enhancing online learning by integrating biosensors and multimodal learning analytics for detecting and predicting student behaviour: a review, *Behaviour & Information Technology* (2025) 1–26.
- [21] J. Steiss, T. Tate, S. Graham, J. Cruz, M. Hebert, J. Wang, Y. Moon, W. Tseng, M. Warschauer, C. B. Olson, Comparing the quality of human and ChatGPT feedback of students' writing, *Learning and Instruction* 91 (2024) 101894.
- [22] T. Wan, Z. Chen, Exploring generative AI assisted feedback writing for students' written responses to a physics conceptual question with prompt engineering and few-shot learning, *Physical Review Physics Education Research* 20 (2024) 010152.
- [23] N. Magyar, S. R. Haley, Balancing learner experience and user experience in a peer feedback web application for MOOCs, in: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–8.
- [24] S. Sajadi, M. Huerta, O. Ryan, K. Drinkwater, Harnessing generative AI to enhance feedback quality in peer evaluations within project-based learning contexts, in: *International Journal of Engineering Education*, 2024.
- [25] C.-L. Yang, A. Uhde, N. Yamashita, H. Kuzuoka, Understanding and supporting peer review using AI-reframed positive summary, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–16.
- [26] A. Becerra, Z. Mohseni, J. Sanz, R. Cobos, A generative ai-based personalized guidance tool for enhancing the feedback to mooc learners, in: 2024 IEEE Global Engineering Education Conference (EDUCON), IEEE, 2024, pp. 1–8.
- [27] A. Becerra, R. Cobos, R. Daza, A multimodal dataset of student oral presentations with sensors and evaluation data, *arXiv preprint arXiv:2601.07576* (2026).
- [28] A. Becerra, D. Andres, P. Villegas, R. Daza, R. Cobos, Mosaic-f: A framework for enhancing students' oral presentation skills through personalized feedback, in: Proceedings of the Learning Analytics Summer Institute Spain 2025 (CEUR Workshop Proceedings, Vol. 4148), 2025, pp. 1–10.