Explainable Feedback for Learning Based on Rubric-Based Multimodal Assessment Analytics with AI

Karsten D. Wolf [1], Fatima Maya [2], and Lisanne Heilmann [1]

Abstract: Providing timely formative feedback to students is very important to support self-regulated learning and deep learning strategies. Feedback has been shown to increase student engagement, satisfaction and learning outcomes, especially in generative learning tasks such as ePortfolios and other forms of multimodal compositions. However, the provision of detailed formative feedback places high demands on teachers' resources. It would be highly beneficial if Large Language Models (LLM) could be used to support the feedback process. Therefore, this paper first describes a general architecture for multimodal formative assessment analysis and feedback generation. It is based on assessment rubrics, which are then used to build task-specific AI analysis pipelines to generate explainable assessment metrics, which are then used to produce helpful feedback. An example of a feedback pipeline for student video submissions in an ePortfolio is given, along with a prompting chain for feedback generation. The paper describes further steps necessary to evaluate and optimise this process in real classroom scenarios..

Keywords: ePortfolios, Explanatory Videos, Multimodal Formative Assessment, Large Language Models (LLM), Prompt Engineering, Few-Shot-Learning, Assessment Rubrics

1 Introduction

The central aspect of studying in higher education is to gain in-depth knowledge, to develop competences, and to acquire a complex understanding of the relevant contexts. To achieve these goals, self-regulated learning (SRL) is a central and necessary skill for academic learning and success [Th21]. Not only does SRL often increases the student's satisfaction with their own academic pathways [Ku14], but it also supports students in their learning process and is correlated with their academic performance [VKB20]. In particular, moments of deeper learning (as conceptualised by [DHO21]) are a prerequisite for successful learning [HD16]. Equal access to such a deep and successful learning can be seen as an imperative foundation of an equal society prepared for 21st century challenges [DHDHB22]. Deeper learning depends on both (and the interaction of) the student's capacity for self-regulated learning and the amount and quality of peer and teacher feedback [DHDHB22]. It therefore seems imperative to support students' deep learning strategies in higher education [AG17, MKRM21].

[1] Universität Bremen, ZeMKI Lab Media & Education, Linzer Str. 6, 28359 Bremen, Germany, wolf@uni-bremen.de, https://orcid.org/0000-0002-4151-840X; lisanne.heilmann@uni-bremen.de, https://orcid.org/0000-0001-9096-9306

[2] Universität Bremen, Faculty 03 AG Information Management, Am Fallturm 1, 28359 Bremen, Germany, famaya@uni-bremen.de, https://orcid.org/0000-0002-9412-392X

In today's globalised world, students face new and changing demands in life and in globalised and digitised labour markets. In order to be able to react appropriately to these circumstances and the demands of the 21st century, new competencies are required [Ma18]. Emerging technologies such as generative AI also require new skills. Expectations of established skills such as (academic) writing are changing as AI models are increasingly able to produce not only linguistically sound texts, but also creative and artistic writing such as poems [Sh23]. The presence of AI itself calls for an intense reflection in education on when and how AI should be integrated into teaching, decision making or everyday life as well as the benefits and risks involved [La22, Ng21]. Critical thinking and critical questioning are becoming key competencies [Al20]. However, deep learning and critical thinking are very difficult to assess and teach [Mi21].

2 Formative Feedback as an Integral Part of ePortfolios to Support Deep Learning

Feedback has been shown to increase students' engagement, satisfaction and outcomes in learning; it plays an important role in SRL by enabling learners to understand their own progress, they can reflect on their performance and gain information on how to develop their understanding and skills in relation to a particular topic; while also gaining a sense of ownership of their own learning processes [Ba10, Ga19, Lu21, Ro18]. This is particularly true, when the tasks involve generative learning strategies [Br21].

One way to support students in their SRL is to integrate ePortfolio assignments into the lesson plan [SH21]. Creating an ePortfolio involves integrating different digital artefacts that appeal to different sensory channels, resulting in multimodal compositions [Pa12]. This could mean incorporating a range of media including text, concept maps, images, paintings, photos, illustrations, sketches, figures, charts, diagrams, memes, presentations, documents, data files, spreadsheets, audio recordings, source code, social media posts, blog entries, selfies, vlogs, videos and more into their ePortfolios [MW24]. This combination of different types of learning experiences as well as output types supports student's engagement with the material and, subsequently, their academic performance [MK23].

EPortfolio assignments support and encourage students to engage with their own learning processes early on in the course of a seminar. However, timely and high quality feedback is required to ensure continued student engagement and to maximise the positive effects of ePortfolios [Fi18, HWH20, Ru05]. Ideally, the feedback should be highly informative, and include information on tasks and processes as well as offer support for the student's self-regulated learning [WZH20]. Accurate and timely feedback is particularly important in terms of their learning experience and motivation [Ba14, HHS02].

Such a formative feedback places high demands on the teaching staff's resources, especially in comparison with shorter and less personalised responses [HM10, Wh05] and for large groups of students [HRP19]. Exploring the ways in which new technologies can assist

and support the provision of high quality formative feedback could benefit both students' academic performance and deeper learning experiences, as well as teachers' resources and their potential to support students' learning in other ways. Precisely because the aim of ePortfolios is to promote critical, active and self-directed engagement with a complex subject, the assessment of these ePortfolios presents a major challenge for AI systems.

At least as far as AI learning is concerned, it has been shown experimentally that different AIs are able to give each other enough productive feedback to improve [Fu23]. A study on multimodal assessment that was supported by machine learning using a combination of drawn models and corresponding text-based descriptions was able to assess the corresponding scientific models [ZHK22]. First forays into AI-generated (or AI-assisted) feedback have shown some potential: A comparison between peer feedback and feedback generated by ChatGPT has shown, that the AI was able to generate helpful und formative feedback measured against the feedback that peers were able to give without significant errors and was concluded to be a "feasible alternative for delivering constructive feedback to a larger number of students" [Ba24].

3 General Architecture for Multimodal Formative Assessment Analytics and Feedback Generation

This work presents the on-going development of a general architecture for AI-based formative multimodal assessment analytics to score complex assessment rubrics by a combination of rule-based and machine learning models, and how this scoring can be used to generate highly informative and explainable feedback. The general design of our proposed architecture represented in Figure 1 follows the general formative assessment cycle of eliciting, interpreting, and responding [St23]. Based upon a didactical plan (see Figure 1, Box 1) students are given specific tasks and problems to solve and document their processes of Self-Regulated Learning in an ePortfolio. In order to enable automated scoring, the rating scales of the rubrics need to be carefully operationalised (see Figure 1, Box 2). Some metrics may be easy to define, such as a certain number of words to be written (word count). Other metrics may be much more complicated, such as quality indicators for an explanatory video created by the student, such as relevant topics, explanation quality or good visualisation of central concepts. Based on the human-readable definitions of performance criteria, indicators, metrics and rating scales, appropriate model types for scoring have to be selected, parameterised or prompted (see Fig. 1, Box 3).

The learners' outcomes (eliciting; see Figure 1, Box 4) such as ePortfolio pages with different multimodal artefacts and reflective comments, are extracted for further analysis (see Figure 1, Box 5). In a second step, on the basis of a pre-defined set of assessment metrics (assessment rubric; see Figure 1, Box 2), the outcomes have to be evaluated (interpreting; see Figure 1, Box 6).

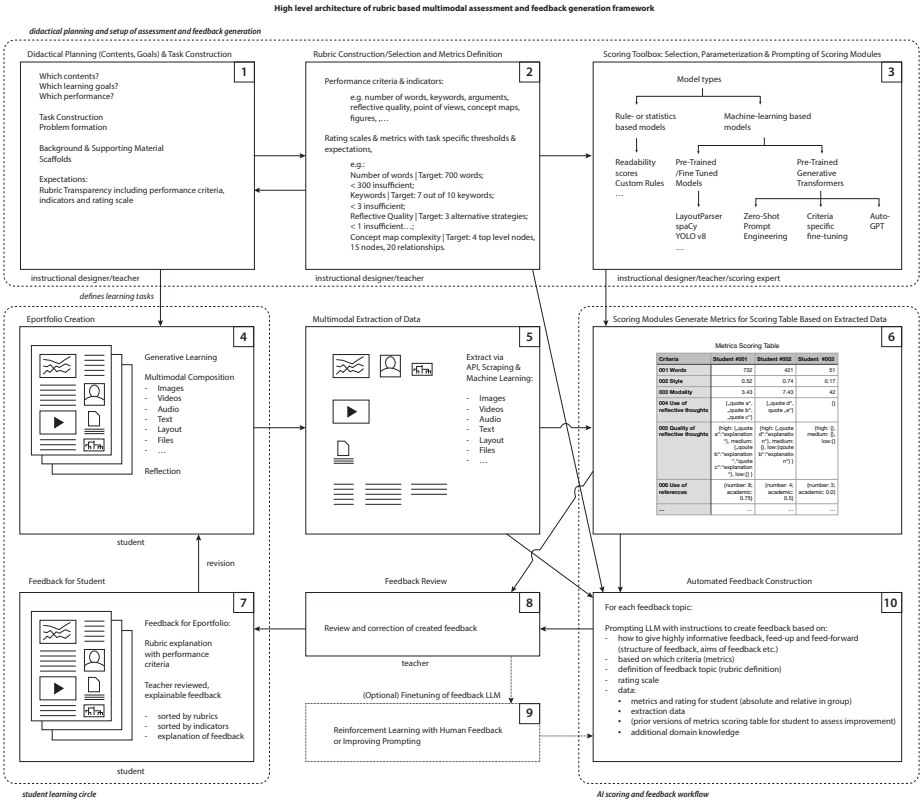

Fig. 1: High level architecture of rubric-based multimodal assessment and feedback generation framework

In a final step, highly informative feedback (responding; see Figure 1, box 10) is generated on the basis of the learner's results and the assessment metrics (see Figure 1, box 6) and made available to the teacher together with the raw metric scores (see Figure 1, box 8), so that the teacher can review, correct and expand the feedback before sending it to the learner (see Figure 1, box 7). Cases of incorrectly assessed metrics will be flagged for further improvement of the assessment modules. Reinforcement learning with human feedback (RLHF) can be used to fine-tune the LLM used for automated feedback generation (see Figure 1, Box 9).

Once all the scoring modules are defined, the extracted data from the ePortfolios can be scored in the metrics scoring table (see Figure 1, Box 6). The scoring table provides the basis for automated feedback generation (see Figure 1, Box 10) and its explainability, which

is necessary for quality review by the teacher (see Figure 1, Box 8) and feedback analysis by the students (see Figure 1, Box 7).

The goal of such a modular architecture instead of a monolithic multimodal LLM is to (a) provide better explainability of scoring and feedback generation [Zh23]; (b) control alignment [Ha23]; (c) control bias [Wa23]; (d) identify hallucinations [Au23, Le23]; and (e) allow the use of a combination of different scoring models, such as rule-based vs. different machine learning models [Li20] or multi-agent conversation [Wu23].

From a technical point of view, the most challenging aspects of such a proposed architectures are (1) multimodal extraction (data basis for assessment scoring), (2) metric scoring (data basis for feedback), and (3) feedback generation based on scoring and extraction data. As an example of further work to be done, this paper describes an early conception of a sub-pipeline within the general architecture described above to generate formative feedback for students' submission of explanatory videos within ePortfolios.

4 Example of Feedback Pipeline for Students' Video Submission in ePortfolios

In the preceding section, we described the concept of a general architecture for the assessment and feedback generation for the various student's artifeacts (texts, illustrations, audio recording, videos etc.) that could be submitted in an ePortfolio. This section aims to demonstrate the feasibility of the proposed architecture in Figure 1, through the design of an end-to-end sub-pipeline for the analysis of video submissions in e-Portfolios and feedback generation based on rubrics. An overview of the pipeline is represented in Figure 2. Following the submission by the student, the video is fetched from the ePortfolio for further processing. The pipeline is composed of three separate sub-pipelines, each of which is responsible for analysing a specific modality of the video: audio, transcript and scenes.

The FFmpeg [3] python wrapper is used to generate the audio file. To assess the quality of the speech, namely the clarity and fluency of the speaker, we chose the following features: speech rate, given by the number of words or syllables spoken per minute; articulation rate, defined by the number of words or syllables spoken per minute, excluding pauses, and average syllable duration (ASD), which is the average time taken to pronounce a syllable. Praat [BW] script was employed to extract these metrics, which they were then stored in a JSON-format.

The second component of the pipeline is designed to analyse the video's content. The extracted audio file is then transcribed using Whisper [Ra23], a state-of-the-art speech recognition system to generate spoken text. The text is transmitted together with the teacher's evaluation rubrics to the Llama 3 70B [4], a recently released large language model (LLM).

[3] https://github.com/kkroening/ffmpeg-python
[4] https://huggingface.co/meta-llama/Meta-Llama-3-8B

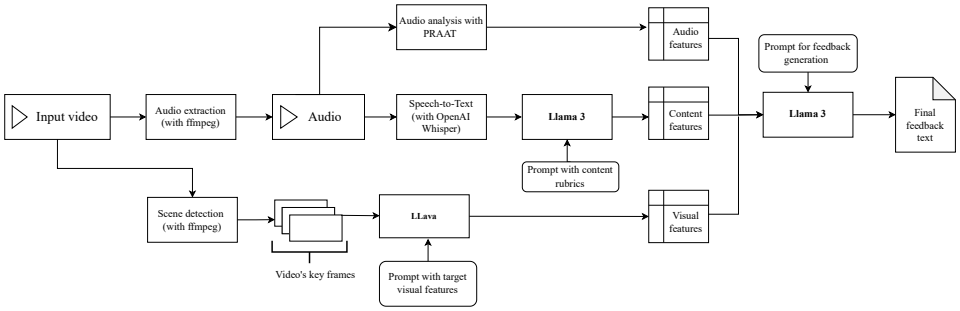

Fig. 2: End-to-end feedback system architecture for the multimodal assessment of explanatory videos in ePortfolios

During the implementation process, a set of prompts were tested and evaluated with the teachers in order to generate the final prompt, which is depicted in Figure 3. Note that, to guide the model to produce the desired output as well as facilitating in context-learning and thus improving the model's accuracy on this specific task, we use the technique of few-shot prompting by providing multiple text snippets with the desired output. The LLM's output (content features) is then parsed in a JSON-format.

For the visual content, we first use PySceneDetect [5], a tool for detecting shot changes in videos, which returns a set of key frames where major changes occur. This reduces the size of the video and allows for frame-based analysis. The key frames are then processed together with a prompt (see Figure 3, middle box) containing target visual elements using LLaVA, a large language vision model [Li23]. These specific quality criteria are inspired by Mayer's principles of multimedia learning [Ma17], which emphasize that text and images should be conveyed together for better learning. In addition, images should support and clarify the meaning. Similar to the previous subsystems, the output is parsed in JSON format. Finally, Llama 3 is used to generate the final feedback. Similar to the generation of content features, we use the few-shot prompting technique, feeding the model with examples of feedback texts related to the generated features from the different modalities. We then feed the model with the prompt and the resulting JSON from the sub-pipelines. The final output is a feedback text organized into rubrics with the corresponding feedback snippets.

Some of the lessons we have learnt during this process are that well-defined prompts based on the rubrics with clear examples of the expected output help to produce better LLM responses with improved consistency. However, this requires iterative optimisation with the help of teachers.

[5] https://github.com/Breakthrough/PySceneDetect/tree/main

Content Prompt

<system>
You are a teacher who asses student's video transcript and response in a JSON format.

<user>
Evaluate this text <text> based on these rubrics <rubrics>. For each rubric, answer
with fulfilled, partially fulfilled or not fulfilled, providing the justification from the text content.
Anwer in this JSON format <json format>.

Prompt LLaVA

Analyse this video frame and identify its key elements such text blocks; images; diagrams; tables. Additionally, in case the image contain text and am image or graphic, determine wether these are related to the text content or they are just decorative.

Feedback Prompt

<system>
You are a teacher who gives feedback to students based on their submitted video analytics .

<user>
Based on the video analytics <video analytics json>, teacher's rubrics, defined rules for visual features <visual rules> and defined rules for audio features <audio rules>, write a feedback text to the student covering what is fulfilled, partially fulfilled or not fullfilled and propose improvement suggestions.

Fig. 3: Left – prompt for content features extraction used with Llama 3; middle – Prompt used for visual features extraction with LLaVA; right - prompt for feedback generation with Llama 3.

5 Conclusion and Future Work

This paper highlighted the importance of integrating ePortfolios in learning settings, allowing students to creatively incorporate diverse artefacts and thereby support deeper learning. However, the success of ePortfolios relies on the provision of timely and high-quality feedback, which places a significant demand on teaching staff resources, especially for large student groups. While previous work has mainly focused on automating the assessment of text-based artefacts, this work takes a step further to assess more complex artefacts through proposing a high level architecture for a mutimodal assessment and feedback system for diverse artefacts in ePortfolios. The system leverages rubric-based assessment through deploying various AI-models (LLMs, computer vision, multimodal LLMs, NLP etc.) to generate assessment analytics, which are then used to produce formative feedback. We also presented an example of a sub-pipeline from the general architecture designed to assess explanatory videos by evaluating the quality of the three modalities: audio, content, and visuals. As this work is still in its early stages, our next steps will involve deploying the proposed sub-pipeline for assessing videos in a real scenario, where the generated feedback texts will be evaluated using inter-rater reliability assessments between system and human ratings. Additionally, we will conduct a survey to collect student feedback on the AI-generated feedback for further improvement of the pipeline. In the long term, similar to the video assessment scenario, different sub-pipelines for the various modalities will be iteratively implemented and evaluated using human raters and student feedback.

Bibliography

[AG17] Asikainen, Henna; Gijbels, David: Do Students Develop Towards More Deep Approaches to Learning During Studies? A Systematic Review on the Development of Students' Deep and Surface Approaches to Learning in Higher Education. Educational Psychology Review, 29(2):205–234, June 2017.

[Al20] Alsaleh, Nada J.: Teaching Critical Thinking Skills. Literature Review. Turkish Online Journal of Educational Technology, 19(1):21–39, 2020.

[Au23] Augenstein, Isabelle; Baldwin, Timothy; Cha, Meeyoung; Chakraborty, Tanmoy; Ciampaglia, Giovanni Luca; Corney, David; DiResta, Renee; Ferrara, Emilio; Hale, Scott; Halevy, Alon; Hovy, Eduard; Ji, Heng; Menczer, Filippo; Miguez, Ruben; Nakov, Preslav; Scheufele, Dietram; Sharma, Shivam; Zagni, Giovanni: Factuality Challenges in the Era of Large Language Models, 2023.

[Ba10] Baeten, Marlies; Kyndt, Eva; Struyven, Katrien; Dochy, Filip: Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. Educational Research Review, 5(3):243–260, January 2010.

[Ba14] Bayerlein, Leopold: Students' feedback preferences: how do students react to timely and automatically generated assessment feedback? Assessment & Evaluation in Higher Education, 39(8):916–931, November 2014.

[Ba24] Banihashem, Seyyed Kazem; Kerman, Nafiseh Taghizadeh; Noroozi, Omid; Moon, Jewoong; Drachsler, Hendrik: Feedback sources in essay writing: peer-generated or AI-generated feedback? International Journal of Educational Technology in Higher Education, 21(1):23, April 2024.

[Br21] Brod, Garvin: Generative Learning: Which Strategies for What Age? Educational Psychology Review, 33(4):1295–1318, December 2021.

[BW] Boersma, Paul; Weenink, David: Praat: doing phonetics by computer. Place: Amsterdam.

[DHDHB22] Darling-Hammond, Kia; Darling-Hammond, Linda; Byard, Eliza: The civil rights road to deeper learning: five essentials for equity. Multicultural education series. Teachers College Press, New York, 2022.

[DHO21] Darling-Hammond, Linda; Oakes, Jeannie: Preparing Teachers for Deeper Learning. Harvard Education Press, February 2021.

[Fi18] Filius, Renée Marianne; Kleijn, Renske A. M. de; Uijl, Sabine G.; Prins, Frans J.; Rijen, Harold V. M.; Grobbee, Diederick E.: Promoting deep learning through online feedback in SPOCs. Frontline Learning Research, 6(2):92–113, November 2018.

[Fu23] Fu, Yao; Peng, Hao; Khot, Tushar; Lapata, Mirella: Improving Language Model Negotiation with Self-Play and In-Context Learning from AI Feedback. 2023.

[Ga19] Garcia, Elaine; Moizer, Jonathan; Wilkins, Stephen; Haddoud, Mohamed Yacine: Student learning in higher education through blogging in the classroom. Computers & Education, 136:61–74, July 2019.

[Ha23] Haendler, Thorsten: Balancing Autonomy and Alignment: A Multi-Dimensional Taxonomy for Autonomous LLM-powered Multi-Agent Architectures, 2023.

[HD16] Hattie, John A. C.; Donoghue, Gregory M.: Learning strategies: a synthesis and conceptual model. npj Science of Learning, 1(1):1–13, August 2016.

[HHS02] Higgins, Richard; Hartley, Peter; Skelton, Alan: The Conscientious Consumer: Reconsidering the role of assessment feedback in student learning. Studies in Higher Education, 27(1):53–64, February 2002.

[HM10] Hourigan, Tríona; Murray, Liam: Using blogs to help language students to develop reflective learning strategies: Towards a pedagogical framework. Australasian Journal of Educational Technology, 26(2), Apr. 2010.

[HRP19] Henderson, Michael; Ryan, Tracii; Phillips, Michael: The challenges of feedback in higher education. Assessment & Evaluation in Higher Education, 44(8):1237–1252, November 2019.

[HWH20] Haughney, Kathryn; Wakeman, Shawnee; Hart, Laura: Quality of Feedback in Higher Education: A Review of Literature. Education Sciences, 10(3):60, March 2020.

[Ku14] Kuo, Yu-Chun; Walker, Andrew E.; Schroder, Kerstin E.E.; Belland, Brian R.: Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. The Internet and Higher Education, 20:35–50, January 2014.

[La22] Laupichler, Matthias Carl; Aster, Alexandra; Schirch, Jana; Raupach, Tobias: Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and Education: Artificial Intelligence, 3:100101, 2022.

[Le23] Leiser, Florian; Eckhardt, Sven; Knaeble, Merlin; Maedche, Alexander; Schwabe, Gerhard; Sunyaev, Ali: From ChatGPT to FactGPT: A Participatory Design Study to Mitigate the Effects of Large Language Model Hallucinations on Users. In: Proceedings of Mensch Und Computer 2023. MuC '23, Association for Computing Machinery, New York, NY, USA, p. 81–90, 2023.

[Li20] Likmeta, Amarildo; Metelli, Alberto Maria; Tirinzoni, Andrea; Giol, Riccardo; Restelli, Marcello; Romano, Danilo: Combining reinforcement learning with rule-based controllers for transparent and general decision-making in autonomous driving. Robotics Auton. Syst., 131:103568, 2020.

[Li23] Liu, Haotian; Li, Chunyuan; Wu, Qingyang; Lee, Yong Jae: Visual Instruction Tuning. Advances in Neural Information Processing Systems, 36:34892–34916, December 2023.

[Lu21] Lu, Hwangji: Electronic Portfolios in Higher Education: A Review of the Literature. European Journal of Education and Pedagogy, 2(3):96–101, July 2021.

[Ma17] Mayer, R.e.: Using multimedia for e-learning. Journal of Computer Assisted Learning, 33(5):403–423, 2017.

[Ma18] Martin, John P.: Skills for the 21st Century: Findings and Policy Lessons from the OECD Survey of Adult Skills. IZA Policy Paper 138. IZA, 2018.

[Mi21] Miterianifa, Miterianifa; Ashadi, Ashadi; Saputro, Sulistyo; Suciati, Suciati: Higher Order Thinking Skills in the 21st Century: Critical Thinking. January 2021.

[MK23] Mathias, Brian; Kriegstein, Katharina von: Enriched learning: behavior, brain, and computation. Trends in Cognitive Sciences, 27(1):81–97, January 2023.

[MKRM21] Mthethwa-Kunene, Khetsiwe; Rugube, Talent; Maphosa, Cosmas: Rethinking Pedagogy: Interrogating Ways of Promoting Deeper Learning in Higher Education. European Journal of Interactive Multimedia and Education, 3(1):e02204, December 2021.

[MW24] Maya, Fatima; Wolf, Karsten D.: An Architecture for Formative Assessment Analytics of Multimodal Artefacts in ePortfolios Supported by Artificial Intelligence. In (Sahin, Muhittin; Ifenthaler, Dirk, eds): Assessment Analytics in Education, pp. 293–312. Springer International Publishing, Cham, 2024.

[Ng21] Ng, Davy Tsz Kit; Leung, Jac Ka Lok; Chu, Samuel Kai Wah; Qiao, Maggie Shen: Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2:100041, 2021.

[Pa12] Palmeri, Jason: Remixing Composition: A History of Multimodal Writing Pedagogy. Southern Illinois University Press, March 2012.

[Ra23] Radford, Alec; Kim, Jong Wook; Xu, Tao; Brockman, Greg; McLeavey, Christine; Sutskever, Ilya: Robust speech recognition via large-scale weak supervision. In: Proceedings of the 40th International Conference on Machine Learning. volume 202 of ICML'23, JMLR.org, Honolulu, Hawaii, USA, pp. 28492–28518, July 2023.

[Ro18] Roehling, Patricia V.: Flipping the College Classroom. Springer International Publishing, Cham, 2018.

[Ru05] Rushton, Alison: Formative assessment: a key to deep learning? Medical Teacher, 27(6):509–513, September 2005.

[SH21] Segaran, Meerita Kunna; Hasim, Zuwati: SELF-REGULATED LEARNING THROUGH ePORTFOLIO: A META-ANALYSIS. Malaysian Journal of Learning and Instruction, 18(1):131–156, January 2021.

[Sh23] Shidiq, Muhammad: The Use of Artificial Intelligence-based Chat-GPT and its Challenges for the World of Education; from the Viewpoint of the Development of Creative Writing Skills. Proceedings ofInternational Conference on Education, Society and Humanity, 1(1):353–357, 2023.

[St23] Stanja, Judith; Gritz, Wolfgang; Krugel, Johannes; Hoppe, Anett; Dannemann, Sarah: Formative assessment strategies for students' conceptions—The potential of learning analytics. British Journal of Educational Technology, 54(1):58–75, 2023. Place: United Kingdom Publisher: Wiley-Blackwell Publishing Ltd.

[Th21] Theobald, Maria: Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. Contemporary Educational Psychology, 66:101976, July 2021.

[VKB20] Viberg, Olga; Khalil, Mohammad; Baars, Martine: Self-regulated learning and learning analytics in online learning environments: a review of empirical research. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. ACM, Frankfurt Germany, pp. 524–533, March 2020.

[Wa23] Wan, Yixin; Pu, George; Sun, Jiao; Garimella, Aparna; Chang, Kai-Wei; Peng, Nanyun: "Kelly is a Warm Person, Joseph is a Role Model": Gender Biases in LLM-Generated Reference Letters. In (Bouamor, Houda; Pino, Juan; Bali, Kalika, eds): Findings of the Association for Computational Linguistics: EMNLP 2023. Association for Computational Linguistics, Singapore, pp. 3730–3748, December 2023.

[Wh05] White, Edward M.: The Scoring of Writing Portfolios: Phase 2. College Composition and Communication, 56(4):581–600, 2005.

[Wu23] Wu, Qingyun; Bansal, Gagan; Zhang, Jieyu; Wu, Yiran; Li, Beibin; Zhu, Erkang; Jiang, Li; Zhang, Xiaoyun; Zhang, Shaokun; Liu, Jiale; Awadallah, Ahmed Hassan; White, Ryen W; Burger, Doug; Wang, Chi: AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation, 2023.

[WZH20] Wisniewski, Benedikt; Zierer, Klaus; Hattie, John: The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. Frontiers in Psychology, 10, January 2020.

[Zh23] Zhao, Haiyan; Chen, Hanjie; Yang, Fan; Liu, Ninghao; Deng, Huiqi; Cai, Hengyi; Wang, Shuaiqiang; Yin, Dawei; Du, Mengnan: Explainability for Large Language Models: A Survey, 2023.

[ZHK22] Zhai, Xiaoming; He, Peng; Krajcik, Joseph: Applying machine learning to automatically assess scientific models. Journal of Research in Science Teaching, 59(10):1765–1794, 2022.