BERA 2025

9 September 2025 to 11 September 2025University of Sussex

Exploring the feasibility of using Generative AI to emulate teacher feedback for GCSE English Language assessments

Claire Tupling, Lucy Howarth, Anagha Varma, Katie Gilmartin

AQA, United Kingdom

Abstract

Generative artificial intelligence (AI) presents researchers with responsibilities of understanding potential educational uses beyond much-discussed scenarios of plagiarism and other forms of cheating (Selwyn, 2024). Use cases are needed to provide educators and researchers with an understanding of how AI might be used to support learning, teaching and assessment.

AQA, Imperial College London and the English Department of an independent school in England collaborated to explore whether generative AI could provide useful feedback for GCSE level English assessments. The research team used prompt engineering of a Large Language Model (LLM) and generative AI with the aim of delivering rapid, detailed and personalised feedback to descriptive English Language tasks by Year 10 students. Students submitted assessments to 'Lambda Feedback', an assessment portal developed by colleagues at Imperial College London. In designing this research, we view AI as part of a sociotechnical system where computational artefacts interact with humans in social and organisational contexts (Baxter and Sommerville, 2011). Viewing AI as sociotechnical enables us to understand how the feedback generated in this study is a result of the interaction between researchers, teachers, and students (Johnson and Verdicchio, 2024).

Researchers worked with the school's English teachers to understand the detail, quality and structure of the feedback they routinely provide to their students. This identified two key focus areas that the feedback would cover: content & organisational skills and technical accuracy. Researchers used this knowledge to develop and refine prompts for the LLM, enabling it to provide feedback. A key aim of the research was to emulate teacher tone, and we attempted to achieve this by giving the LLM a character, so it sounded like a teacher. We sought teacher input throughout the process, identifying and refining focus areas. These included using prompts to ask open-ended questions to expand students' imagination, ensuring a distinction between narrative and descriptive writing.

Researchers supported c.110 year 10 students and their teachers during English lessons as they inputted assessments on to 'Lambda Feedback' and received an AI-generated response. To explore the extent to which the AI feedback emulated what the students would have expected to receive from their teacher, we took a two-stage approach to evaluation. This involved carrying out focus groups with teachers, as well as requesting students to complete a survey about their thoughts on the feedback within the platform. Data analysis and evaluation focused on user experience, comprehensibility of the feedback and alignment of the AI and teacher feedback.

Most students found the feedback helpful and easy to understand. However, some struggled to understand AI-generated feedback due to its use of sophisticated language, while other students found the feedback too lengthy. This suggests further work is needed to effectively mimic teacher tone. Many suggested simplifying the language to make it more age-appropriate, as well as adding summaries and visuals to improve accessibility. Overall, students commented that the feedback made them feel positive and had increased their confidence.

Teachers found the AI feedback to be helpful and detailed, often exceeding what they could provide themselves. However, the feedback was frequently too long, thus overwhelming students and sometimes failing to address key strengths or weaknesses in their work. Teachers commented that some students dismissed AI feedback, seeing it as less valuable than teacher input, especially in subjective areas like creative writing. Teachers commented that some SEND students may struggle with its length and complexity, while higher-ability students may find it too basic. Teachers agreed they would use the AI feedback as a supplemental tool but emphasised the need for human supervision of students when inputting work.

In this paper researchers from AQA will explore how the outcomes from this project align with broader educational research on AI for assessment. There are common touchpoints around subject-matter knowledge, mark justification, trainability, data-processing capabilities and ethical use. The findings from this research are valuable when defining what helpful feedback for future learning looks like (Winstone, et al, 2016).

References

- Baxter G, Sommerville I (2011) 'Socio-technical systems: from design methods to systems engineering', Interacting with Computers, 23(1):4–17

- Johnson, D.G. and Verdicchio, M., (2024) 'The sociotechnical entanglement of AI and values', AI & Society, http://dx.doi.org/10.1007/s00146-023-01852-5
- Selwyn, N., (2024) 'On the limits of artificial intelligence (AI) in education' Nordisk tidsskrift for pedagogikk og kritikk, 10(1), pp.3-14
- Winstone, N. E.; Nash, R. A.; Parker, M. & Rowntree, J. (2016) 'Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes', Educational Psychologist, 52(1), pp. 17–37