

## **Peer Review Activity**

**Paper 1: Can AI give good feedback on essay-type assignments? An explorative case study of LLMs in higher education (De Wet, Da Silva & Bohnsack, 2025)**

**Familiarise yourself with the purpose, problem, objective or research question of each paper. Are they in line with your experience or thoughts on the topic, contributing to the collective body of knowledge in this area?**

De Wet, Da Silva & Bohnsack's (2025) study adds to the body of knowledge of the use of ChatGPT in education. The main findings – that students prefer a mix of human and LLM-based feedback, as opposed to human-only or LLM-only – are unexpected, personally, as I thought that students would prefer what I thought would be the more nuanced, focussed human-only feedback. I believe these findings to be insightful and so contribute to the collective body of knowledge in this area.

**Is the research methodology utilised in each paper appropriate for the stated purpose or question?**

De Wet, Da Silva & Bohnsack (2025) chose a mixed-method approach for their research – a quantitative survey and qualitative open-ended questions regarding This approach was able to ascertain preferences of feedback, however the study used OpenAI's ChatGPT-3.5 for the master's courses and ChatGPT-4 for the undergraduate course. In relation to these choices, I believe that the research question – 'Can AI give good feedback on essay-type assignments? An explorative case study of LLMs in higher education' - was too broad as only one LLM model is being tested here (ChatGPT). A more appropriate title would have been 'Can ChatGPT give good feedback on essay-type assignments? An explorative case study of an LLM in higher education'. However, I believe that the researchers should have used the same version of ChatGPT to increase reliability of results.

Students were not told which feedback was LLM-generated. I agree with this approach, as I believe that pre-conceptions about LLMs could affect ratings of usefulness.

**In terms of data collection and analysis, is this also appropriate for the stated purpose or question? (We will discuss this further in upcoming units.)**

The data collection and analysis was appropriate as the open-ended questions regarding feedback added insight into reasoning behind preferences and the use of

quantitative methods in the form of scored survey questions using a 4-point Likert scale allowed for further numerical data analysis.

**Does each paper support its claims and conclusions with explicit arguments or evidence?**

De Wet, Da Silva & Bohnsack (2025) found that the LLM feedback (provided by various models of Chat GPT) tended to include repetitive and formulaic language and sometimes illogical output. They also found that, although it often took repeated feedback cycles for students to identify that their feedback was provided by an LLM, when they did discover this, their confidence in the feedback tended to decrease. However, overall students found LLM feedback helpful: 78% rated it as helpful for task improvement; 73% rated it positively for logical structure; 75% rated it as easy to follow.

Most students (64%) preferred a mixed feedback format of human and LLM feedback, whereas 32% preferred human-only and just 4% preferred exclusively LLM feedback.

**How would you enhance the work/paper?**

I would enhance De Wet, Da Silva & Bohnsack's (2025) paper by using one AI model – the latest available, changing the research question to reflect this, as stated above. I would also have sought to gain an equal spread of students across course levels.

**Paper 2: Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT (Jeon & Lee, 2023)**

**Familiarise yourself with the purpose, problem, objective or research question of each paper. Are they in line with your experience or thoughts on the topic, contributing to the collective body of knowledge in this area?**

The findings – that ChatGPT seems to be used best in complementing teaching practice, rather than a replacement for teaching – are in line with my own experience of using ChatGPT in teaching. They appear to add to the body of knowledge in terms of categorising different uses, according to role played (Interlocutor, Content Provider, Teaching Assistant, and Evaluator).

**Is the research methodology utilised in each paper appropriate for the stated purpose or question?**

Jeon & Lee (2023)'s qualitative approach allowed for some insights into teachers' different uses of ChatGPT within education. However, although they chose teachers with a good range of teaching experience, I feel that the age range used wasn't varied enough and only 2/11 participants were males. Furthermore, the fact that the teachers had not used ChatGPT in their teaching prior to this study limits the insight into how teachers may be using the technology in classrooms. Finally, the fact that the study only looks at elementary English language teachers in Korea further limits the levels of insight to be gained.

I feel that a follow-up, mixed qualitative and quantitative study may better address the research purpose, surveying teachers with prior experience in integrating ChatGPT into their practice; the questions should ask which of the study's uses of the LLM they employed, with an extra section asking how else these experienced teachers used ChatGPT.

**In terms of data collection and analysis, is this also appropriate for the stated purpose or question?**

The research used qualitative individual semi-structured interviews and gained additional insight through use of the teachers' chatbot-use logs. The qualitative analysis of the interview data through a coding scheme identified recurring patterns and emergent themes, with a focus on specific applications of ChatGPT and teacher roles. These methods were appropriate for discovering how the teachers found use of ChatGPT within their recent exploration of the technology and went some way to fulfilling the paper's purpose, however, as previously stated, I believe that more widespread data, including that gained from teachers who have prior experience of using ChatGPT in the classroom, is needed to fully address the purpose.

**Does each paper support its claims and conclusions with explicit arguments or evidence?**

The study found that their teachers had a dynamic and complementary relationship with ChatGPT, wherein teachers were required to orchestrate different resources with ChatGPT, ensure students were active and not passive in their interactions with the technology, and raise ethical awareness regarding its use. It is suggested that teachers should be trained in line with these findings.

The findings are based on the patterns and themes extracted from the interview analysis. They found the following roles that ChatGPT can play in the classroom:

- **Interlocutor:** ChatGPT as a role-player or interactive game partner.
- **Content Provider:** ChatGPT facilitates the production and recommendation of materials, customization of materials, and provision of cultural knowledge.
- **Teaching Assistant:** ChatGPT assists by functioning as a grammar checker or online dictionary or background knowledge activator to preview topics.
- **Evaluator:** ChatGPT can provide initial grading of students' writing and produce testing materials.

### **How would you enhance the work/paper?**

As alluded to previously, to enhance this paper, I would seek to survey teachers who already have prior experience in using ChatGPT in the classroom, so as to uncover additional uses of the technology that may have developed organically over time (as opposed to more short-term uses that the studied teachers developed under pressure). I would also ensure that a wider age bracket of teacher is accounted for, while ensuring that males and females were equally represented. Finally, a focus on multiple subjects, spanning the arts and sciences, would give further insight into how ChatGPT is complementing education practice.

### **References**

De Wet, M., Da Silva, M.O. and Bohnsack, R. (2025) 'Can AI give good feedback on essay-type assignments? An explorative case study of LLMs in higher education', *Innovations in Education and Teaching International*, pp.1-16. Available at: 10.1080/14703297.2025.2536609 (Accessed 14 August 2025).

Jeon, J. and Lee, S. (2023) 'Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT', *Education and Information Technologies*, 28, pp. 15873–15892. Available at: 10.1007/s10639-023-11834-1 (Accessed 14 August 2025).