Automated Feedback vs. Human Feedback: A Study on AI-Driven Language Assessment

Mohammad Aliakbari[1], Pooria Barzan[1*], Mehran Sayyadi[2]

[1] Department of English Language and Literature, Faculty of Humanities, Ilam University, Ilam, Iran
[2] Department of English Language and Literature, Faculty of Humanities, Ilam University, Iran

* Corresponding author email address: pooriabarzan@yahoo.com

1. Round 1

1.1 Reviewer 1

Reviewer:

The sentence "AI offers scalable, immediate, and data-driven solutions..." is compelling but too general. Consider briefly exemplifying one AI tool (e.g., ChatGPT or Write & Improve) that is widely recognized in the field.

The sentence "These systems deliver rapid, personalized feedback..." conflates multiple capabilities. Please distinguish clearly between feedback speed, personalization mechanisms, and linguistic domains (grammar, vocabulary, etc.).

The limitation regarding "socio-pragmatic nuances" is mentioned but could benefit from a concrete example (e.g., irony, sarcasm, or politeness strategies).

The sentence "Human feedback retains a critical edge…" could be strengthened by citing specific research that quantifies human feedback advantages in pronunciation.

The manuscript states that AI "struggles to interpret socio-pragmatic nuances," but fails to mention progress in affective computing or emotion recognition. Consider briefly acknowledging these advancements.

While AI literacy is discussed as essential, there is no mention of current teacher training programs or frameworks. Please add examples of initiatives or models for developing AI literacy in educators.

The use of the Oxford Placement Test is noted, but the version (e.g., paper-based or online) and scoring criteria are missing. Include this information to enhance reproducibility.

The description of ChatGPT's feedback lacks technical details. Specify which version of ChatGPT was used and whether prompts were standardized or adapted per learner.

You mention a "standardized rubric," but do not provide it. Please consider including the rubric in an appendix or describing its main criteria (e.g., accuracy, fluency, cohesion).

The phrase "learners' appreciation for personalized guidance" is supported by survey results, but no quotes are included. Integrate at least one illustrative quote to substantiate the theme.

Author revised the manuscript and uploaded the updated document.

1.2   Reviewer 2

Reviewer:

You mention "labor intensity" and "bias" in human feedback but do not elaborate. Please briefly describe types of biases (e.g., halo effect, cultural bias) to enhance the argument.

The term "constructivist and sociocultural theories" is invoked without citation or elaboration. Please reference foundational theorists (e.g., Vygotsky) or specific principles such as the Zone of Proximal Development.

The phrase "consistent error detection" could be misleading. Please clarify whether this consistency applies across proficiency levels and languages, as some AI tools underperform with non-standard input.

Standard deviations are provided, but confidence intervals would strengthen the statistical interpretation. Consider adding these or clarifying if assumptions for t-tests (e.g., normality) were tested.

Themes are described, but the methodology for thematic analysis is underdeveloped. Indicate whether coding was inductive or deductive and whether inter-rater reliability was assessed.

You state that "ChatGPT's real-time feedback" was effective, yet spoken task feedback was delivered within 24 hours. Please clarify this discrepancy in timing.

Author revised the manuscript and uploaded the updated document.

2.   Revised

Editor's decision after revisions: Accepted.
Editor in Chief's decision: Accepted.