

Review Form 3

Journal Name:	Asian Journal of Education and Social Studies
Manuscript Number:	Ms_AJESS_125085
Title of the Manuscript:	AI Versus Human Graders: Assessing the Role of Large Language Models in Higher Education
Type of the Article	

[Review Form 3](#)

PART 1: Review Comments

Compulsory REVISION comments	Reviewer's comment	Author's Feedback <i>(Please correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)</i>
<p>Please write a few sentences regarding the importance of this manuscript for the scientific community. Why do you like (or dislike) this manuscript? A minimum of 3-4 sentences may be required for this part.</p>	<p>Relevance and Timeliness: The paper addresses a critical issue in education as AI becomes more pervasive. The comparison between AI grading and human grading touches on the efficiency, scalability, and fairness of automated systems in educational assessment.</p> <p>Clear Objectives: The study's objectives are well-defined, aiming to assess the viability of LLMs as grading tools in higher education and to explore any potential biases that may arise from AI-based assessments.</p> <p>Comprehensive Literature Review: The manuscript demonstrates a good understanding of the existing body of research, providing a solid background on the use of AI in education, including prior studies on automated grading systems.</p> <p>Use of Quantitative and Qualitative Approaches: The research methodology uses both quantitative data (accuracy and grading consistency metrics) and qualitative analysis (examining biases and fairness), offering a well-rounded approach to the topic.</p>	
<p>Is the title of the article suitable? (If not please suggest an alternative title)</p>	<p>OK</p>	
<p>Is the abstract of the article comprehensive? Do you suggest the addition (or deletion) of some points in this section? Please write your suggestions here.</p>	<p>OK</p>	
<p>Are subsections and structure of the manuscript appropriate?</p>	<p>OK</p>	
<p>Please write a few sentences regarding the scientific correctness of this manuscript. Why do you think that this manuscript is scientifically robust and technically sound? A minimum of 3-4 sentences may be required for this part.</p>	<p>The manuscript is generally scientifically robust and technically sound, but with some limitations that need addressing:</p> <p>Methodological Soundness: The study employs a clear methodology, using datasets from real-world student assessments to test the performance of LLMs. However, the sample size of assignments and grading tasks needs to be discussed more thoroughly to ensure the results are statistically significant.</p> <p>Bias Evaluation: The paper raises important concerns about potential bias in AI grading but lacks a deep exploration of the nature of such biases. There should be more detail on how bias was measured and controlled during the study, as well as examples of specific instances where bias was detected in AI grading compared to human grading.</p> <p>Control Groups and Comparative Analysis: While the comparison between AI and human graders is an essential part of the research, the study could benefit from a more detailed explanation of the control mechanisms used to ensure fairness in the comparison. Were the human graders selected based on their expertise or familiarity with AI? Additionally, providing detailed statistics on where AI performed significantly better or worse than human graders would enhance the study's credibility.</p>	
<p>Are the references sufficient and recent? If you have suggestions of additional references, please mention them in the review form.</p>	<p>OK</p>	

Review Form 3

Minor REVISION comments Is the language/English quality of the article suitable for scholarly communications?	The paper is generally well-written, with a clear structure and logical flow. The introduction effectively sets the stage for the research questions.	
Optional/General comments	The manuscript presents an intriguing and timely investigation into the comparison between AI-powered large language models (LLMs) and human graders in higher education settings. The topic is highly relevant given the increasing integration of artificial intelligence (AI) in education, especially in automating tasks like grading. The paper attempts to evaluate the potential of LLMs to perform grading tasks efficiently and reliably compared to traditional human graders, considering factors such as accuracy, fairness, and bias.	

PART 2:

	Reviewer's comment	Author's comment (if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)
Are there ethical issues in this manuscript?	<i>(If yes, Kindly please write down the ethical issues here in details)</i>	

Reviewer Details:

Name:	James Boogere
Department, University & Country	Gulu University, Uganda