Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Volume 16. Number 3, September 2025 Pp.175- 187
DOI: https://dx.doi.org/10.24093/awej/vol16no3.10

A Comparative Study of Human and AI-Assisted Assessment using ChatGPT: The Case of Moroccan EFL Learners

Hajar Ennadir
Department of English, Faculty of Letters and Humanities Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fes, Morocco
Corresponding Author: hajar.ennadir@usmba.ac.ma

Salma Habiballah
Department of English, Faculty of Letters and Humanities Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fes, Morocco
Email : salma.habiballah@usmba.ac.ma

Received:06/02/2025 Accepted: 09/12/2025 Published:09/20/2025

Abstract
Artificial intelligence (AI) is capable of performing various tasks and has been applied across multiple fields, including higher education. AI is increasingly used in teaching, learning, and classroom assessment. The primary aim of this study is to compare human assessments and AI-assisted assessments in Moroccan higher education, focusing on summative e-assessments. The primary question of the study is: How does AI-assisted assessment differ from human assessment in evaluating assignments in Moroccan higher education? The study yields data specific to the Moroccan context, filling the gap in this concern; it also provides educators and policymakers with a clear view of the similarities and differences between human and AI-assisted assessment. Writing assignments from EFL classes in two departments at the Faculty of Letters and Humanities, Dhar El Mahraz, were graded by their respective instructors and ChatGPT-4. The results of this study showed that AI grading produced higher scores than human grading, with differences that are both statistically significant in the department of French language studies, while in the department of psychology, the differences between human grading and AI grading are neither statistically significant nor practically large, suggesting that AI grading is comparable to human grading. The study's findings imply that the grading system for AI might be different and is, therefore, incapable of replacing the human aspect of the grading process in Moroccan higher education, opening the possibility to enhance AI models of assessing EFL and align them with human standards of assessing EFL.

Keywords: Artificial intelligence, assessment, ChatGPT, e-assessment, Moroccan higher education, writing assignment

How to cite: Ennadir, H., & Habiballah, S. (2025). A Comparative Study of Human and AI-Assisted Assessment using ChatGPT: The Case of Moroccan EFL Learners *Arab World English Journal, 16*(3): 175-187. https://dx.doi.org/10.24093/awej/vol16no3.10

Introduction

Artificial Intelligence (AI), since its emergence, has been developing to be able to perform multiple actions. It has been used in a number of activities in different fields such as speech recognition, autonomous learning and scheduling, game playing, spam fighting, logistics planning, machine translation, etc. (Stuart & Peter, 2010). With its massive advancement impacting several fields and domains, AI has also come to affect the field of education in general and the field of higher education in particular. Research predicted that in the upcoming years, one of the most useful technologies will be artificial intelligence (AI) (Chai et al., 2021). No field was exempt from the AI invasion; teaching and learning are a great example of that, notably in the aspect of classroom assessment (Jurane-Bremane, 2023).

In Moroccan educational contexts, research shows that concerns were raised about the use of AI in the classroom, particularly in early childhood education. Yet, using AI as technical support seems to have optimistic implications (Douali et al., 2022). Moreover, Moroccan higher education has institutionalized the adoption of AI, with Mohammed VI Polytechnic University (UM6P) being the first university in Africa to formally implement ChatGPT Edu as an assistant in teaching, research, and administration (Hespress EN, 2024). Statistically, Moroccan educators are eager to use AI in the classroom, with 64% of them thinking it would be simple to learn the required skills. Nonetheless, 92% of educators acknowledge that users' intentions may have an impact on the adoption of AI (Moukhliss et al., 2024).

This study focuses on summative e-assessment in the context of Moroccan higher education. It aims to compare human and AI grading in EFL classes to unravel the accuracy of each one. Writing assignments of two departments at the faculty of letters and humanities Dhar El Mahraz were graded by their respective instructors and ChatGPT-4 abiding by five criteria: grammar, spelling, punctuation, clarity, and structure.

The paper raises the research question “to what extent is AI summative e-assessment accurate compared to the human one in the context of Moroccan EFL education?”. The significance of this study addresses the issue of the impact of AI tools on summative e-assessment in Moroccan universities. By focusing on comparing AI and human grades, this paper discusses practical issues of the use of AI in education and offers empirical insights into the Moroccan breakthrough regarding educational technology. Most studies tackle the issue of AI and education by simply highlighting its advantages and pitfalls when used as a learning support tool rather than an instrument for assessing academic assignments. This creates a research gap that this study addresses to enhance the understanding of AI-assisted assessment accuracy.

Literature Review

Artificial Intelligence

Artificial intelligence can be defined as a field of study and the resulting innovations and developments that have culminated in computers, machines, and other artifacts having human-like intelligence characterized by cognitive abilities, learning, adaptability, and decision-making capabilities (Chen, 2020). The term Artificial Intelligence was first coined by John McCarthy in 1956 (Sumakul et al., 2022) and can be defined as the existence of thinking and intelligent reasoning, which can go into intelligent machines (Turing, 2009 as cited in Jen & Salam, 2024). Thanks to its impact on the field of higher education, AI has eventually led to a new field of research named Artificial Intelligence in Higher Education (AIED). Zawacki-Richter et al. (2019) classify research in AIED into four broad areas, which are Profiling and Prediction, Intelligent

Tutoring Systems (ITS), Adaptive Systems, and Personalisation and Assessment and Evaluation. The latter is the focus of this paper.

Profiling and prediction

This area of research focuses on how AI is used to make informed decisions about students' academic journey (Zawacki-Richter et al., 2019). The use of AI prediction models to determine student profiles has increased dramatically. These models focus on the result of student performance to draw a line of action that allows the student to strengthen their deficiencies and gradually achieve their performance objectives (Y. Chen, 2018 as cited in Rincon-Flores, 2022). AI allows for the analysis of vast amounts of data, including academic records, attendance patterns, in addition to social and emotional factors, which provide a comprehensive view of the academic profile of students (Katarya, R. 2019 as cited in Maulana, A., Idroes, 2023).

Intelligent Tutoring Systems (ITS)

This domain is concerned with the use of AI to provide tailored instructional interventions. This domain is concerned with the use of AI to provide tailored instructional interventions (Zawacki-Richter et al., 2019). An ITS provides computer-based tutorials on topics in well-defined structured subjects. ITS delivers a sequence of information, activities, and quizzes adapted to each individual student. While the student engages with a particular activity, the system captures thousands of data points such as clicks, typing, correct answers, and misconceptions. These data are analysed in order to determine the next information, activity, and quiz to be delivered (Holmes and Tuomi, 2022). An ITS can monitor students' progress, provide feedback, hints, and scaffolding, and select appropriate problems or tasks for students to practice. Besides, they can model the psychological state of students, such as motivation, emotion, and cognition, as well as their prior knowledge, skills, and preferences (Lin et al., 2023).

Adaptive Systems and Personalisation

Adaptive systems and personalisation focus on the use of AI in personalising the educational experience of learners (Zawacki-Richter et al., 2019). AI provides specialized support and raises knowledge-gap awareness, enabling instructors to teach effectively and efficiently through personalized and adaptive instruction (Guan et al., 2020 as cited in Chen, 2021). AI is a strong toolkit for expanding the adaptivity provided to students. It can develop a technology that meets students where they are, builds on their strengths, and grows their knowledge (U.S. Department of Education, 2023).

Assessment and Evaluation

The final area is AI-based assessment and evaluation, which is the focus of this study. AI in assessment and evaluation focuses on the automation of evaluation using AI (Zawacki-Richter et al., 2019). According to Mirchi et al., 2020 as cited in González-Calatayud, 2021, the researchers used AI for simulation-based training in medicine, and they created a Virtual Operative Assistant in order to provide automatic feedback to students based on metrics of performance. From a formative educational perspective, virtual reality and AI are integrated to classify students in relation to proficiency performance benchmarks, and the system gives feedback to help them improve (Janpla, 2020 as cited in González-Calatayud, 2021). Some autograders are claimed to diagnose the type of error and suggest corrections, while others, depending on the domain, are

claimed to score student answers correctly with an accuracy of about 90 percent (Hsu et al., 2021 as cited in Holmes and Tuomi, 2022).

AI and Assessment

There is a strong association between AI and human grades, according to research on AI-based essay grading systems in higher education. However, its successful incorporation depends on ethical considerations, appropriate instruction, and the improvement of AI-generated material. Enhancing AI writing tools and investigating new avenues are necessary to improve writing in higher education (Aljuaid, 2024).

Assessment is one of the most affected areas by AI in higher education. AI assessment tools provide personalized feedback through the analysis of students' responses to assignments, helping in identifying their areas of strength and weakness. Eventually, providing educators with important insights into instructional effectiveness, student performance, and learning outcomes (Nazaretsky et al., 2022).

Similarly, Huang et al. (2023) state that AI tools can provide insights about the effectiveness of the instruction and identify areas where teaching strategies may need to be adjusted. Besides, the study adds that AI-powered tools can help in automating different aspects of the assessment process, saving time and reducing the burden on teachers. AI-powered grading tools are an example; these tools can analyse essays of students and provide feedback on grammar, structure, and content; hence, reducing teachers' time grading assignments.

Assessment and E-assessment

According to Gierl et al. (2014), the process of assessment is deemed challenging by teachers, especially when it comes to hand-marking essays. Manual grading requires careful consideration, critical analysis, and tailored criticism, which adds to the lengthy and highly resource-intensive grading process as a whole. Furthermore, the difficulties of manual assessment are made worse by the vast number of sheets, particularly in light of the expanding student population.

Particularly in higher education, e-assessment is becoming a daily need in the educational process. Applications, learning management systems, or specialist software can provide it. Appiah and Van Tonder (2018, p. 1454) assess that "e-assessment involves the use of any technological device to create, deliver, store and/or report students' assessment marks and feedback".

Formative vs Summative Assessment

In most research studies, the focus lies on formative assessment. The concept of employing AI for assessment is founded on the assistance that this technology may offer teachers with a large class size (Bai et al., 2022) or on the notion that learning is enhanced by the rapid feedback that students get (Podsiad and Havard, 2020). When the learning process is finished, summative assessment takes place. Tests, projects, presentations, and performance assignments are a few types of summative assessments. The goal behind summative assessment is to demonstrate how well a student has grasped the unit's concepts, knowledge, and abilities. The latter is the focus of this paper.

AI in EFL Writing Assessment

Research emphasizes the importance of automated scoring, particularly in EFL contexts. Globally, Uyar and Büyükahıska (2025) argue that despite ChatGPT's potential as a scoring tool, the discrepancies indicate that it is still not proficient enough for practical use. Furthermore, Yavuz et al. (2024) conclude that although ChatGPT and Google's Bard showed a high degree of consistency and grading proficiency, further work and modification are required to gain a more sophisticated comprehension of non-objective essay requirements. In Morocco, generative AI is not only used for assessment but also for test generation. A study conducted by Far-hat and Ouchouid (2024) revealed that Moroccan EFL instructors' effectiveness when utilizing ChatGPT to create formal exams was positively impacted by providing them with a guide that included prompt designing and test construction guidelines. Among other things, it was suggested that their professional training incorporate AI's application in planning and teaching.

Method

The study compares human and AI-assisted assessment in English writing assignments. The objective of the study is to examine the extent to which human assessment is similar to or different from AI-assisted assessment. In order to meet the objective of the study, the paper used ChatGPT-4. This version of ChatGPT is designed for detailed analysis, nuanced understanding, and structured responses, making it well-suited for language correction and teaching. Besides, the study used the standard ChatGPT interface. A writing assignment was assigned to the participant in which they were asked to describe a person of their choice, like a friend or a family member. The pieces of writing were then corrected by both ChatGPT-4 and two English language teachers to assess the difference between the two methods of assessment.

Participants

The sampling approach adopted was convenience sampling. The sample was composed of 30 participants from the Faculty of Letters and Humanities Dhar El Mahraz at the University of Sidi Mohamed Ben Abdellah in Fez, Morocco. Eighteen participants were from the department of French language studies; eleven were female participants, and seven were male. Twelve participants were from the department of psychology; eleven were female participants, with one male.

Procedures

To conduct the comparison, participants were asked to write a paragraph in which they describe a person of their choice, like a friend or a family member. They had twenty minutes to do the assignment on a Google Form. First, the assignment was graded by two teachers focusing on five criteria, which are grammar, spelling, punctuation, clarity, and structure. Then, the same pieces of writing were graded using ChatGPT-4. ChatGPT-4 was given the following prompt: "Grade the following paragraph out of 10 based on the following criteria: Grammar, spelling, punctuation, clarity, and structure. Each criterion should be graded out of 2." The total grade was 10 out of 10, with 2 grades for each criterion. The grammar criterion focused on correct sentence structure, subject-verb agreement, proper use of articles, and pronoun accuracy. The criterion of spelling checked the correct spelling of each word. For punctuation, it evaluated capitalization at the beginning of sentences and for proper nouns, as well as the use of periods to separate ideas. The clarity criterion examined whether the intended meaning could be understood by the reader

without confusion. The final criterion was structure; it focused on the organization of ideas into logical, separate sentences and paragraphs.

Results

This section lays out the data in tables and charts and provides descriptive analysis accounting for observations, patterns, or crossovers. The section presents an interpretation for both the French and psychology departments.

Findings of the French department

Table 1. *French department's students' human and AI grades*

Human Grading	AI Grading	Difference	Degree of Difference
4	8	Different	-4
6	7.5	Different	-1.5
6	8	Different	-2
10	9.5	Different	0.5
7	9.5	Different	-2.5
4	7.5	Different	-3.5
6	8	Different	-2
9	10	Different	-1
5	8	Different	-3
4	8	Different	-4
9	10	Different	-1
7	9	Different	-2
3	10	Different	-7
7	10	Different	-3
5	8	Different	-3
5	10	Different	-5
9	10	Different	-1
7	10	Different	-3

The table represents the grades provided by the teacher (Column 1) and the grades provided by ChatGPT (Column 2) to the 18 students of the French department. The third column accounts for whether the grades were different or similar among the assessors while the fourth column shows the difference scores.

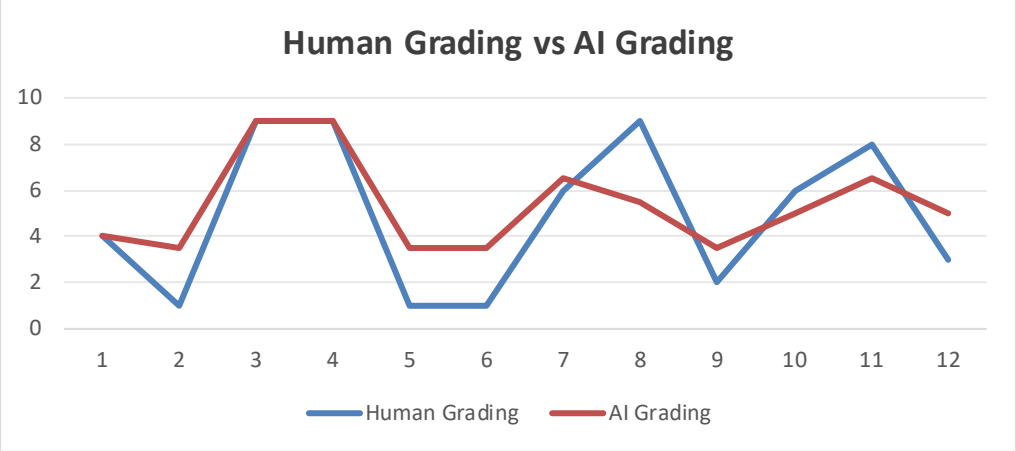


Figure 1. Comparison of human vs. AI grading

The figure illustrates a line chart comparing the human and AI grades among the 18 students of the French department. The x-axis represents the students, and the y-axis indicates the grades out of 10. The chart clearly shows that the AI grading is, overall, higher than the human one. The lowest AI grade is 7.5 while the lowest human grade is 3. Grading paragraphs consistently as 7.5 and higher raises questions of credibility of AI tools.

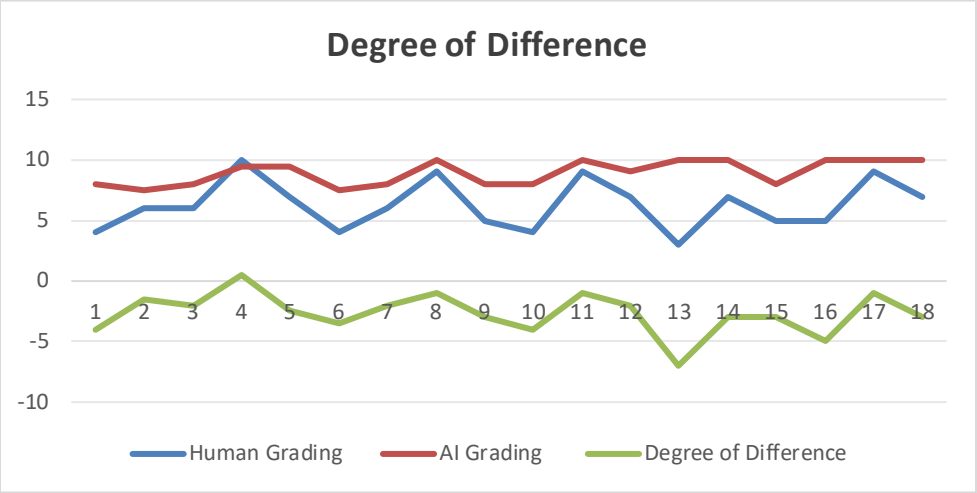


Figure 2. Degree of difference between human and AI assessors

The line chart reflects the degree of difference between human and AI assessment in the context of French department students. The difference scores range between -7 and 0.5, representing a wide variance among assessors. For example, student n14 was given a 3 by the human assessor and a 10 by the AI tool. This suggests that ChatGPT can, sometimes, overlook the assessment criteria provided in the prompt. In other cases, student n4 and student n8 were graded nearly similarly by the two assessors, shrinking the grade gap by 0.5 and 1.

To point out the significant differences between human and AI grading in the French department, the researchers ran several statistical tests. The paired-samples t-test gave a p-value of (> .001). Since the value is below the standard 0.05 significance level, it shows that the difference in average grades (6.39 for human grading and 9.07 for AI grading) is statistically significant. This means the difference is due to there being a consistent gap between the two grading methods.

Effect size was also measured using Cohen's d, which was calculated at −1.55, indicating a very large and substantial difference. The negative value shows that ChatGPT routinely assigned higher grades than the human instructors.

These results show that AI grading produces consistently higher scores than human grading, with differences that are both statistically significant. This sheds light on the impact of AI on assessment outcomes as well as the urgent need for further research regarding grading accuracy.

Findings of the psychology department

Table 2. *Psychology department's students' human and AI grades*

Human Grading	AI Grading	Difference	Degree of Difference
4	4	Match	0
1	3.5	Different	-2.5
9	9	Match	0
9	9	Match	0
1	3.5	Different	-2.5
1	3.5	Different	-2.5
6	6.5	Different	-0.5
9	5.5	Different	3.5
2	3.5	Different	-1.5
6	5	Different	1
8	6.5	Different	1.5
3	5	Different	-2

This chart summarises the results of the Department of Psychology. 12 students were given a writing task, the pieces of writing were graded by a human assessor, who was the teacher of the subject, referred to as Human Grading (Column 1), they were then graded by an AI assessor, which was ChatGPT, referred to as AI Grading (Column 2). Three grades were similar while the other grades were not (column 3). Column 4 represents the degree of difference between Human grading and AI grading, which varies from 3.5 to -2.5.

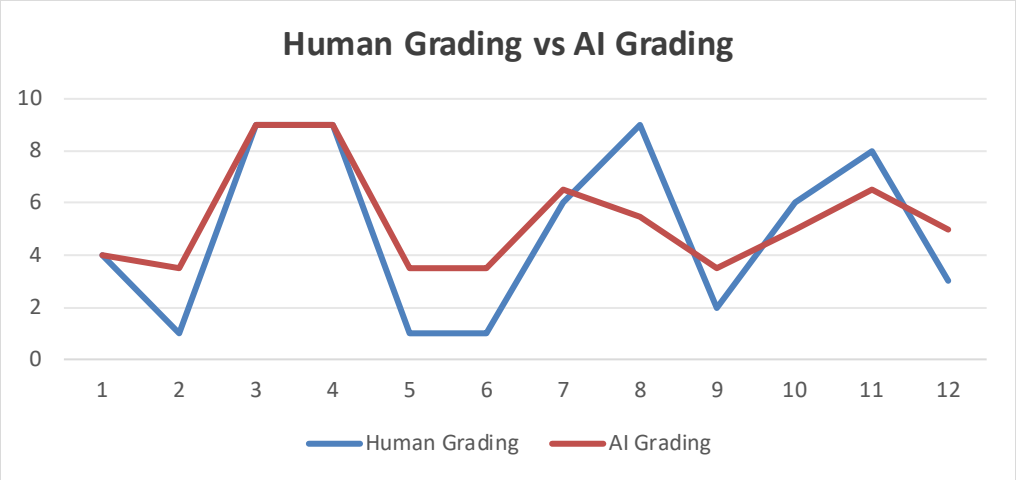


Figure 3. Comparison of human vs. AI grading

The graph represents both Human Grading and AI grading. Two instances from the data show that human grading and AI grading may match and may not. Both the teacher and ChatGPT gave 4 grades for the first writing, which shows that human grading and AI grading may match. The second piece of writing got 1 grade from the teacher and 3.5 grades were given by the AI tool (ChatGPT), which shows that there might be a difference between the grading of teachers and AI tools.

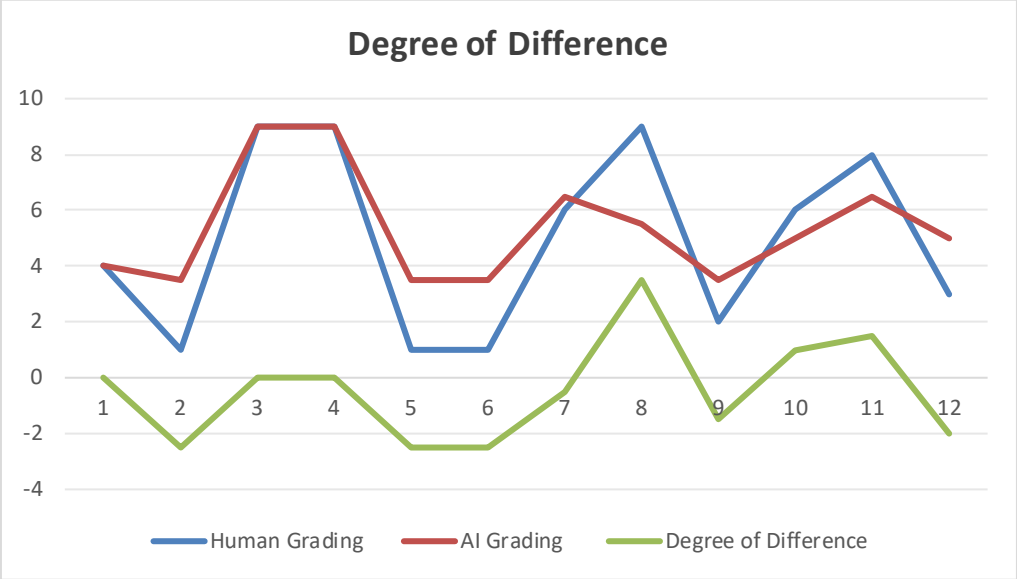


Figure 4. Degree of difference between human and AI assessors

Graph number two is a reflection of the degree of difference between the two grades; the difference ranges between 3.5 and -2.5. For students number 1,3, and 4, the degree of difference is 0 as the grades that were given by the teacher and ChatGPT were, respectively, 4 and 9,9. For student number 8 with a grade of 9 from the teacher and a grade of 5.5 from ChatGPT, the degree of difference is 3.5 grades; the teacher's grade was 3.5 grades higher. On the other hand, student number 2 was given 1 grade by the teacher and 3.5 grades by the AI tool, resulting in a -2.5 degree of difference; the teacher's grading was lower by 2.5 grades compared to the AI tool.

In order to examine whether a significant difference exists between human and AI-assessment in the department of psychology, the researchers conducted a series of statistical analyses. In the first place, a paired t-test was performed to determine significance; the test yielded a p-value of 0.41. Given that this value is greater than the standard 0.05 significance level, the difference in the mean grades, 4.92 for human grading and 5.38 for AI-grading, is not statistically significant. This shows that there is a high probability that the observed difference is due to random chance and not a systematic variation between the two assessment methods.

In addition to significance, the study used Cohen's d to measure the effect size and to quantify the magnitude of the difference. The calculated Cohen's d was −0.25, which is considered a small effect size. This shows that despite the fact that the AI grades were slightly higher on average, the practical difference is minimal. The negative sign simply reflects that the AI's mean grade was higher than the human's. Eventually, while the human grading and AI-grading are not identical, their differences are neither statistically significant nor practically large, suggesting that the AI-grading is comparable to human grading.

Based on the comprehensive analysis of the data, the AI grading system appears to be a reliable alternative to human grading. While there are some minor differences in the average grades and the overall score spread, these are not statistically significant and have a small practical effect. This suggests that ChatGPT can provide grades that are largely consistent with a human grader. Therefore, the AI grading system is a comparable and potentially valuable tool.

Discussion

According to this study's comparison of AI and human grading in the context of Moroccan higher education, AI grades tend to be marginally higher than human-assigned grades. It's interesting to see, in some cases, that ChatGPT provided higher grades, especially when human graders gave lower ones. This aligns with findings of the mixed-method study conducted by El Messaoudi et al. (2025), which shows that students appreciate ChatGPT for its efficiency (73.2% reported time-saving) and idea generation capabilities (68.3% reported frequent benefit), reflecting that ChatGPT is rewarding different qualities than human graders. This study's findings imply that the grading system for AI might be different and is, therefore, incapable of replacing the human aspect of the grading process. This opens the possibility to enhance AI models to be aligned with the human standards of assessment. These concerns were highlighted by Khoual et al. (2025) who used predictive modeling to report that 395 students from various Moroccan faculties reported issues in regard to ethics of AI use in education.

To verify the correctness of AI technology in academic assessment, similar studies were conducted. For instance, Grivokostopoulou et al.'s (2017) work compared the outcomes of AI-assisted assessment with those of manual assessments. In their study, the findings showed a strong correlation and were very comparable. It was found that the results were quite similar when compared to the grade levels provided by the teachers (very low, low, medium, good, and excellent). The instructors, however, identified the highest scores as outstanding, but the AI tool used did not. Moreover, research indicates that AI-assisted assessment can be particularly effective for grammatical accuracy, which might explain some of the discrepancies observed. For instance, a comparison between the two modes of assessment made by Bouziane and Bouziane (2024) indicated that a balanced approach to effective essay correction is suggested by the fact that while ChatGPT performs better in grammar, spelling, sentence structure, relevance, and supporting evidence, human evaluators do better in thematic consistency.

Conclusion

AI assessment tools such as ChatGPT are one of the most remarkable advances that have come to influence the educational field, particularly at the level of assessment. This study was an attempt to compare human assessment with AI assessment. Pieces of writing of two groups of students from the faculty of Letters and Humanities, Dhar El Mahraz, Fes, one group belongs to the department of French language studies, and the other to the department of psychology. Teachers of the two groups, in addition to ChatGPT-4 assessed and graded the pieces of writing provided by the participants. According to the results, the grading of human assessors and AI grading in the department of French language studies are not different, while in the department of psychology, they are similar. In the department of French language studies, ChatGPT produced higher scores than human grading, with differences that are both statistically substantial, while in the department of psychology, the differences between human grading and ChatGPT are neither statistically significant nor practically large, suggesting that AI-grading is comparable to human grading.

The results of this study indicated that AI-assessment tools such as ChatGPT might provide grades that are significantly different in some contexts, while in other contexts, the difference is not statistically significant. Therefore, any use of AI-assessment tools should be implemented with some precaution. Hence, Moroccan teachers, educators, and tool developers should work hand in hand to develop AI-assessment tools that can provide comparable grades even in different

contexts. For a reliable use of AI-assessment tools, they should provide grades that align with human standards of assessment. Besides, teachers and educators are encouraged to dive into the art of prompting, as it is a crucial aspect for tools such as ChatGPT, as it might affect the whole process of assessment and grading. Moreover, stakeholders in Moroccan higher education, including teachers, educators, administrative staff, and tool developers, should benefit from initiatives such as the National Strategy: Digital Morocco 2030. One of the aims of this initiative is to use AI to enhance educational outcomes, foster digital literacy, and improve the efficiency of academic institutions.

The focus of the study was to compare human and AI assessment tools in grading writing, with special focus on grammar, spelling, punctuation, clarity, and structure. Future research could address this comparison, concentrating on aspects such as cohesion, coherence, vocabulary, and conciseness. Researchers are recommended to address research questions, such as to what extent AI and human assessment are similar or different in assessing cohesion in writing? To what extent is AI similar or different from teachers in assessing conciseness in writing? Besides, researchers are encouraged to address different aspects of EFL, like comprehension and grammar, in their comparison between AI and human assessment. They could address research questions similar to the following: In what way are AI and human assessment tools similar or different in assessing comprehension? To what extent does AI assess grammar similarly or differently from human assessment?

Funding: This research is not funded.
Acknowledgments: Not applicable.
Conflicts of Interest: The authors declare no conflict of interest.
Authenticity: This manuscript is an original work.
Artificial Intelligence Statement: ChatGPT was used to enhance the readability and quality of the paper.

Authors Biography

Hajar Ennadir is a second year PhD candidate at the Faculty of Letters and Human Sciences, Dhar El Mahraz, Fez, Morocco, with a focus on education, EFL teaching and learning, AI in education, and Academic writing. She serves as a part time EFL instructor. ORCid ID: https://orcid.org/0009-0003-8121-4640

Salma Habiballah is a PhD holder in applied linguistics. Her research interests include examining the effectiveness of online education as perceived by both teachers and students, as well as discovering the ways different members of educational boards use to enhance their digital learning experiences. ORCid ID: https://orcid.org/0000-0002-4122-7477

References

Aljuaid, H. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal (AWEJ) Special Issue on ChatGPT*, 26-55. https://dx.doi.org/10.24093/awej/ChatGPT.2

Appiah, M., & Van Tonder, F. (2018). E-assessment in higher education: A review. *International Journal of Business Management & Economic Research*, *9*(6), 1454-1460.

Bai, J. Y., Zawacki-Richter, O., Bozkurt, A., Lee, K., Fanguy, M., Cefa Sari, B., & Marín, V. I. (2022). Automated essay scoring (AES) systems: Opportunities and challenges for open and distance education. *In Proceedings of The Tenth Pan-Commonwealth Forum on Open Learning (PCF10)*. https://doi.org/10.56059/pcf10.8339

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education. *Computers & Education*, *151*, 103855. https://doi.org/10.1016/j.compedu.2020.103855

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology & Society*, 25 (1), 28-47.

Douali, L., Selmaoui, S., & Bouab, W. (2022). Artificial intelligence in education: Fears and faiths. *International Journal of Information and Education Technology*, 12(7), 650-657. doi: 10.18178/ijiet.2022.12.7.1666

El Messaoudi, M., Lamiae, A., & Daoudi. R. (2025). Student Satisfaction with ChatGPT as a Writing Tool in Moroccan Higher Education: A Mixed Methods Study. *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence:*193-219. https://dx.doi.org/10.24093/awej/AI.11

Gierl, M. J., Latifi, S., Lai, H., Boulais, A. P., & De Champlain, A. (2014). Automated essay scoring and the future of educational assessment in medical education. *Medical education*, *48*(10), 950-962. https://doi.org/10.1111/medu.12517

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied sciences*, *11*(12), 5467. https://doi.org/10.3390/app11125467

Hespress En. (2024, May 31). "Study: 75% of students see AI as an asset for learning". https://en.hespress.com/85975-study-75-of-students-see-ai-as-an-asset-for-learning.html

Huang, A. Y., Lu, O. H., & Yang, S. J. (2023). Effects of Artificial Intelligence–Personalized Recommendations on Learners' Learning Engagement, Motivation, and Outcomes in a Flipped Classroom. *Computers & Education*, *194*, 104684. https://doi.org/10.1016/j.compedu.2022.104684

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European journal of education*, *57*(4), 542-570. https://doi.org/10.1111/ejed.12533

Jen, S. L., & Salam, A. R. (2024). Using Artificial Intelligence for Essay Writing. *Arab World English Journal (AWEJ) Special Issue on ChatGPT*, April 2024: 90-99. https://dx.doi.org/10.24093/awej/ChatGPT.5

Khoual, M., Elkaimbillah, Z., Mcharfi, Z., & El Asri, B. (2025). AI in Moroccan education: Evaluating student acceptance using machine learning classification models. *Indonesian Journal of Electrical Engineering and Computer Science*, 37(1), 452-462. http://doi.org/10.11591/ijeecs.v37.i1.pp452-462

Lin, CC., Huang, A.Y.Q. & Lu, O.H.T. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learn. Environ.* 10, 41 (2023). https://doi.org/10.1186/s40561-023-00260-y

Maulana, A., Idroes, G. M., Kemala, P., Maulydia, N. B., Sasmita, N. R., Tallei, T. E., ... & Rusyana, A. (2023). Leveraging Artificial Intelligence to Predict Student Performance: A

Comparative Machine Learning Approach. *Journal of Educational Management and Learning, 1*(2), 64-70. https://doi.org/10.60084/jeml.v1i2.132

Maryam, F., Jamaa, O. (2025). An Empirical Assessment of Moroccan EFL Teachers' Use of Generative AI for EFL Formal Assessment: An Intervention Study. *Social Sciences*, 14(1), 8-21. https://doi.org/10.11648/j.ss.20251401.12

Moukhliss, G., Lahyani, K., & Diab, G. (2024). The Impact of Artificial Intelligence on Research and Higher Education in Morocco. *Journal of Education and Learning (EduLearn)*, 18(4), 1292-1300. https://doi.org/10.11591/edulearn.v18i4.21511

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology, 53*(4), 914-931. https://doi.org/10.1111/bjet.13232

Podsiad, M., & Havard, B. (2020). Faculty acceptance of the peer assessment collaboration evaluation tool: A quantitative study. *Educational Technology Research and Development, 68*, 1381-1407. https://doi.org/10.1007/s11423-020-09742-z

Rincon-Flores, E. G., Lopez-Camacho, E., Mena, J., & Olmos, O. (2022). Teaching through Learning Analytics: Predicting Student Learning Profiles in a Physics Course at a Higher Education Institution, 82-89. https://doi.org/10.9781/ijimai.2022.01.005

Sumakul, D. T. Y., Hamied, F. A., & Sukyadi, D. (2022). Artificial intelligence in EFL classrooms: Friend or foe?. *LEARN Journal: Language Education and Acquisition Research Network, 15*(1), 232-256.

U.S. Department of Education, Office of Educational Technology, Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations, Washington, DC, 2023.

Uyar, A. C., & Büyükahıska, D. (2025). Artificial intelligence as an automated essay scoring tool: A focus on ChatGPT. *International Journal of Assessment Tools in Education*, 12(1), 20-32. https://doi.org/10.21449/ijate.1517994

Yavuz, F., Çelik, Ö., & Yavaş Çelik, G. (2025). Utilizing large language models for EFL essay grading: An examination of reliability and validity in rubric-based assessments. *British Journal of Educational Technology*, 56(1), 150-166. https://doi.org/10.1111/bjet.13494

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education–Where are the Educators?. *International journal of educational technology in higher education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0