AI-Assisted Grading – A Study on Efficiency and Fairness

Dr. Fazil T. Najafi, University of Florida

For more than 40 years, Dr. Najafi has worked in government, industry, and education. He earned a BSCE 1963 from the American College of Engineering, University of Kabul, Afghanistan. In 1966, Dr. Najafi earned a Fulbright scholarship and did his B.S., MS, and Ph.D. degree in Civil Engineering at Virginia Polytechnic Institute and State University, Blacksburg, Virginia; his experience in industry and government includes work as a Highway Engineer, Construction Engineer, Structural, Mechanical, and Consultant Engineer. Dr. Najafi taught at Villanova University, Pennsylvania, and was a visiting professor at George Mason University and a professor at the University of Florida, Department of Civil and Coastal Engineering. He has received numerous awards, such as Fulbright scholarship, teaching awards, best paper awards, community service awards, and admission as an Eminent Engineer into Tau Beta Pi. The Florida Legislature adopted his research on passive radon-resistant new residential building construction in the HB1647 building code of Florida. Najafi is a member of numerous professional societies and has served on many committees and programs; and continuously attends and presents refereed papers at international, national, and local professional meetings and conferences. Lastly, Najafi attends courses, seminars, and workshops and has developed courses, videos, and software packages during his career. Najafi has more than 300 refereed articles. His areas of specialization include transportation planning and management, legal aspects, construction contract administration, public works, and Renewable Energy.

Vani Ruchika Pabba, University of Florida

Vani Ruchika Pabba holds a Master of Science in Computer Science from the Herbert Wertheim College of Engineering at the University of Florida, where she served as a Graduate Research Assistant. Her research focuses on artificial intelligence in education, including natural language processing for automated grading and feedback generation, multi-modal learning (integrating vision and language models), and generative AI. Her broader interests include sustainable computing, IoT, and the development of smart cities and connected environments. Prior to her graduate studies, she accumulated three years of professional experience as a Software Engineer in India, specializing in software design and development for enterprise applications. She is committed to advancing educational technology and addressing real-world challenges through innovative computing solutions.

Dr. Rajarajan Subramanian, Pennsylvania State University, Harrisburg, The Capital College

Rajarajan Subramanian currently holds the position of Associate Teaching Professor of Civil Engineering and Construction (SDCET) programs at Pennsylvania State University at Harrisburg. He has 25 years of experience in academia, in teaching roles, including 10 years at Annamalai University in India and three years at Linton Institute of Technology in Malaysia. He also has 10 years of professional engineering experience. Prior to joining Pennsylvania State University, he worked as a Transportation Engineer at the Maryland State Highway Administration.

Dr. Sofia M Vidalis, Pennsylvania State University, Harrisburg, The Capital College

Sofia M. Vidalis is an Associate Professor in the Department of Civil Engineering and Engineering Technology at Penn State Harrisburg. She received her Ph.D., Masters, and Bachelor's in Civil Engineering from the University of FL. Her background is in construction management, transportation planning, and operations. She has had industry experience as a Transportation Engineer at Florida Design Consultants and as a consultant for Applied Research Associates.

©American Society for Engineering Education, 2025

AI-Assisted Grading – A Study on Efficiency and Fairness

Abstract: The adoption of AI-assisted grading systems in education is gaining attention due to their potential to streamline assessment processes. While AI has proven effective in automating objective tests, its application to theoretical and research-based assignments has been explored less. This study investigates the efficiency and fairness of using AI, specifically ChatGPT, to grade theoretical understanding and research paper assignments in undergraduate and graduate courses. The research was conducted in two phases. In the first phase, we assessed ChatGPT's performance in grading assignments, focusing on time efficiency, consistency, and grading patterns. We compared AI-assisted grading with traditional human grading methods in the second phase. We then analyzed variations in scores, potential biases, and feedback's perceived usefulness. We conducted surveys to gather perceptions from both students and educators regarding AI-based grading.

The results indicated that AI-assisted grading reduced grading time and provided more consistent feedback, especially for assignments with clear rubrics. However, challenges arose in handling nuanced or subjective responses, such as those in research papers, leading to concerns about fairness and bias in grading. These limitations highlighted the need to investigate AI's capabilities further in evaluating complex academic work. This research contributed to the growing conversation on AI's role in education, particularly its potential to automate grading for complex assignments. The findings offer insights into the benefits and drawbacks of AI in educational settings, offering valuable guidance for educators considering AI integration to enhance efficiency while maintaining fairness in grading practices.

Keywords: AI-Driven Academic Evaluation, Artificial Intelligence in Education, Grading Bias, Theoretical Knowledge Assessment, ChatGPT, Automated Grading.

1. Introduction

Adopting artificial intelligence (AI) in education is revolutionizing traditional teaching and learning processes, with applications ranging from personalized learning platforms to automated grading systems. AI-assisted grading has garnered significant attention for its potential to streamline assessment processes, particularly for large-scale courses where manual grading is resource-intensive. Automated systems have been successfully implemented for objective tasks like multiple-choice quizzes, but their applicability to more complex assessments, such as theoretical understanding and research-based assignments, remains underexplored [1].

While effective in evaluating nuanced and subjective responses, traditional grading methods are often labor-intensive and prone to inconsistencies due to human biases and variability. Factors such as grader fatigue, subjectivity, and varying interpretations of rubrics can lead to score discrepancies and feedback quality [2]. These challenges underscore the need for scalable solutions that enhance grading efficiency while maintaining fairness and accuracy.

Recent advancements in natural language processing (NLP) models, such as ChatGPT, have demonstrated remarkable capabilities in generating coherent, contextually relevant text, raising the question of their potential utility in grading assignments. ChatGPT and similar large language models (LLMs) leverage deep learning techniques to process and analyze written responses, offering opportunities to automate grading processes for assignments that require

theoretical and analytical understanding [3].

However, deploying AI in this capacity introduces challenges. For example, LLMs may struggle to handle nuanced, context-dependent answers or subjective interpretations common in research-based assignments. Concerns about fairness and grading bias persist, as AI systems may unintentionally replicate or exacerbate societal biases embedded in training data [4]. These issues necessitate a rigorous evaluation of AI-assisted grading systems to determine their suitability for academic use.

This study investigates the efficiency and fairness of using AI, specifically ChatGPT, to grade theoretical and research-based assignments in undergraduate and graduate courses. We aim to provide insights into this technology's benefits, limitations, and perceptions by comparing AI-assisted grading with traditional human grading. The findings will contribute to the growing discourse on AI's role in education, offering practical guidance for educators seeking to integrate AI into assessment processes while ensuring equitable outcomes.

2. Literature Review

Artificial Intelligence (AI) has been increasingly adopted in educational contexts, primarily for automating routine tasks and enhancing teaching and learning experiences. AI systems have shown promise in improving efficiency and consistency in assessment, particularly for large-scale courses. While multiple-choice and other objective test formats are well-suited for automation, the application of AI in grading complex assignments, such as essays and research papers, remains challenging [5].

2.1. AI in Educational Assessment

AI-driven grading systems have demonstrated significant potential in automating objective assessments. For example, automated essay scoring (AES) tools such as e-Rater have been widely used to evaluate written assignments based on linguistic features and coherence [6]. These systems have shown efficiency in reducing grading time and providing consistent scores, particularly in standardized testing scenarios.

Recent advancements in natural language processing (NLP) and machine learning have expanded the capabilities of AI systems, allowing them to handle more complex tasks like essay grading. Large language models such as OpenAI's GPT series have been trained on vast text corpora, enabling them to analyze and generate highly accurate human-like text [3]. Studies have highlighted the ability of these models to assess writing quality, summarize content, and provide meaningful feedback [7].

2.2. Benefits of AI-Assisted Grading

AI-assisted grading systems offer several advantages over traditional human grading. They can significantly reduce the time and effort required to evaluate assignments, especially in courses with high enrollment [8]. These systems also provide greater consistency in grading by applying the same criteria uniformly across all assignments, reducing the risk of variability associated with human graders [9].

Additionally, AI-based systems can provide immediate feedback to students, enabling them to identify and address learning gaps more effectively. This feedback can be tailored to specific aspects of the assignment, offering detailed suggestions for improvement [10]. These features

make AI-assisted grading particularly appealing for formative assessments.

2.3. Challenges and Limitations

Despite their benefits, AI-assisted grading systems face significant challenges in evaluating subjective and nuanced assignments. Research has shown that while AI can effectively assess objective criteria like grammar, structure, and style, it struggles with subjective elements such as creativity, critical thinking, and originality [6]. These limitations raise concerns about the fairness and reliability of AI in grading assignments requiring higher-order cognitive skills.

Bias is another critical issue in AI grading. Since AI models are trained on existing data, they may inadvertently replicate biases in the training datasets [4]. For instance, studies have reported disparities in grading outcomes for non-native English speakers and minority groups, raising ethical questions about the use of AI in academic assessment [11].

2.4. Comparing AI and Human Grading

Comparative studies between AI-assisted and human grading have yielded mixed results. While some studies report high levels of agreement between AI and human scores, others highlight discrepancies in evaluating complex responses [5]. Research suggests that AI systems perform best when used with human oversight, where they can complement human graders by handling routine tasks and providing preliminary scores [9].

2.5. Research Gap

Most existing studies focus on AI's application in automating objective assessments, with limited exploration of its potential for grading theoretical understanding and research-based assignments. Additionally, few studies have examined the perceptions of educators and students regarding AI-assisted grading systems, particularly regarding fairness and usefulness of feedback. This research addresses these gaps by evaluating AI's efficiency, fairness, and perceptions in grading complex academic work.

3. Research Objectives and Questions

3.1. Research Objectives

The primary goal of this research is to evaluate the potential of AI-assisted grading systems, explicitly using ChatGPT, for assessing theoretical and research-based assignments in undergraduate and graduate courses. This study focuses on five key objectives: evaluating the efficiency of AI-assisted grading compared to traditional human grading, analyzing its consistency in applying rubrics uniformly, exploring its fairness by identifying potential biases, assessing the quality of feedback it provides, and understanding the perceptions of educators and students regarding its advantages and limitations.

3.2. Research Questions

The study aims to answer specific research questions to achieve the above objectives. Regarding efficiency, it investigates how AI grading compares to human grading in terms of time savings and whether it reduces educators' workload, particularly in large-scale courses. For consistency, it examines the grading patterns of AI systems across various assignments and whether they reduce variability compared to human graders. The study also explores fairness by identifying potential biases in AI grading, especially for responses from students

with diverse linguistic and cultural backgrounds, and compares these biases to those present in human grading. The quality of feedback is another critical focus, particularly its usefulness, alignment with rubrics, and impact on student learning. Lastly, the study seeks to understand stakeholder perceptions, exploring the views of educators and students on the efficiency, fairness, and acceptability of AI-assisted grading and whether these perceptions vary between undergraduate and graduate stakeholders.

3.3. Broader Implications

The research aims to contribute to the growing literature on AI in education, offering practical insights for adopting AI-assisted grading systems. By addressing these objectives and questions, the study seeks to inform best practices for integrating AI into academic workflows while ensuring fairness, consistency, and quality in grading [10].

4. Methodology

This study adopts a mixed-methods approach to evaluate the efficiency, fairness, and perceived usefulness of AI-assisted grading systems, explicitly using ChatGPT for theoretical and research-based assignments. The research was conducted in two phases: (1) an experimental phase focusing on performance metrics of AI grading and (2) a comparative phase analyzing differences between AI-assisted and traditional human grading methods. Surveys and interviews will supplement quantitative findings by capturing the perceptions of educators and students.

The research design combines quantitative grading performance analysis with stakeholders' qualitative insights. This dual approach ensures a comprehensive understanding of AI-assisted grading systems' strengths, limitations, and perceptions [12].

4.1. Phase 1: AI Grading Performance Analysis

Objective: Assess ChatGPT's performance in grading assignments with respect to time efficiency, consistency, and feedback quality.

- **Dataset Collection:**
    - Assignments from undergraduate and graduate courses were collected across disciplines, including essays, theoretical problem-solving exercises, and research-based assignments. Rubrics for these assignments were predefined to standardize evaluation criteria [5].
    - A 100–150 assignments dataset was randomly sampled to ensure diversity in content and complexity.
- **Metrics for Analysis:**
    1. **Grading Time Efficiency:** Measure the time ChatGPT takes to grade each assignment compared to human graders.
    2. **Grading Consistency:** Evaluate ChatGPT's ability to apply rubric consistently across all assignments, using inter-rater reliability (IRR) as a metric.

3. **Feedback Analysis:** Analyze the quality and relevance of feedback provided by ChatGPT, focusing on its alignment with rubric criteria and its usefulness for student improvement.

4.2. Phase 2: Comparative Analysis

Objective: Compare AI-assisted grading to traditional human grading methods, focusing on variations in scores, fairness, and feedback usefulness.

- **Human Grading Process:**
    - Experienced educators will grade assignments using the same rubrics applied to AI grading. Grading time and feedback provided by educators were recorded.
- **Comparison Criteria:**
    1. **Score Variations:** Analyze discrepancies between AI and human scores using statistical tests like paired t-tests and Bland-Altman analysis.
    2. **Fairness:** Investigate potential biases in AI grading, such as discrepancies in scores for non-native English speakers or students from diverse backgrounds, using fairness metrics from machine learning.
    3. **Feedback Perception:** Conduct sentiment analysis on feedback from both AI and human graders to assess perceived quality and relevance.

4.3. Stakeholder Perceptions

Objective: Gather insights from educators and students on AI-assisted grading systems' perceived effectiveness, fairness, and limitations.

- **Survey Design:**
    - Structured surveys were administered to students and educators to capture their perceptions of AI grading. Questions will address feedback usefulness, perceived fairness, and willingness to adopt AI-based grading.
- **Interviews:**
    - Semi-structured interviews will be conducted with educators to explore their views on integrating AI into grading workflows. Focus groups with students will gather insights on the acceptability and impact of AI-generated feedback.

4.4. Data Analysis

- **Quantitative Data:**
    - Grading efficiency and consistency metrics will be analyzed using descriptive statistics and inferential tests. Bias analysis will involve subgroup comparisons based on demographic characteristics [13].
- **Qualitative Data:**
    - Survey responses and interview transcripts were coded thematically to

identify recurring patterns and perceptions. Triangulation will ensure validity by comparing findings across data sources [12].

5. Results

This section presents the findings of the study, focusing on key aspects of AI-assisted grading using ChatGPT. The results highlight the system's performance in terms of grading efficiency, consistency, and feedback quality. Additionally, comparisons between AI-generated and human-assigned scores are discussed alongside the perceptions of students and educators. The section also identifies key challenges, including bias, limitations in handling creative responses, and issues related to grading AI-generated content, concluding with the broader implications of these findings for academic assessment.

5.1. Grading Time Efficiency

Preliminary findings indicate that AI-assisted grading using ChatGPT significantly reduces grading time compared to human grading. ChatGPT graded assignments 70% faster than human graders on average while adhering to predefined rubrics. For example, assignments requiring approximately 30 minutes of human grading were completed by ChatGPT in under 10 minutes. It aligns with previous studies highlighting AI's ability to handle repetitive tasks quickly and accurately [9]. This efficiency was particularly beneficial in large classes where grading is time-intensive. Surveyed educators acknowledged that the time saved allowed for a greater focus on personalized instruction. However, feedback from both educators and students emphasized the importance of human oversight in reviewing subjective responses.

5.2. Grading Consistency

Analysis of grading consistency revealed that ChatGPT applied rubrics uniformly across all assignments, achieving an inter-rater reliability (IRR) score of 0.89. This score was comparable to or higher than the consistency levels typically achieved by human graders, whose IRR scores averaged 0.82, as reported in educational studies [8]. However, some inconsistencies were observed in cases where rubric criteria required the interpretation of nuanced arguments or creative responses. For instance, in an assignment for an Introduction to Thermodynamics course, students were asked to:

Question: *"Explain the concept of entropy and its implications for the second law of thermodynamics. Provide a real-world engineering example where the principle of entropy increase is observed."*

One non-native English-speaking student responded:

"Entropy is the natural outcome of system processes tending to disorder, and this shows why heat engines can't ever reach perfection. Think of how heat radiates from an engine when running—just as spilled water can't be scooped back perfectly without some loss, heat loss is inevitable. It's the same as how spilled water disperses when the glass tips over."

While the student's analogy of "spilled water" was creative and demonstrated an understanding of irreversible processes and energy dissipation, ChatGPT flagged the analogy as "off-topic" because it did not explicitly reference an industrial engineering example, as prompted. The AI-generated feedback such as:

"Your example is unclear and does not relate directly to a thermodynamic system. Consider using a heat exchanger or a Carnot cycle to explain entropy's impact."

The above illustrates ChatGPT's difficulty in interpreting metaphorical language and unconventional analogies, particularly when these align conceptually but do not match the

expected academic format. In contrast, human graders commended the analogy for being relatable and demonstrating a foundational understanding of entropy's principle of irreversibility. The inconsistency highlighted the model's limitations in processing diverse linguistic styles and non-standard expressions that convey valid engineering concepts. This example underscores the importance of refining AI models to recognize valid interpretations that may deviate from traditional terminology, especially for submissions from non-native speakers or students employing creative explanations.

5.3. Feedback Quality

ChatGPT's feedback was well-structured and aligned with rubric expectations in 85% of cases. Survey responses indicated that 78% of students found the feedback "clear and actionable," particularly for assignments with well-defined criteria. However, feedback for subjective assignments, such as research papers, occasionally lacked depth or failed to address specific nuances, consistent with prior findings on AI's limitations in subjective evaluations [7]. Educators noted that while the AI-generated feedback effectively addressed grammatical and structural issues, it sometimes overlooked deeper analytical or interpretive aspects. One example involved a student's literature review on artificial intelligence in education, where the AI's feedback focused predominantly on formatting and surface-level content rather than critiquing the strength of the argumentation. Students expressed that while the feedback was efficient, it often felt "generic" and less tailored to individual learning needs.

5.4. Score Comparisons Between AI and Human Grading

Preliminary comparisons between AI-generated and human-assigned scores show a strong correlation ($r = 0.91$), indicating substantial agreement. However, discrepancies were noted in assignments that involved subjective judgment, with AI tending to over-penalize unconventional responses that deviated from typical patterns found in its training data [4]. In contrast, human graders were more lenient in acknowledging creative interpretations that adhered to the assignment objectives. For example, one research-based assignment, where the student integrated interdisciplinary perspectives, received a lower grade from ChatGPT due to its unfamiliar format despite receiving high praise from human graders for originality and depth. These findings highlight the need for human oversight in cases requiring nuanced interpretation.

5.5. Perceptions of Educators and Students

Survey results from 50 students and 10 educators reveal a generally positive perception of AI-assisted grading.

- **Students:** 78% of students found AI feedback helpful in improving their assignments, particularly for objective and theoretical tasks. However, 60% expressed concerns about fairness in grading subjective responses. Feedback from non-native English-speaking students reflected apprehension about potential biases and emphasized the need for feedback that acknowledges diverse linguistic styles.
- **Educators:** 85% of educators appreciated the time saved with AI grading but emphasized the need for human review to ensure fairness and contextual accuracy, particularly for complex assignments. One educator stated, "The AI speeds up the grading process, but it doesn't always capture the depth of analysis required for

research assignments." Several educators suggested that AI could be most effective as an initial grading tool, with human graders reviewing more complex cases.

5.6. Challenges Identified

Preliminary findings also highlight several challenges in using AI-assisted grading systems:

1. **Bias in Grading:** Bias was observed, particularly in grading assignments with non-native English phrases, echoing prior research.
2. **Handling Creativity:** ChatGPT struggled to fairly evaluate unconventional or creative responses that fell outside typical patterns, consistent with earlier critiques of AI in subjective assessments.
3. **Feedback Customization:** While feedback was consistent, it lacked the level of personalization that human graders can provide, particularly in addressing individual student needs.
4. **Grading AI-Generated Work:** ChatGPT faced challenges in detecting and fairly evaluating AI-generated content, as it tended to treat well-structured but machine-like responses as authentic without questioning their origin. This challenge raised concerns about the potential for "grading loops," where AI systems grade assignments that similar language models may have produced.

5.7. Implications

These preliminary findings suggest that while ChatGPT is highly effective in reducing grading time and ensuring consistency for structured assignments, its limitations in evaluating subjective and nuanced responses must be addressed. Combining AI with human oversight could mitigate these challenges, providing a balanced approach to grading that leverages AI's efficiency while maintaining fairness and contextual sensitivity.

6. Conclusion

This study highlights the potential of AI-assisted grading systems, such as ChatGPT, in addressing some of the persistent challenges in academic assessment. Preliminary findings demonstrate that AI can significantly reduce grading time and provide consistent evaluations, particularly for assignments with well-defined rubrics. Students and educators recognize AI's benefits in streamlining assessment processes and delivering actionable feedback for structured tasks.

However, limitations in handling nuanced, subjective, and creative responses underline the need for human oversight in AI-assisted grading. Issues such as bias in grading non-native English responses and the inability to accurately evaluate unconventional submissions emphasize that AI systems are not yet fully equipped to replace human graders for complex assignments. These challenges highlight the importance of integrating AI as a complementary tool rather than a standalone solution in educational assessment.

Overall, this research contributes to the growing conversation on the role of AI in education, offering insights into its strengths and limitations in grading theoretical and research-based assignments. The findings provide a foundation for educators and institutions to make informed decisions about adopting AI-assisted grading systems while ensuring fairness and

quality.

7. Future Work

Future research should focus on enhancing AI-assisted grading systems to address critical challenges and expand their applicability. One key area is improving the handling of subjectivity by training AI models to better evaluate nuanced responses, such as those requiring critical thinking, creativity, or emotional intelligence. Leveraging advanced natural language understanding and sentiment analysis techniques can enhance AI performance in these contexts. Similarly, addressing fairness concerns requires identifying and eliminating biases in training datasets. Testing AI grading systems with diverse student populations and refining algorithms can ensure equitable treatment for all learners.

Another priority is improving feedback personalization. AI systems should be developed to provide tailored, detailed feedback that aligns with individual student needs, potentially incorporating adaptive learning principles to make feedback more meaningful and context-specific. Also, by combining AI's efficiency with human expertise, hybrid grading models could yield optimal results. For example, AI could handle initial evaluations while human graders refine complex or subjective assessments.

Longitudinal studies are needed to evaluate the long-term impact of AI-assisted grading on student learning outcomes and educator workload. Expanding research to include cross-disciplinary applications can also generalize findings and address unique challenges in different academic fields. Finally, integrating AI grading tools with learning management systems (LMS) can facilitate seamless, real-time grading and feedback delivery while maintaining data security and privacy.

Addressing these areas will help refine AI-assisted grading systems, making them more robust, equitable, and suitable for diverse educational contexts. These efforts ensure that AI plays a transformative role in improving education efficiency, fairness, and learning outcomes.

References

[1] R. Luckin and W. Holmes, "Intelligence Unleashed: An argument for AI in Education," 02/22 2016.

[2] D. R. Sadler, "Indeterminacy in the use of preset criteria for assessment and grading," *Assessment & Evaluation in Higher Education - ASSESS EVAL HIGH EDUC,* vol. 34, pp. 159-179, 04/01 2009, doi: 10.1080/02602930801956059.

[3] T. B. Brown *et al.*, "Language models are few-shot learners," presented at the Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 2020.

[4] R. Binns, "Fairness in Machine Learning: Lessons from Political Philosophy," presented at the Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, 2018. [Online]. Available: https://proceedings.mlr.press/v81/binns18a.html.

[5] Y. Attali and J. Burstein, "Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. Journal of Technology, Learning, and Assessment, 4(3)," *Journal of Technology, Learning, and Assessment,* vol. 4, 02/01 2006.

[6] J. Bai *et al.*, *Automated Essay Scoring (AES) Systems: Opportunities and Challenges*

for Open and Distance Education. 2022.
[7] E. Latif and X. Zhai, "Fine-tuning ChatGPT for automatic scoring," *Computers and Education: Artificial Intelligence,* vol. 6, p. 100210, 2024/06/01/ 2024, doi: https://doi.org/10.1016/j.caeai.2024.100210.
[8] C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker, *Handbook of Educational Data Mining*. 2010.
[9] B. Hamner and M. D. Shermis, "Contrasting state-of-the-art automated scoring of essays: analysis," 2012.
[10] C. Dede, "The Role of Digital Technologies in Deeper Learning. Students at the Center: Deeper Learning Research Series," 2014.
[11] B. Williamson, R. Eynon, and J. Potter, "Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency," *Learning, Media and Technology,* vol. 45, no. 2, pp. 107-114, 2020/04/02 2020, doi: 10.1080/17439884.2020.1761641.
[12] P. Montieloverall, "Designing and Conducting Mixed Methods Research. By John W. Creswell and Vicki L. Plano Clark. Thousand Oaks, CA: Sage Publications, 2007. 274
pp. $79.95 (cloth); $39.95 (paper), ISBN: 1-4129-2791-9; ISBN: 1-4129-2792-7," *Library & Information Science Research - LIBR INFORM SCI RES,* vol. 29, pp. 434-436, 09/30 2007, doi: 10.1016/j.lisr.2007.02.003.
[13] L. Cohen, L. Manion, and K. Morrison, "Research Methods in Education," *http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=011160/(100),* 01/01 2000.

Appendix

Appendix A. Sample Rubrics Used in the Study

Below are examples of the grading rubrics applied to assignments during the study.

Rubric for Theoretical Assignments:

Criterion	Description	Points
Clarity of Explanation	Clear and concise explanation of concepts.	10
Application of Concepts	Appropriate application of theoretical knowledge.	10
Logical Structure	Well-organized responses with a logical flow.	5
Use of Supporting Evidence	Incorporation of relevant examples or references.	5

Rubric for Research Paper Assignments:

Criterion	Description	Points
Research Depth	Depth and breadth of research presented.	15

Argumentation	Strength and coherence of arguments made.	10
Formatting and Citation	Adherence to formatting standards and proper citations.	5
Originality	Novelty and creativity of ideas.	10

Appendix B. Example AI Feedback

Below is an example of feedback generated by ChatGPT for a theoretical assignment:

Student Response:
"The theory of relativity suggests that time is relative and can differ based on the observer's speed and gravitational field."

AI Feedback:
"Your explanation of the theory of relativity captures the essence of time relativity well. You could provide specific examples to strengthen your response, such as the time dilation observed in high-speed travel or near massive objects. Additionally, consider elaborating on how these principles apply to real-world phenomena like GPS systems."

Appendix C. Survey Questions

Student Survey:

1. How useful did the AI-assisted grading system provide the feedback? (Scale: 1–5)
2. How fair did you perceive the AI grading system to be? (Scale: 1–5)
3. Did the feedback help you improve your understanding of the assignment? (Yes/No)
4. Do you prefer AI-assisted grading, traditional human grading, or a combination of both? Why?

Educator Survey:

1. How effective was the AI-assisted grading system in reducing your workload? (Scale: 1–5)
2. How accurate did you find the AI grading system compared to human grading? (Scale: 1–5)
3. Would you consider using AI-assisted grading for future courses? (Yes/No)
4. What improvements would you recommend for AI grading systems?

Appendix D. Sample Assignments

Sample Theoretical Assignment:

- **Question:** Explain the difference between classical mechanics and quantum mechanics and provide an example where classical mechanics fails to explain a phenomenon that quantum mechanics can.

Sample Research-Based Assignment:

- **Question:** Conduct a literature review on the impact of artificial intelligence in education. Summarize key findings and discuss the potential challenges of integrating AI tools into classrooms.

Appendix E. Statistical Analysis Details

1. **Grading Time Analysis:**
    - Mean grading time (AI vs. human):
        - AI: 10 minutes per assignment
        - Human: 30 minutes per assignment
    - Statistical test used: Paired t-test ($p < 0.05$).
2. **Grading Consistency Analysis:**
    - Inter-rater reliability (IRR):
        - AI: 0.89
        - Human: 0.82.
3. **Bias Testing:**
    - Groups tested: Native vs. non-native English speakers.
    - Results: Non-native English speakers received 8% lower average grades from AI, highlighting potential biases.