



IMPROVING THE FAIRNESS AND DEPTH OF AUTOMATED ESSAY SCORING SYSTEMS THROUGH ADVANCED MACHINE LEARNING TECHNIQUES

Fatimah Nowh Aljaroudi

2024

Contents

1. Abstract.....	2
2. Introduction	2
3. Problem Statement	3
4. Research Objectives:	3
5. Research Scope.....	4
6. Literature Review	4
6.1.Table:.....	4
6.2. Handbook of Open, Distance and Digital Education:.....	6
6.3. Automated Essay Scoring and the Deep Learning Black Box:.....	7
6.4. Reducing Manual Grading Workload through Machine Learning Techniques:.....	7
6.5. From Automation to Augmentation: Large Language Models Elevating Essay Scoring Landscape:.....	8
6.6. A Review of Deep-Neural Automated Essay Scoring Models:	8
6.7. Systematic Review of AI-Driven Feedback in Educational Psychology:	9
6.8. The Eras and Trends of Automatic Short Answer Grading:	9
6.9. Computer or human: a comparative study of automated evaluation scoring and instructors' feedback on Chinese college students' English writing:.....	10
6.10. Stacking Neural Network Models for Automatic Short Answer Scoring:	11
6.11. Toward intelligent open-ended questions evaluation based on predictive optimization:	12
7. Research Methodology:.....	13
8. Expected outcomes	15
9. Project Timeline.....	16
10. Conclusion.....	16
11. Bibliography:	17

1. Abstract

The research aim is to advance Automated Essay Scoring (AES) through pioneering machine learning techniques to improve accuracy and fairness. The study addresses a mixed-methods model, integrating both quantitative and qualitative analyses. Quantitative assessments involve the performance of newly developed ML models comparing with traditional AES systems and human raters. Qualitative feedback is gathered from educational parties to evaluate the effectiveness and fairness of the feedback provided by AES. This multifaceted model has the goal of enhancing the depth and reliability of automated scoring systems in educational settings.

Keywords: Automated Essay Scoring, Machine Learning, Natural Language Processing, Educational Technology, Fairness, Evaluation Methods

2. Introduction

Automated Essay Scoring (AES) systems have witnessed significant progression, starting from primitive manual methods in the 1960s to the use of the advanced Natural Language Processing (NLP) and Machine Learning (ML) technologies. It was first designed for simple tasks such as word counting and basic grammar checks. Now, modern AES tools are proficiently analyzing complex dimensions like context organization, thematic coherence, and argumentative structure. However, AES systems still struggle with capturing the restraints of human writing, facing difficulties like limited training datasets which can cause systematic biases and an overreliance on superficial features. These issues not only affect the reliability of AES but also its acceptance in academic environments. The challenges in the understanding the full spectrum of human expressions highlight a crucial demand for research that aims at developing more robust, equitable, and in-depth AES methodologies. This research is focusing on addressing these gaps by leveraging cutting-edge ML techniques to update the training processes and evaluation metrics, thus improving the system's ability to perform nuanced assessments that are both fair and academically rigorous.

3. Problem Statement

During the past few years, Automated Essay Scoring (AES) has significantly enhanced from basic evaluation such as word counting to sophisticated metrics using Natural Language Processing (NLP) and Machine Learning (ML). However, current AES systems still face substantial challenges, especially in evaluating the full meaning of human expressions, like evaluating argument strength, creativity, and nuanced content. These limitations are often encountered because of the inadequate training datasets, which leads to potential systematic biases and over-reliance on certain features. These issues compromise the fairness and accuracy of AES, affecting academical grading and student feedback. Current challenges highlight the needs of further research to develop more advanced AES systems that can handle different academical contexts and content forms effectively.

4. Research Objectives:

- To analyze the effect of the diversity of training dataset on the performance and fairness of AES systems.
- To evaluate the effectiveness of existing AES technologies in processing different forms of assignments and feedback, focusing on strengths and weaknesses.
- To develop a new ML model that includes a diverse range of linguistic and thematic elements to enhance the understanding and scoring complex assignments.
- To propose improvements in current AES models that reduce biases and over-reliance on non-essential features.
- To validate the proposed ML model against traditional and current AES systems through thorough testing with diverse assignments datasets.

5. Research Scope

The research scope for the study on improving Automated Essay Scoring (AES) systems through advanced machine learning techniques emphasizes on several key areas:

1. **Technological Enhancement:** Explore and integrate the latest advancements in Natural Language Processing (NLP) and machine learning algorithms to enhance the accuracy and depth of AES systems.
2. **Fairness and Bias Reduction:** Investigate methods to mitigate biases inherent in current AES systems, ensuring fair scoring across diverse student populations and writing styles.
3. **Evaluation of Effectiveness:** Assess how different machine learning models influence the performance and reliability of AES when applied to varied academic writing formats.
4. **Training Data Diversity:** Examine the impact of using a more diverse set of training data on the performance and fairness of AES systems.
5. **Comparative Analysis:** Conduct comparative studies between improved AES models and traditional human grading to validate effectiveness and reliability.

This scope will be a guideline for the research methodology, ensuring a comprehensive analysis of the potential of advanced machine learning techniques to transform AES systems.

6. Literature Review

6.1. Table:

Year	Author(s)	Methodology	Key Findings	Weaknesses/Limitations
2023	D. Ifenthaler et al.	Review of AES technologies, NLP, Machine Learning	Evolution of AES, advancements in NLP and machine learning, identification of limitations	Challenges in depth of analysis, potential biases
2020	Kumar, V.S., Boulanger, D.	Deep learning models on ASAP dataset	High QWK scores, improved interpretability of AES	Limited dataset diversity

2022	Weegar, Idestam-Almquist	Two-step ML grading with manual review and clustering	64-74% reduction in grading workload	High dependency on manual review for accuracy
2024	Xiao et al.	Experiments with LLMs	Effective use of LLMs in scoring essays	Similar content, limited new insights
2021	Masaki Uto	Comprehensive review of DNN-based AES models	DNN-AES models offer an effective alternative to traditional AES by eliminating the need for manual feature engineering. The review categorizes and details models for four types of AES tasks.	The paper does not specify weaknesses or limitations.
2022	Sebastian Gombert, Aron Fink, Tornike Giorgashvili, Ioana Jivet, Daniele Di Mitri, Jane Yau, Andreas Frey, Hendrik Drachsle	Controlled study comparing AI-driven feedback (using NLP models GBERT and T5) with traditional feedback methods on student essays.	AI-driven feedback was more helpful and enhanced student reflection and understanding compared to traditional feedback.	Challenges in accurately interpreting diverse responses; variability in feedback reception and dependence on comprehensive training data.
2014	Steven Burrows, Iryna Gurevych, Benno Stein	Historical analysis, component analysis, trend analysis, categorization of ASAG systems across five eras.	Identified five eras in ASAG research, emphasizing the latest trend towards evaluation and standardization across the field.	Historical focus; lacks empirical contributions for newer, rapidly evolving ASAG methods.
2022	Huimei Chen, Jie Pan	Mixed methods, comparison of AES system "Aim Writing" with human instructors' feedback, data triangulation	Aim Writing provides timely grammar corrections, improving lower-level learners' grammar and vocabulary, but lacks depth in feedback on content and organization.	AES feedback is less personalized and context-sensitive compared to human feedback, leading to potential misinterpretations by students.

2021	Rian Adam Rajagede & Rochana Prih Hastuti	Stacking neural networks and XGBoost, SMOTE for data balancing, TPE for hyperparameter tuning	The stacking model with SMOTE achieved a combined F1-score of 0.821, outperforming previous models in the Ukara 1.0 Challenge. The approach handles imbalanced data effectively.	The study relies on a small dataset, which may limit the generalizability of the results.
2023	Faisal Jamil & Ibrahim A. Hameed	Deep neural networks, particle swarm optimization, gradient descent, semantic similarity using WordNet and Growbag	The ISE system achieved a high accuracy (97.54%) in grading open-ended questions, outperforming traditional methods by incorporating dynamic optimization.	Limited to a single dataset; potential challenges in generalizing across different academic contexts.

6.2. Handbook of Open, Distance and Digital Education:

The book chapter written by D. Ifenthale elaborated this particular area. The chapter aimed to go through; the evolution, operational mechanisms, and current challenges of automated essay scoring systems. They were designed to automate the evaluation of text-based essays. To achieve this goal, the writer reviewed AES technologies and methodologies, focusing mainly on improvements in natural language processing (NLP) and machine learning (ML) which develop the capability of such systems to evaluate and grade the written content.

The chapter addresses three important findings. First, AES systems have enhanced significantly since the 1960s, starting from first systems such as Project Essay Grader (PEG) that focused on basic metrics like counting words and grammar check. Also, current developments have engaged more complicated NLP and ML technologies, allowing for a more detailed analysis of assignments. For example: content, style, organization, and techniques. In spite of these advancements, AES systems still facing issues related to the ability of capturing the full depth and nuance of human viewpoints, especially in argument strength, creativity, and deeper content analysis.

The author effectively outlines the advantages of AES systems, which includes their capability to reduce the work load of grading and provide effective feedback. On the other hand, it critically highlights the weaknesses as well, including the disadvantage of over-reliance on non-

important features and some potential of systemic biases. This would be encountered if the training data are not sufficient enough. (Zawacki-Richter & Jung, n.d.)

6.3. Automated Essay Scoring and the Deep Learning Black Box:

The research paper written by Vivekanandan S. Kumar & David Boulanger presents a detailed exploration of Automated Essay Scoring (AES), by operating deep learning technologies. The main goal of this paper is to investigate the accuracy and effectiveness of automated essay scoring according to a

grading rubric. The method used was introduces dataset of a manual graded essay, then applying seep neural networks to predict rubric scores. The deep learning model showed a higher QWK than the previous AES systems, including a significant enhancement comparing with human assessment.

The study focused on the interpretability of AES systems, proposing that transparency in how grades are measured could improve the acceptability and reliability of automated scoring, which elaborate the novelty of this study.

The strength of this scientific paper is that it provides a detailed analysis automated essay scoring by using significant dataset. However, the limitations of using a single source for the dataset considered as a weakness for this paper.(Kumar & Boulanger, 2021)

6.4. Reducing Manual Grading Workload through Machine Learning Techniques:

This article investigates the major load of manual grading in educational field, particularly in computer science courses.

The main goal of this study is to determine the integration of machine learning (ML) to reduce the workload of manual regular grading, without reducing the quality of grading.

The method used by the researchers involves a two-step automated grading process, the initial step of the automated grading process is collecting a subset of a manual graded answers as a

data to train ML models. Then the trained ML models start to grade the remaining of student answers. This article achieved a workload reduction by 64% to 74% compared to manual grading, increasing the efficiency and accuracy of grading.

Examining the practical application of ML in educational grading, providing detailed analysis of different strategies to implement automated grading system, are considered as a strength of the essay. While the limitation of this study that it mostly relies on the accuracy of ML models, and it can't be generalized on all course contexts.(Weegar & Idestam-Almquist, 2024)

6.5. From Automation to Augmentation: Large Language Models Elevating Essay Scoring Landscape:

Xiao and his colleagues examine specifically how the two models of GPT-3.5 and GPT-4 can improve the efficiency and accuracy of grading by comparing novice and expert graders. The objective of this article is the evaluation of the effectiveness of both GPT-3.5 and GPT-4 in automating scoring process using the experiment method in this essay to examine LLMs on datasets and expert graders. This essay emphasizes the significant enhancement of grading accuracy and efficiency, illustrating two potentials solutions, standalone grading and support grading system to support humans.

Moreover, it is possible to summarize the strengths and weaknesses into two main points to each one. The strengths are, detailed empirical analysis and incorporation of human-AI interpretations, while the weaknesses are, limited demographic focus and potential bias in AI interpretations.(Changrong Xiao et al., 2024)

6.6. A Review of Deep-Neural Automated Essay Scoring Models:

This article reviews various deep learning models used in automated essay scoring, mainly concentrating on their effectiveness across different grading tasks. The main objective of this article is to review different deep learning models used in automatic essay scoring, using critical reviewing the literature and absorbing on the effectiveness of Deep Neural Networks (DNNs)

and Recurrent Neural Network (RNNs). It is also illustrated the promising future in the efficiency and accuracy of essay scoring using the DNNS.

The Comprehensive overview and critical analysis of DNNs and RNNs models specifically, are the strengths for this article. On the other hand, the weaknesses are, potential reliability issues and confined to a limited educational format. (Gombert et al., 2024)

6.7. Systematic Review of AI-Driven Feedback in Educational Psychology:

The cases study was implementing an AI-driven feedback system for German-speaking university students in an educational psychology course. The system technology used is the Natural Language Processing (NLP), GBERT and T5 models. It is used to provide detailed and personalized feedback on student assignments regarding in related to the learning tips provided to the students. The effectiveness of this evolutionary AI-driven feedback was compared with assessor's feedback methods. It has been shown that the new technology was recognized as more helpful by students, enhancing their understanding and reflective practices. (Uto, 2021)

6.8. The Eras and Trends of Automatic Short Answer Grading:

Automatic Short Answer Grading (ASAG) is a continually improving field which focuses on the computational evaluation of short, natural language answers to objective questions. ASAG systems goal is to highlight the challenges of manual grading, like subjectivity, inconsistency, and the time-intensive nature of the process. During the past years, the field had a significant growth, with numerous methodologies being evolved and tested. This article focused on the historical improvements on this area.

The field of ASAG had the journey of development starting from concept mapping (1996-2002) to information extraction (2003-2008), corpus-based methods (2004-2010), and machine learning (2008-Present). Although each era brought improvements to this field, machine learning is currently leading this evolution because of its generalization capabilities and ability to handle unseen new data.

Many evaluation models have been used. Kappa statistics and Pearson correlation were the most common used. The key theme across studies was the negotiation which model is best, rule-based methods, which better for repeated assessments, or statistical methods, that is better for generalization.

Another area of improvement was the use of Natural Language Processing Techniques from basic tokenization to advanced semantic analysis. Considering the effectiveness of ASAG systems relies on the choice of these techniques.

The most highlighted challenge which we can find on almost all of the related articles is the limited availability of open datasets. This limitation has restricted meaningful comparisons and reliability among studies.

An important point to be addresses is that the field is still struggling in balancing interpretability with the need for generalization and scalability. The Early rule-based systems were more interpretable, however machine learning models proved more effective in handling unseen data.

Choosing the ASAG methodology should be context-dependent. Rule-based systems are effective for repeated assessments, while machine learning models excel in dynamic and diverse environments.(Burrows et al., 2015)

6.9. Computer or human: a comparative study of automated evaluation scoring and instructors' feedback on Chinese college students' English writing:

The literature on automated feedback systems for English as a Foreign Language (EFL) learners emphasizes the evolving role of Automated Essay Scoring (AES) systems like Aim Writing in conjunction with traditional teacher feedback. In tradition, teacher was the base of writing instruction and providing comprehensive guidelines on grammar and content organization. The growing number of students and the emphasis on grammar in countries like China have made it difficult for teachers to provide detailed and timely feedback. This has led to the adoption of AES systems, which offer quick, accessible comments on students' writing, specifically on

grammatical errors. This helped in improving the grammatical competence of lower-level English learners.

Despite the advantages of AES systems in providing timely feedback, they are still having significant limitations, such as evaluating the higher-order aspects of writing. For instance, idea development, argument organization, and overall coherence. AES systems cannot provide context-sensitive feedback, that is critical for improving the logical flow and content of essays. This limitation is highlighted as a big issue for advanced learners, who require nuanced feedback to develop their writing skills.

To overcome these challenges, it has been recommended to use a hybrid feedback model that combines the strengths of both AES systems and traditional teacher. In this model, AES systems can be used to provide initial feedback on grammar and vocabulary, which allows the students to make some corrections before receiving detailed and personalized feedback from teachers on the other aspects. This model not only improves the efficiency of the feedback process but also ensures that students receive the comprehensive guidelines they need to enhance their writing proficiency. As AES technology is under development, it will be important for future research to focus on improving these systems' ability to assess the logical flow and content of essays. (Chen & Pan, 2022)

6.10. Stacking Neural Network Models for Automatic Short Answer Scoring:

In the article, the authors navigate through the application of stacking models, joining neural networks and XGBoost. This is to enhance the efficiency of automated short answer scoring systems. This approach had some key challenges, like data imbalance and the complexity of hyperparameter tuning. The study illustrates a significant enhancement in scoring accuracy, achieving a combined F1-score of 0.821 on the Ukara 1.0 Challenge dataset, by using the Synthetic Minority Over-sampling Technique (SMOTE) for balancing data and the Tree-structured Parzen Estimator (TPE) for optimizing hyperparameters.

The research acknowledged the effectiveness of the stacking model in outperforming single models, such as neural networks or XGBoost alone. Especially, in handling the variant student

responses. The study's methodology included the use of sentence embeddings and word count features to represent student responses, providing an in-depth approach to the classification process. The model ability to deal with class imbalance was enhanced more by the implementation of SMOTE, particularly in cases where the number of correct responses was significantly lower than incorrect ones.

Similar to other studies, this article highlights certain challenges and limitations. Such as the dependency on a small dataset, which may affect the generalizability of the results. Moreover, the research recommends upcoming exploration of multi-task learning methods and different features for different types of questions to develop the performance of automated scoring systems(Rajagede & Hastuti, 2021)

6.11. Toward intelligent open-ended questions evaluation based on predictive optimization:

The article "Toward intelligent open-ended questions evaluation based on predictive optimization" by Faisal Jamil and Ibrahim A. Hameed introduces an innovative Intelligent Student Evaluation (ISE) system designed for grading text-based questions. This system utilizes advanced deep learning and optimization techniques to acknowledge the limitations of traditional grading methods, like inflexibility and limited accuracy. The ISE system integrates semantic similarity analysis using both the WordNet library and Growbag dataset to evaluate student answers, improving the precision of grading by considering the question's type, structure, and related keywords.

The research emphasizes the use of particle swarm optimization (PSO) and gradient descent (GD) within a deep neural network (DNN) framework to automate the grading process. The study was built on real exam data from Norwegian University students, examining the system's ability to adapt to dynamic cases, thus outperforming static, pre-trained models. The authors conducted extensive experiments to validate the system's effectiveness, demonstrating significant improvements in grading accuracy, precision, recall, and F1-score, compared to other machine learning methods.

The study also addressed the limitations which is also the case on other studies, such as the reliance on a single dataset and the challenges of generalizing results to variant academic areas. Future researchers have to consider the incorporation of more diverse datasets and the exploration of additional optimization algorithms to update the grading methodology furthermore. The ISE system shows a significant advancement in automated grading, offering a measured and flexible solution for educational parties. (Jamil & Hameed, 2023)

The study illustrates the ability of AI to deliver personalized and effective feedback. However, and similar to other studies, challenges are the AI's protentional to interpret different student responses accurately and the variability in feedback reception.

The main important challenge addressed on AES systems reviews is dealing with scoring rubric ambiguity and grading subjectivity. This study suggests that the advances NL processing capabilities of the LLMs would solve this issue via providing trusted and personalized feedback.

The exploration of Automated Essay Scoring (AES) systems across the studies reviewed in this essay shows the increase demands on the advanced technologies including ML, deep learning, and Large Language Models (LLMs) to develop the efficiency of essay grading. They provided a detailed understanding of how these methods enhancements have reform the grading systems. This was mostly elaborated on both D. Ifenthaler's historical and operational analysis of AES and the application of deep learning models in AES as detailed by Vivekanandan S. Kumar & David Boulanger.

Although the studies demonstrated the importance of AES to educational assessment revolutionary, they also discussed multiple challenges. This can be summed up on one main issue which are the biases encountered due to inadequate training datasets, over-reliance on specific features, and the generalizing across multiple educational criteria. Furthermore, future improvements would be needed on areas like ML accuracy and training data diversity.

7. Research Methodology:

This study will adopt a mixed-methods approach to examine the effectiveness and fairness of Automated Essay Scoring (AES) systems boosted by advanced machine learning techniques.

1. Quantitative Analysis:

- **Data Collection:** Collect diverse datasets including essays from different academic levels and subjects to guarantee a wide representation of language use and essay styles.
- **Experimental Design:** Implement various AES models, including both current used systems and newly developed models that integrate advanced ML algorithms to reach the aim of this study.
- **Statistical Methods:** Use correlation analysis to compare AES system scores and human examiners scores to test the consistency and fairness. Advanced statistical examinations, such as regression analysis, will test the influence of training dataset diversity on AES performance.

2. Qualitative Analysis:

- **Interviews:** Conduct semi-structured interviews with educators and students to collect feedbacks on the perceived fairness and effectiveness of AES feedback.
- **Focus Groups:** Manage discussions with educational technology experts to widen the knowledge of the potential biases and limitations of current AES systems and the improvements achieved with the new models.

3. Model Development:

- **Algorithm Testing:** Develop and refine machine learning models using natural language processing to enhance the AES's ability to understand complex human expressions and reduce limitations.
- **Validation:** Implement cross-validation techniques to test the robustness and reliability of the new AES models across various datasets.

4. Ethical Considerations:

- Ensure data privacy and consent for all participants involved in the study, particularly during data collection and interviews.

5. Data Analysis Tools:

- Utilize software such as SPSS for statistical analysis

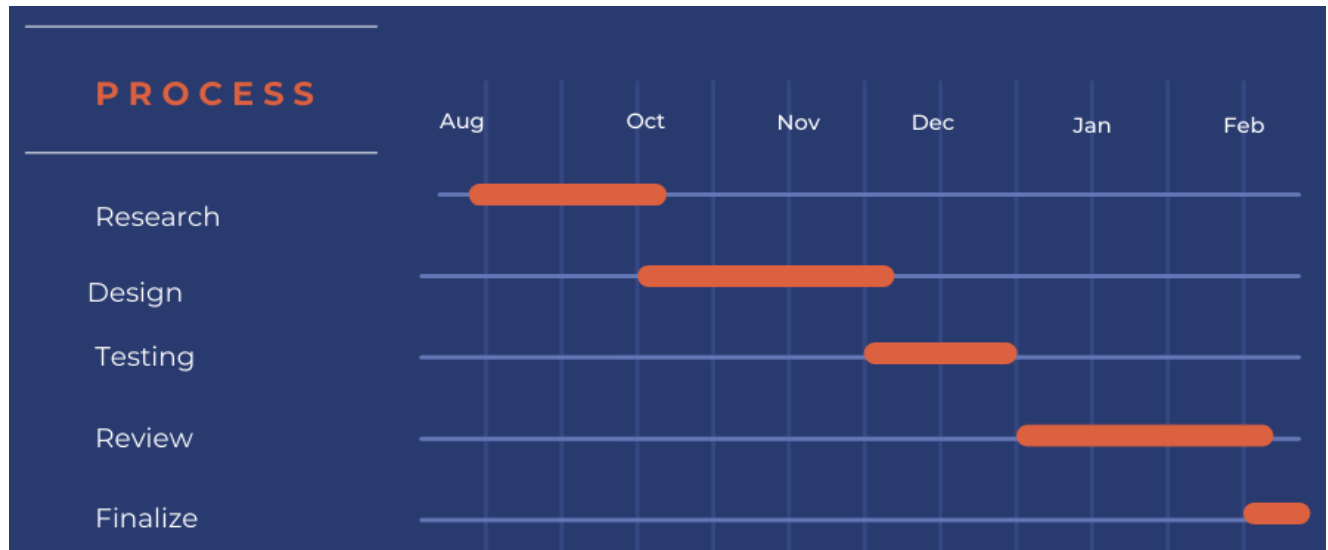
The research goal is to advance the capabilities of AES systems by integrating the mentioned methodologies and making them more adaptive, fair, and capable of handling the complex distinctions of student writing.

8. Expected outcomes

The expected outcomes of the research on enhancing Automated Essay Scoring (AES) systems using advanced machine learning techniques include:

1. **Improved Accuracy:** Enhanced accuracy in essays scoring by integrating machine learning models which leads to closer matching to human grading ideals.
2. **Increased Fairness:** Reduction of limitations in AES systems, ensuring equitable scores across various writing styles.
3. **Enhanced Understanding:** Deeper analysis abilities of AES systems to understand complexity of the writing elements such as argument strength, creativity, and nuanced content.
4. **Robust Validation:** Empirical evidence supporting the superiority of new AES models over traditional methods through rigorous testing and comparison.
5. **Practical Implementations:** Recommendations for integrating advanced AES systems in educational settings to improve grading efficiency and feedback quality.

9. Project Timeline



10. Conclusion

In conclusion, the benefits of integrating AES into educational systems would be summarized on reducing instructors' workload and providing timely, personalized feedback. However, it is suggested the need of balancing such technologies with human oversight to guarantee the fairness and comprehensiveness in grading.

11. Bibliography:

- Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. In *International Journal of Artificial Intelligence in Education* (Vol. 25, Issue 1, pp. 60–117). Springer New York LLC. <https://doi.org/10.1007/s40593-014-0026-8>
- Changrong Xiao, Wenxing Ma, Sean Xin Xu, & Kunpeng Zhang. (2024). *From Automation to Augmentation: Large Language Models Elevating Essay Scoring Landscape*. Beijing Xicheng Educational Research Institute.
- Chen, H., & Pan, J. (2022). Computer or human: a comparative study of automated evaluation scoring and instructors' feedback on Chinese college students' English writing. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1). <https://doi.org/10.1186/s40862-022-00171-4>
- Gombert, S., Fink, A., Giorgashvili, T., Jivet, I., Di Mitri, D., Yau, J., Frey, A., & Drachsler, H. (2024). From the Automated Assessment of Student Essay Content to Highly Informative Feedback: a Case Study. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-023-00387-6>
- Jamil, F., & Hameed, I. A. (2023). Toward intelligent open-ended questions evaluation based on predictive optimization. *Expert Systems with Applications*, 231. <https://doi.org/10.1016/j.eswa.2023.120640>
- Kumar, V. S., & Boulanger, D. (2021). Automated Essay Scoring and the Deep Learning Black Box: How Are Rubric Scores Determined? *International Journal of Artificial Intelligence in Education*, 31(3), 538–584. <https://doi.org/10.1007/s40593-020-00211-5>
- Rajagede, R. A., & Hastuti, R. P. (2021). Stacking Neural Network Models for Automatic Short Answer Scoring. *IOP Conference Series: Materials Science and Engineering*, 1077(1), 012013. <https://doi.org/10.1088/1757-899x/1077/1/012013>
- Uto, M. (2021). A review of deep-neural automated essay scoring models. In *Behaviormetrika* (Vol. 48, Issue 2, pp. 459–484). Springer Japan. <https://doi.org/10.1007/s41237-021-00142-y>

- Weegar, R., & Idestam-Almquist, P. (2024). Reducing Workload in Short Answer Grading Using Machine Learning. *International Journal of Artificial Intelligence in Education*, 34(2), 247–273. <https://doi.org/10.1007/s40593-022-00322-1>
- Zawacki-Richter, O., & Jung, I. (n.d.). *Handbook of Open, Distance and Digital Education*.