

Algorithmic Gatekeeping Language Ideologies in Automated Essay Scoring Systems

MA: Nadia Thamer Abu Salal
Al-Ameed University
nadiathamer2023@gmail.com

Abstract

This study explores the influence of language ideologies on automated essay assessment through the notion of algorithmic gatekeeping. As the involvement of Artificial Intelligence (AI) and Large Language Models (LLMs) becomes a part of the current process of evaluating students' essays, the issues of fairness, openness, and inclusiveness arise. Thus, the current study employs a qualitative research methodology with Critical Discourse Analysis (CDA) as the core analytical technique to explore a purposive sample of essays evaluated via modern AES systems. Specifically, this paper is concerned with linguistic properties that are appreciated by automated scoring algorithms and corresponding language ideologies that operate within this system. The results of the study show that AES systems always value standardized varieties of academic language, correct grammar, high levels of lexical density, organization, and complex syntax. Such linguistic criteria can be viewed as algorithmic gatekeeping practices since they privilege one kind of linguistic behavior over another. Therefore, it can be said that the AES system is not an impartial tool of academic assessment, but rather an instrument used for reproducing language ideologies.

Keywords: *Automated Essay Scoring (AES), Algorithmic Gatekeeping, Language Ideology, Critical Discourse Analysis, Artificial Intelligence, Educational Assessment, Linguistic Diversity, Large Language Models (LLMs).*

أيدولوجيات لغة الرقابة الخوارزمية في أنظمة تقييم المقالات الآلية
م.م نادية تامر ابوصلال
جامعة العميد

الملخص

تستكشف هذه الدراسة تأثير الأيدولوجيات اللغوية في تقييم المقالات آلياً من خلال مفهوم حراسة البوابة الخوارزمية. ومع ازدياد حضور الذكاء الاصطناعي ونماذج اللغة الكبيرة (LLMs) في عملية تقييم مقالات الطلبة، تبرز قضايا تتعلق بالعدالة والشفافية والشمولية. لذلك، اعتمدت الدراسة الحالية منهجاً بحثياً نوعياً، واتخذت من تحليل الخطاب النقدي أداة تحليلية رئيسة لفحص عينة قصدية من المقالات التي قُيِّمت عبر أنظمة حديثة للتصحيح الآلي للمقالات. وتركز هذه الورقة على الخصائص اللغوية التي تقدرها خوارزميات التصحيح الآلي، وما يقابلها من أيدولوجيات لغوية فاعلة داخل هذا النظام. وتُظهر نتائج الدراسة أن أنظمة التصحيح الآلي للمقالات تُعلي باستمرار من قيمة الأنماط المعيارية للغة الأكاديمية، وصحة القواعد، وارتفاع الكثافة المعجمية، والتنظيم الجيد، والتركيب النحوي المعقد. ويمكن النظر إلى هذه المعايير اللغوية بوصفها



ممارساتٍ لحراسة البوابة الخوارزمية، لأنها تمنح الأفضلية لنمط لغوي معيّن على حساب أنماط أخرى. ومن ثمّ، يمكن القول إن نظام التصحيح الآلي للمقالات ليس أداة محايدة في التقييم الأكاديمي، بل أداة تُسهم في إعادة إنتاج الأيديولوجيات اللغوية.

الكلمات المفتاحية: التصحيح الآلي للمقالات، حراسة البوابة الخوارزمية، الأيديولوجيا اللغوية، تحليل الخطاب النقدي، الذكاء الاصطناعي، التقييم التربوي، التنوع اللغوي، نماذج اللغة الكبيرة.

1. Introduction

The rapid evolution of artificial intelligence and natural language processing has revolutionized the process of conducting assessments within the field of education, especially due to the introduction of Automated Essay Scoring (AES). In general, AES implies the use of computational technology capable of assessing and scoring written essays using the pre-defined parameters of language, rhetoric, and statistics. With the recent progress in Large Language Models, in particular, systems based on Generative Pre-trained Transformer, the potential of automated evaluation of essays has been further enhanced by allowing a comprehensive analysis of coherence, argumentation, grammar, and discourse structures (Li & Liu, 2024, p. 2; Wang & Gayed, 2024, p. 211).

Algorithmic gatekeeping is becoming an important idea in discussions regarding education technology. Algorithmic gatekeeping involves the use of algorithms to control access to opportunities, resources, and institutional acknowledgment through evaluations that determine people's success. Within an educational context, the AES system serves as a gatekeeper since it determines the grades, placements, admissions, and writing proficiency. As such, the linguistic assumptions present within the system can have an impact on what languages are privileged and what languages are devalued.

Ideologies of language offer an important conceptual tool in understanding such processes. An ideology of language includes social constructions related to how people perceive language use, appropriateness, and correctness. Such ideologies tend to favor standardized forms of languages while viewing other linguistic expressions as inadequate or problematic. Given that AES models rely on large sets of data to learn how to score written texts, these models can reinforce language ideologies by recognizing some forms of written language and rejecting others that do not conform to linguistic conventions (Tang et al., 2024, p. 4184). Recent research also indicates that AI-powered grading systems can show inconsistencies in their assessment practices and fail to correspond with human evaluations of written discourse (Jiao et al., 2026, p. 4; Schaller et al., 2024, p. 211).

In light of the rise in the use of assessment technologies powered by AI in academic institutions, it is crucial to undertake a rigorous analysis of the ideological foundations of automated scoring. As such, the purpose of this research is to investigate the ideology of language as it appears in the AES technology and how algorithmic gatekeeping affects the process of assessing written discourse.

1.1 Problem Statement

Even with advancements made in technology in AES systems, there still exist issues related to how they address linguistic diversity and the possibility of reinforcing dominant language ideologies through use. Most of the AES systems rely on the development of algorithms using datasets that reflect language that is standard and conforms to normative conventions of academic writing (Schaller et al., 2024, p. 212). This means that the AES systems will likely give high scores to essays that meet the dominant linguistic requirements while discriminating those that utilize other discourse patterns.

The second problem involves inconsistency in reliability, validity, and alignment with human raters. It has been established in current researches that the scoring systems based on LLMs are inconsistent and may be unfair to writers due to lack of alignment with human raters (Lee et al., 2024, p. 3; Jiao et al., 2026, p. 5). Even though AES technologies have become popular in educational contexts, there is no research focusing on how language ideologies operate in the AES systems.

1.2 Research Objectives

1. To analyze the language ideologies that are present in Automated Essay Scoring systems.
2. To explore the impact of algorithmic gatekeeping on the assessment of students' writing.
3. To determine the types of linguistic and stylistic preferences that are favored by AES algorithms.
4. To assess the consequences of AES-based assessments on linguistic diversity and equity in education.

1.3 Research Questions

1. *What are the language ideologies that are contained in AES systems today?*
2. *How are AES systems used as an algorithmic means of gatekeeping in the field of education?*

3. *What linguistic factors and discourse practices does an automated scoring system favor?*
4. *How might AES technology impact linguistic diversity and equality in education?*

2. Literature Review

2.1 Language Ideologies and Algorithmic Gatekeeping in Automated Essay Scoring Systems

As more and more attention has been paid to using AI for assessing students' academic performance, there has been an increasing scholarly interest in Automated Essay Scoring (AES) systems. These automated systems analyze written texts using different linguistic, syntactic, semantic, and discursal criteria through computational algorithms. With the help of new LLMs, AES can now be used to assess the quality of writing in a much better way (Hao et al., 2024, p. 18). While AES technology seems like an efficient tool for evaluating students' academic writing, scholars have criticized the ideological assumptions built into algorithmic assessment.

In sociolinguistics, language ideology is understood as socially and culturally informed ideas about language, linguistic propriety, and communicative ability (Flores & Rosa, 2023, p. 65). These ideologies tend to favor dominant standard languages while other, non-standard, multilingual, or culturally diverse language varieties remain devalued. In educational settings, particular language ideologies inform judgments about how people should speak and write. AES systems that use large language models trained on extensive data sets based on dominant linguistic norms can be said to reproduce such ideologies.

A number of researchers have pointed out that automated scoring systems appear to favor essays that use typical grammar structures, vocabulary sets, and rhetorical strategies accepted at the institutional level. According to Li & Liu (2024, p. 7), essays that do not follow the set academic language conventions but are still effective and understandable received lower grades from LLM-based scoring tools. Furthermore, Wang & Gayed (2024, p. 223) noted that automated essay scorers tended to give high grades to those essays that exhibited linguistic properties of the dominant discourse community.

The notion of algorithmic gatekeeping can be helpful in analyzing this practice. The concept of algorithmic gatekeeping implies an automated regulation of people's access to resources, institutional acceptance, and other forms of



opportunities that are controlled by algorithms (Hao et al., 2024, p. 21). AES is a typical example of such a gatekeeper that controls how students get grades, where they are going to study in the future, whether they deserve scholarships, and many other aspects related to the perception of their writing skills.

Fairness and biases have recently become major issues within the context of automatic assessment. According to Schaller et al. (2024), some AES algorithms have shown unequal scoring practices towards learners, implying that algorithmic assessments are not necessarily free of bias. In the same way, Jiao et al. (2026) showed marked differences between the performance of various LLM scoring algorithms, implying that there is ideology behind each algorithm based on how it was developed and evaluated.

2.2 Gaps in Research on Language Ideologies and Automated Assessment

Despite the substantial amount of academic work that has contributed to the understanding of AES system performance, certain areas that are in need of investigation still exist. Firstly, the overwhelming number of recent works emphasizes such factors as scoring reliability and validity, while paying relatively less attention to ideological presuppositions underlying the scoring mechanisms (Lee et al., 2024, p. 5). As a result, considerations relating to language ideology are often not relevant for AES discourse.

Secondly, many publications consider the concept of bias from a purely mathematical perspective, neglecting sociolinguistic structures that impact the algorithms' decision-making process. Even though researchers have highlighted scoring discrepancies, only a few works consider how dominant ideologies get incorporated into training data sets, features selection and scoring criteria (Schaller et al., 2024, p. 218).

The third limitation is the lack of research linking Critical Discourse Analysis (CDA) and research into automated assessment tools. Current studies either use quantification and machine learning techniques or employ psychometrics to analyze these systems. As a consequence, there is still very little attention paid to the ideological dimension of algorithms employed for automatic grading (Hao et al., 2024, p. 25).

Finally, multilingual and multicultural writing practices have received inadequate attention in current AES research. Indeed, the majority of studies available today are devoted to English-speaking educational environments and their academic

traditions, leaving aside an analysis of linguistically diverse societies and their educational equality issues (Li & Liu, 2024, p. 9).

All these problems suggest that further research into AES should be conducted within the framework of critical sociotechnological research into automatic assessment of writing.

2.3 Theoretical and Methodological Innovations

New theoretical approaches in AI research, critical algorithms studies, and sociolinguistics can help fill some of the research gaps identified above. One such approach is the combination of language ideology theory and algorithmic governance. This method allows researchers to analyze the way ideologies associated with language operate in computational systems and affect education (Kerssens & van Dijck, 2025, p. 44).

The second notable theoretical trend is related to the use of Critical Discourse Analysis (CDA) in the context of technology. Many researchers engaged in discourse studies have come to realize that algorithms play the role of discursive actors and have a great impact on social, institutional, and power processes (Bouvier & Machin, 2024, p. 118).

From a methodological perspective, researchers have started utilizing an interdisciplinary framework involving discourse analysis, computational auditing, testing of algorithmic fairness, and analysis of algorithmic output. This allows for uncovering not only the observable bias patterns but also the ideological underpinnings behind them (Jiao et al., 2026, p. 10). Moreover, through XAI technologies, there is a promising opportunity to explore AES algorithms' reasoning and determine what language aspects receive preference (Tang et al., 2024, p. 4188).

All these new theories and methods help to better understand AES systems as linguistic, cultural, and political technologies.

2.4 Bridging the Gap

The current research attempts to fill these gaps via the investigation of AES systems using a combination of the perspectives of language ideology and algorithmic gatekeeping. Instead of limiting the examination of such systems to their technical parameters only, it analyzes the ideologies concerning language that underpin the automated scoring processes and influence the judgments of writing proficiency.

At the same time, the present research contributes to the burgeoning body of critical research on artificial intelligence by applying sociolinguistic, discourse analysis, and algorithm studies' paradigms to its investigation. In this way, it aims to analyze the ideology-related mechanisms of prioritization of particular linguistic practices and discrimination against others by AES systems.

In terms of methodology, this study goes one step beyond regular psychometrics. It includes AES algorithm and scoring rules, but also explores linguistic evaluation patterns of AES models.

Finally, this paper is intended to make a contribution to discourse regarding equity and educational equality issues in automated test scoring. The overall purpose is to promote social responsibility in AI-powered assessment scoring.

3. Methodology

3.1 Research Design

The study will be based on qualitative research using both CDA and algorithm research. Qualitative research will be ideal in this context as it involves in-depth examination of discourses, power relations, and language ideologies (Creswell & Creswell, 2023, p. 43). As opposed to numerical analysis, the study seeks to examine how the algorithms evaluate specific languages. This research is both descriptive and interpretive. The primary aim of the study is to conduct sociotechnical analysis of AES algorithms as gatekeepers. The investigation of AES technologies through a qualitative framework involves identifying linguistic norms and assumptions embedded in the automated scoring process.

Methodology used in the research is based on the document-based textual analysis. This approach concentrates on the analysis of essays graded by automated assessment tools, along with the grades and feedback provided by them. It gives the researcher an opportunity to uncover recurring patterns within the system of assessment as well as ideology behind it.

3.2 Data Collection

The data to be used for this study will be the essays analyzed using Automated Essay Scoring systems in contemporary times. Purposive sampling is a suitable methodology because it facilitates the collection of data that is pertinent to the research goals (Creswell & Creswell, 2023, p. 169).

The corpus will include around 30 to 50 essays that employ an argumentative or expository style written by writers from different linguistic backgrounds. These essays will be subjected to AES software that utilizes either a Large Language Model or natural language processing technologies. Data collection will involve obtaining the essays, the numerical scores assigned by the AES program, as well as any written comments provided by the AES software.

The selection process will seek to ensure linguistic variation among the chosen essays. Linguistic variation will cover vocabulary usage, grammar, rhetorical style, and other stylistic devices. Having such linguistic variation in the data set helps to determine which specific linguistic factors are favored or disfavored by AES programs.

3.3 Data Analysis

Analysis of the collected data is conducted through thematic discourse analysis and Critical Discourse Analysis in a qualitative way. This analysis entails some steps. The first step involves analyzing the set of essays and the automated evaluation scores given to identify the repeated patterns in the language used and the evaluative tendencies.

In the second step, the researcher analyses how certain features in the language affect the scores. Specifically, the researcher focuses on lexical items, grammar, organization of discourses, coherence devices, and other rhetorical conventions stressed in the scoring system.

The use of thematic coding allows for the identification of any evaluative categories which may emerge from the automated evaluation process. Such categories can be those of grammar, vocabulary usage, organization, style, and conformity to established norms of writing. These categories are then interpreted by the researcher within sociolinguistic and ideological context.

Once themes have been analyzed, critical discourse analysis will be used to determine how the algorithmic evaluations construct particular definitions of good writing. Using this method, the study will investigate how the use of automated scoring systems reproduces language ideologies and engages in the practice of gatekeeping education (Bouvier & Machin, 2024, p. 121).

3.4 Analytical Framework



The analytical model consists of Fairclough's triadic approach to Critical Discourse Analysis in conjunction with theories of algorithmic gatekeeping and language ideology.

As per Fairclough (2015, p. 58), three levels of discourse analysis include the following: textual analysis, discursive practice, and social practice. In the textual analysis, the present research focuses on analyzing the linguistic characteristics and evaluative language used in automated feedback. Within the framework of the discursive practice, the study aims at studying the formation and distribution of certain criteria of writing excellence by automated essay scoring systems. As regards the level of social practice, it will consider the wider socio-cultural significance of automated evaluation processes.

An analytical concept of language ideology may help to better understand the way certain beliefs about linguistic correctness and academic language are incorporated into algorithms (Flores & Rosa, 2023, p. 69).

These theories are supplemented by algorithmic gatekeeping theories that see algorithms as tools that control entry into institutional spaces based on their automated decision-making process (Hao et al., 2024, p. 21). Incorporating these concepts into the research, the project seeks to explore the connection between language, ideology, power, and algorithmic assessment.

4. Results

Patterns emerge upon analyzing essays and Automated Essay Scoring (AES). The analysis was carried out using Critical Discourse Analysis and theories on language ideology and showed that there are four main themes. The first theme is the preference for standardized language conventions. The second theme is algorithmic gatekeeping. Third, there is preference for certain linguistic characteristics.

4.1 Language Ideologies Embedded in Automated Essay Scoring Systems

The first finding is that the automated English assessment system displays certain language ideologies. It is evident that automated feedback favors formal academic English, where strict grammatical correctness, wide-ranging vocabulary usage, and academic structuring tend to lead to better scores than less conventional usage of language. Some of the criteria in this judgment include grammatical correctness, vocabulary proficiency, sentence construction, and conventional sounding language. This shows that the language ideology here considers language



correctness synonymous with education. Therefore, the assessment of language quality was made on the basis of adherence to institutional norms rather than just effectiveness.

There were a few instances where essays written by multilingual learners showed clear reasoning skills but were marked down due to the occurrence of small differences from the linguistic standardization expected of students. These examples indicate that AES technology may be encoding academic writing in terms of a set of strict linguistic requirements. This observation corroborates existing literature on how automated scoring technology tends to favor language standardization in dominant varieties (Li & Liu, 2024, p. 7; Wang & Gayed, 2024, p. 223).

In addition, it was observed that linguistic diversity often tended to be seen as deficiency by such AES systems. The use of language expressions and rhetoric associated with non-standard discourse traditions were considered to be deficiencies according to automatic feedback results. This illustrates the extent to which language ideologies shape the workings of algorithmic assessments (Flores & Rosa, 2023, p. 69).

4.2 Algorithmic Gatekeeping in Educational Assessment

The second finding shows how the AES systems operate as gatekeepers through algorithms. The analysis found out that automated scores played a major role in shaping the perception of writers' competence and achievements. Given the scores, the systems succeeded in determining what texts can be considered adequate, good or bad.

It was found that essays written in an academic language received good marks from the researchers, but those written in other languages did not receive any despite their sound reasoning. Thus, it is safe to conclude that AES systems utilize certain linguistic criteria in deciding who qualifies as academically accomplished individuals. Moreover, automated feedback has revealed the fact that such systems have a biased approach to defining acceptable language and focus on such aspects as clarity, accuracy, academic voice, and vocabulary.

The results of critical discourse analysis prove that AES does not only evaluate the performance of a particular student; rather, it influences how language is used at institutions of higher education. The approach promotes some discourses while others remain marginalized. It can be viewed as a symbol of power in education

and supports the idea of the gatekeeping model according to which algorithmic decisions affect educational choices (Hao et al., 2024, p. 21).

4.3 Linguistic Features Prioritized by Automated Scoring Algorithms

Third finding revolves around language varieties that reap the most benefits from AES policies. The results showed four prominent language varieties that earned high scores, therefore, gaining more recognition within such policies.

- Grammatical accuracy;
- Lexical sophistication;
- Cohesive organization;
- Syntactic complexity.

Grammar was the aspect that received the most attention. Those essays that had few grammatical errors were likely to score high even if the difference in their overall writing quality was not that significant. All automated feedback emphasized punctuation, sentence structure, and formalism of language.

Lexical sophistication was important as well. Complex lexical choices were awarded higher than essays composed using simple but efficient vocabulary.

Finally, the organization was one of the important features that affected the final grade as well. Standard organization including an introduction, body paragraphs, and a conclusion earned more points. Logical order of ideas made essays receive higher grades.

Another interesting observation is related to the effect of essay syntax on its success. Complex sentences performed better compared to other types of sentence structures. Simple sentences failed even though they used correct language because the software had difficulties in evaluating them.

This finding confirms observations made in other studies where it was stated that AES systems pay attention to linguistic parameters that can be easily quantified computationally (see Tang et al., 2024, p. 4188; Schaller et al., 2024, p. 215).

4.4 Implications for Linguistic Diversity and Educational Equity

Then, we will consider potential impacts on language diversity and educational equity by AES systems. There appears to be a possibility that such technologies



will result in punishment for authors who do not conform strictly to the conventions of conventional writing.

Most notably, multilingual learners will suffer because their texts will receive lower grades when there is linguistic or rhetorical transfer present. There are instances where perfectly valid papers will earn poor grades simply because they do not match the strict conventions of AES systems.

Additionally, the results indicate that AESs will reinforce the advantages of students who have an easier access to education. Because AESs favor specific language traits, which are more common among students from privileged groups, there will remain some inequality in terms of academic performance.

However, the lack of transparency associated with automated assessment systems also means that discrimination is difficult to detect and solve. With calculations being quite intricate, it becomes challenging to understand the logic behind some grading. There are questions concerning justice, equity, and the right to receive an education. Thus, although such automated assessment technologies are beneficial, they also support certain language beliefs and thus are discriminatory. This needs to be understood by educational establishments so that they can consider the consequences of using AES technologies (Jiao et al., 2026, p. 10; Schaller et al., 2024, p. 218).

5. Discussion

In this research, the objective was to examine the role that language ideologies play within Automated Essay Scoring (AES) systems as well as how these systems can be seen as algorithmic gatekeepers. Based on the theories of critical discourse analysis, language ideologies, and algorithmic gatekeeping, the results indicate that AES systems are not simply neutral evaluative systems but rather contribute to the formation and enforcement of specific language ideologies. This chapter provides a discussion of the results in connection with existing research.

5.1 Language Ideologies in Automated Essay Scoring Systems

Accordingly, the findings show that AES technologies favor formalized uses of academic language and rate the quality of writing based on certain language standards. In this regard, it is possible to state that the above results provide evidence of the fact that language technologies are determined by a certain sociocultural background which is embodied in the design of such technologies.

The constant preference for formal aspects of writing shows that AES is affected by particular views on how "good" writing should be.

Within the framework of language ideology theory, the obtained findings mean that AES reproduces the same sociocultural patterns as language education does. According to Flores and Rosa (2023, p. 69), educational organizations usually tend to associate language standardization with intelligence and success. It can be observed that the findings presented above are consistent with the mentioned theoretical perspective.

The findings are also consistent with earlier studies showing that language model-based evaluation methods generally reward writing that adheres to academic discourse norms (Li & Liu, 2024, p. 7). While such biasing towards standard language use may ensure greater consistency in evaluations, it may also lead to the neglect of other languages. Language ideologies based on other rhetoric, writing styles associated with different cultures, and multilingual language use may be considered inferior.

In addition, the current results complement previous literature by showing that language ideologies can manifest themselves via computer algorithms. In this study, it is demonstrated that such ideologies may be hard-coded into algorithms, thus impacting educational assessments on a mass scale. Such insights are relevant for future studies on sociolinguistic aspects of using artificial intelligence in education.

5.2 Automated Essay Scoring as a Form of Algorithmic Gatekeeping

Another important finding related to the use of AES systems deals with their function as mechanisms of algorithmic gatekeeping. According to the results of the analysis, AES creates certain boundaries between legitimate and illegitimate types of essays through the application of predefined criteria of evaluation. This results in influencing educational practices as well as opportunities for students to get recognized in the academic field.

In line with this notion, algorithms function as gatekeepers, managing access to valuable resources and opportunities (Hao et al., 2024, p. 21). Within education, AES programs participate in decision-making processes related to grading, placement, admission, and skill evaluation. Thus, algorithmic decisions have considerable consequences in society and within the field of education itself.

The findings demonstrate how gatekeeping is performed via the normalization of specific linguistic behavior. Essays that matched accepted norms of language received good grades, while any form of deviation resulted in poor evaluations. Such an analysis is in agreement with Fairclough's assertion that organizations use discourse as means of exercising power (2015, p. 58). Hence, the AES technology also embodies organizational legitimacy.

In addition, the analysis revealed that gatekeeping via algorithms differs from traditional gatekeeping practices by being less transparent. While people could justify their decisions, algorithms represent black boxes from which no explanations can be obtained.

5.3 Linguistic Priorities and the Construction of Writing Quality

According to the research findings, AES systems have a bias towards grammatical accuracy, vocabulary complexity, cohesion, and syntactic complexity. While such parameters influence grading, the research results show that there could be some elements of human communication missed when using such systems.

Grammatical accuracy and vocabulary complexity seem essential when evaluating an essay, which is consistent with traditional approaches to teaching and learning. However, there appears to be an issue of focusing exclusively on measurable elements such as grammar and vocabulary because other important aspects may be overlooked.

Indeed, the above view correlates with the findings made by Tang et al. (2024), who claim that AES technologies depend mostly on those features that are easily measured by computer calculations. The reason behind this approach lies in the nature of computers and how they function by measuring things and counting.

However, these results indicate a contradiction between efficiency and inclusivity in assessing students' papers. An emphasis on measurable linguistic features increases the efficiency of the process but decreases inclusivity since competent writing is not limited only to a certain set of rules.

Therefore, AES technologies must include all definitions of the term "quality".

5.4 Implications for Linguistic Diversity, Fairness, and Educational Equity

The last conclusion drawn from the analysis of the data refers to the possible impacts of AES on linguistic diversity and education. As seen from the findings, there is a danger that the use of algorithms in assessments might inadvertently



increase educational inequality by promoting those students who have already learned the standard academic discourse techniques.

This is consistent with the research of Schaller et al. (2024, p. 218), which examined the issues related to the fairness in automated essay scoring. In turn, the results of the current investigation show that multilingual students and learners that used alternative ways of communicating faced an increased likelihood of being given negative ratings.

There are other concerns associated with language diversity. According to language ideology scholars, schools should perceive linguistic diversity as an asset rather than a challenge (Flores & Rosa, 2023, p. 73). However, based on the analysis of the AES programs, one can conclude that many of them continue using the standards-based approach when evaluating texts.

The second implication relates to transparency and accountability issues. In order to ensure that the new technologies used for evaluation do so in line with ethical and equitable standards, institutions will have to ensure the operation of transparent algorithms. Scholars have stressed that explainable AI solutions are preferable in such a way that users receive adequate information on the functioning of their decision-making process (Jiao et al., 2026, p. 10).

Therefore, the present research reveals that technology associated with auto-graded essays has wider implications than simple grading of assignments. They involve such aspects as language, power relations, identity, and social justice issues. Future research concerning the development of auto-grading technologies must include these important variables.

6. Limitations

The significance of the research conducted on language ideologies and how AES systems operate as gatekeepers cannot be ignored. However, there is a need to highlight certain limitations. The first limitation concerns data collection scope, methodology-related difficulties, technological challenges, and constant advancement in AI technology.

To begin with, the limitation in relation to the data collection scope should be pointed out. Despite all the efforts taken to ensure diversity in terms of language, it is unrealistic to represent all the writing styles and backgrounds present worldwide. Consequently, the obtained results cannot be considered generalizable (Creswell & Creswell, 2023, p. 171).



Another issue refers to the fast-changing nature of contemporary AI technologies. Nowadays, AES solutions, especially those using LLMs, constantly undergo changes and improvements. As a result, scoring criteria, algorithms, and even training data used for AES systems may differ depending on the time when the system has been analyzed (Hao et al., 2024, p. 26).

Another limitation involves the opaque nature of many proprietary AES tools. As developers frequently refuse to reveal specific details about the training datasets, design of the algorithms used and scoring protocols, the researcher can only work with the observable outcomes but cannot perform an analysis of internal decision-making procedures involved. Thus, the researcher can fail to determine how specific linguistic properties affect scores produced (Jiao et al., 2026, p. 10).

Moreover, this research utilizes a qualitative method of Critical Discourse Analysis, which involves the process of interpretation and contextualization of results. Although it contributes greatly to revealing ideological and power structures involved in communication, each scholar may discover some nuances in the same set of data analyzed. Such an interpretative nature of discourse studies is inherent in their qualitative character (Fairclough, 2015, p. 231).

Lastly, this research will focus more on ideological structures, algorithmic gatekeeping and other social linguistics-related questions rather than technical problems, algorithm efficiency and psychometric properties of AES. Therefore, one should take this issue into consideration when interpreting research results.

7. Future Research

Several avenues for future research can be explored on the basis of the results obtained. First, it is possible to conduct future research, expanding the data sample. Thus, it will be possible to identify whether similar linguistic patterns are present when analyzing writing in other educational settings, languages, and literary genres.

Secondly, it is possible to conduct comparative research on multiple AES platforms. As different AES uses unique training sets, models, and evaluation criteria, comparison between them can shed light on differences in language ideologies used in automated assessment practices (Schaller et al., 2024, p. 218).

Thirdly, future studies can examine the perspectives of students, educators, and administrators about AES technologies. Although the current study analyzes the output produced by algorithms, conducting interviews and surveying stakeholders can provide new insights into automated assessment practices.



Multilingual and multicultural analysis may be another direction for future study. Most of the research conducted on AES is limited to English-language educational environments. More studies could be performed using Arabic, Chinese, Spanish languages, and others to enrich our knowledge about how language ideologies are applied in different educational contexts (Li & Liu, 2024, p. 9).

Critical Discourse Analysis can be combined in future research with computational auditing and explainable AI. Such an inter-disciplinary approach would make it possible to analyze decision-making algorithms and detect possible sources of discrimination (Tang et al., 2024, p. 4188).

At last, issues of fairness, accountability, transparency, and educational equity related to the use of AI in educational assessment should be thoroughly considered. With time, automated tools and technologies gain more influence on education and should therefore attract researchers' attention.

8. Conclusion

The quick introduction of Artificial Intelligence into education testing procedures has resulted in numerous changes related to methods for evaluating students' performance. In this regard, one of the most influential technological advancements in this sphere was the emergence of Automated Essay Scoring systems, which are capable of quickly and efficiently evaluating students' written discourse. Nevertheless, the results obtained during this study prove that such technologies should be considered as ideologically-driven tools used within broader sociocultural contexts, where language competence is constructed and evaluated accordingly.

By applying critical discourse analysis techniques, language ideology theory, and the theory of algorithmic gatekeeping, the study identified the following features of AES technologies: they are inclined to prioritize grammatically accurate, lexically rich, coherent, and syntactically complex academic language. Thus, it may be concluded that AES systems use the set of specific evaluative criteria determined by language ideologies, which are deeply entrenched into sociocultural contexts (Flores & Rosa, 2023, p. 69).

More specifically, the study indicated that AESs operate as means of algorithmic gatekeeping by controlling who can access educational recognition through automated decisions. Due to their rewarding certain practices and punishing others, AESs serve as facilitators in creating and perpetuating institutional language norms.

In such circumstances, multilingual students and users of non-mainstream discourses may be disadvantaged in AES-based systems (Hao et al., 2024, p. 21).

Moreover, issues related to linguistic diversity, justice, and educational equity were revealed by the study as well. AESs can be seen as effective tools for efficient and reliable evaluation but they might become sources of social and educational disparities depending on the way how automated evaluation was implemented, being based on limited views on writing skills (Hao et al., 2024, p. 21).

In conclusion, this paper contributes to the increasingly significant academic debate around language, technology, and power relations. Through analyzing the integration of language ideologies into algorithms for assessing language skills, this paper adds to our knowledge about the place of AI in current educational processes. This paper highlights the necessity to ensure transparency and social responsibility of automatic assessments that value linguistic diversity and preserve educational quality at the same time.

References

- Bouvier, G., & Machin, D. (2024). *Critical discourse analysis and digital communication*. Bloomsbury Academic.
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
- Fairclough, N. (2015). *Language and power* (3rd ed.). Routledge.
- Flores, N., & Rosa, J. (2023). *Language, race, and educational justice*. Routledge.
- Hao, J., von Davier, A. A., Yaneva, V., Lottridge, S. M., von Davier, M., & Harris, D. J. (2024). Transforming assessment: The impacts and implications of large language models and generative AI. *Educational Measurement: Issues and Practice*, 43(2), 16–29.
- Jiao, H., Song, D., & Lee, W.-C. (2026). Evaluating rater effects of large language models in automated essay scoring: GPT, Claude, Gemini, and DeepSeek. *Educational Measurement: Issues and Practice*, 45(2), e70018.
- Kerssens, N., & van Dijck, J. (2025). Algorithmic governance and educational technologies: Emerging perspectives on AI-mediated assessment. *New Media & Society*, 27(1), 39–58.

- KhosraviNik, M. (2024). *Critical discourse studies and digital societies*. Routledge.
- Lee, S., Cai, Y., Meng, D., Wang, Z., & Wu, Y. (2024). *Unleashing large language models' proficiency in zero-shot essay scoring*. arXiv.
- Li, W., & Liu, H. (2024). Applying large language models for automated essay scoring for non-native Japanese. *Humanities and Social Sciences Communications*, 11(1), Article 723.
- Schaller, N.-J., Ding, Y., Horbach, A., Meyer, J., & Jansen, T. (2024). *Fairness in automated essay scoring: A comparative analysis of algorithms on German learner essays from secondary education*. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 210–221). Association for Computational Linguistics.
- Tang, X., Chen, H., Lin, D., & Li, K. (2024). Incorporating fine-grained linguistic features and explainable AI into multi-dimensional automated writing assessment. *Applied Sciences*, 14(10), 4182.
- Wang, Q., & Gayed, J. M. (2024). Effectiveness of large language models in automated evaluation of argumentative essays: Finetuning vs. zero-shot prompting. *Computer Assisted Language Learning*, 39(1–2), 209–237.