Characterization of Essay Content for Content-Based Assessment Using Morphological Classification Technique

Kuyoro S. *1, Eluwa J.*2, Awodele O.*3, Ajayi A.*4
Department of Computer Science, Babcock University, Ilishan-Remo, Ogun State, Nigeria.

Abstract
Assessment and ratings are crucial in the educational system. Specifically, in the process of manual scoring for essay which has shown to be inefficient and rigorous. The objective of the current study was to develop an intelligent essay scoring system using ensemble machine-learning techniques. Several grading rubrics have been identified from literature for essay grading and machine learning but they lack the characterization of the essay content for content-based assessment using morphological classification technique. This study undertakes a bibliometric analysis method to characterize and determine the appropriate morphology for content-based assessment. The identified grading rubrics were analysed with the aggregation of each review to create an individual grading rubric description. The identified grading rubrics were characterized into five (5) from a total of twenty-five (25) descriptors using morphological classification technique which was initiated based on syntax, lexical semantics, and pragmatics for content-based assessment. This finding is a step towards providing a clearly defined characterization of essay content to effectively inform the preparation for the development of an intelligent grading process using ensemble machine learning technique.

1. Introduction

Machine Learning is now commonly used in several real-world applications that have a major impact on people's lives, especially when used in Automated Essay Scoring (AES) as part of high-stakes tests [1,2]. Educational software based on Natural Language Processing (NLP) and Automated Essay Scoring (AES) Systems often use a combination of Machine Learning Techniques such as linear regression, neural networks, and ensemble technique to automatically score students' essays, and this has proven to be a strong match or even surpass the reliability of human graders [3].

The ability of a computer program to comprehend spoken human language is referred to as natural language processing (NLP). Natural language processing (NLP) is a form of artificial intelligence (AI). Natural Language Processing (NLP) is a textual data analysis technology that is commonly used. The technology is used to break down the texts into tokens, which can then be used to compute. This method is used to make data conversion easier during study. To guarantee that the expressions are grammatically and factually accurate, these tokens are reviewed and evaluated with the adjacent tokens. Stemming, lemmatization, discourse, and other NLP strategies can be used [4]. As a result, NLP has been discovered to be appropriate and relevant in the sense of an intelligent scoring system.

Writing ability assessment has advanced from students answering multiple-choice tasks to tasks based on constructed-response which requires students to read prompt (s), reflect on major key points in the development of an essay; consequently, their use of language use, writing skills, and critical thinking can be assessed. For the last 30 years, holistic writing evaluation has been the gold standard for evaluating writing abilities. Candidates are asked to write essays on one or more subjects, which are then rated by raters, resulting in subjectivity on their part and ratings that are extremely inaccurate without adequate preparation [5]. The process of manual scoring has its limitations such as weariness, interference, discrepancy of scoring over time, and so on [6]. This inefficient and rigorous way of scoring has brought about the need for fast response as well as computerized scoring based on writing quality analysis. As a result, the Automated Essay Scoring (AES) system was developed to provide writing scores via computer programs.

Assessment and rating are crucial in the educational system. Interest in the creation and the use of Computer-based Assessment Systems (CbAS) has exploded in recent years, thanks to both an increase in the number of students enrolling in universities and the opportunities provided by e-learning strategies for asynchronous and widespread education [7].

Automatic evaluation, also known as intelligent scoring method, is favoured over manual assessment to reduce monotony, unfairness, and inconsistencies while also freeing up resources for the instructor to concentrate on the key operation. As a result, for the educational system, automatic evaluation is necessary. The field of Computer-based Assessment Systems (CbAS) has grown significantly as a result of increased intake by university systems and e-learning systems as a widespread education tool. Computer Assisted Assessment (CAA) has grown in importance as a result of advances in Natural Language Processing (NLP), Information Extraction (IE), and e-learning [8].

Text processing is a big part of intelligent scoring. To construct a text-based grading system, text processing methods must be applied to the text. The three main methods used in a free text evaluation method are keyword analysis, complete natural-language processing, and information retrieval [9,10]. Methods such as ontology, semantic similarity matching, and statistical approaches can be used to evaluate text [11].

ISSN 2229-5518

Since it is difficult to solve problems such as synonymy or polysemy in student responses, keyword analysis has historically been viewed as a poor solution. Full text parsing and semantic analysis, on the other hand, are difficult to achieve and port through languages. As a result, Information Retrieval provides a more cost-effective and reliable solution, using natural language processing (NLP) software to scan texts for contents while avoiding in-depth analysis [12]. Other methods for evaluating student free text responses include integrating keyword-based approaches, pattern recognition techniques, splitting the responses into terms and their semantic connections, Latent Semantic Analysis (LSA) with syntactic and semantic information, Machine Learning methods, and LSA with syntactic and semantic details Machine learning techniques are used in the learning process of the entire Information Discovery in Databases (KDD) data mining system [13,14].

2. Methodology

The methodological approach used mixes bibliometric, content analysis, and morphological classification technique to create a characterization of essay content for content-based assessment in automated essay scoring system. This method was selected based on its effectiveness in defining concepts (quote). In the current study, the first stage focused on clarifying the search term in google scholar alerts. The second stage was to extend the search to other research databases and the last stage was the selection of articles for analysis based on specific inclusion and exclusion criteria. The research methodology chosen for this study was a systematic literature review. The main phases of the study were as follows:

2.1 Stage 1: Essay Grading Systems Research and Classification

The present stage was divided into three namely: identification, screening and inclusion. In stage 1, bibliometric data was collected through identification. Then a screening of the overall result was carried out to identify which documents can be taken into consideration, in line with the research areas deemed interesting and relevant. Finally, inclusion is aimed at selecting the articles to be discussed in detail.

2.2 Stage 2: Literature Analysis

After the completion of stage 1, the next stage is literature analysis. The approach used for the bibliometric analysis included the use of indicators for the parameters studied; and morphological analysis for feature grouping. In the current study, the authors developed an extensive and comprehensive search query to retrieve all potential documents focusing on rubrics utilized in automated essay scoring.

The search terms were determined through a process of refining an alert in google scholar on automated essay scoring. Google scholar was selected because it has incentives for quality, visibility and open access [quote 11]. Based on this process of refining, the following search alerts were created: “Grading Rubrics”, “Essay Grading”, “Automated Essay Scoring”.

2.3 Stage 3: Literature Discussion

After the second stage, a third and final one followed, where the literature results were discussed, and conclusion were drawn on the feature grouping. In Figure 1, the main stages and steps followed for the characterization are shown.

3. Result of the Bibliometric Analysis

3.1 Stage 1: Essay Grading Systems Research and Classification

The first stage consisted of the search of articles, which included the activities of collecting the materials belonging to the academic world. This first stage was divided into three parts as follows:

3.1.1 Identification (Step 1)

For a comprehensive survey of the phenomenon, an investigation on the electronic and hand searching with citation tracking were carried out within MDPI, Taylor & Francis, Elsevier, John Wiley & Sons, Springer and Copernicus. Based on the research, the process of refining was with the keywords “Grading Rubrics” AND “Essay Grading” AND “Automated Essay Scoring”, as shown in Table 1.

In order to maintain the consistency of the results, the same keywords were used in all databases and a time horizon of 20 years was chosen, from 2000 to 2020.

The choice of the keywords for performing the survey was based on the awareness that grading rubrics can be an important tool in the effort to adopt an objective essay grading in the context of automatic essay grading. In this regard, it is worthy to note extracting the right features will allow the classification algorithm to create a reliable predictive model that will influence the accuracy of the system. Therefore, a set of features that are meaningful to the task of essay scoring must be created and characterized for content-based assessment using morphological classification technique.

Table 1: Keywords and time period

Keywords	**Time Period**
Grading Rubrics	2000 - 2020
Essay Grading	
Automated Essay Scoring	

The search returned in total 1437 potentially useful documents. The result extracted Elsevier (489) are numerically superior to the other databases namely Springer (291), Taylor & Francis (212), MDPI (156), Copernicus (147), John Wiley & Sons (142), (Table 2).

Table 2: Total results of articles on electronic database

Documents Carried out on 2020	
Source of Articles	Results
Elsevier	489

Springer	291
Taylor & Francis	212
MDPI	156
Copernicus	147
John Wiley & Sons	142

The result underlines that most of the documents are articles (37.78%) and subsequently conference papers (26.02%). All the document types are filled in Table 3.

Table 3. Distribution of documents types

Documents Carried out on 2020		
Document Types	**Records**	**Contribute %**
Article	543	37.78
Conference Paper	374	26.02
Review	351	24.42
Proceedings Paper	143	9.95
Book Chapter	13	0.90
Editorial	8	0.55
Book	5	0.34

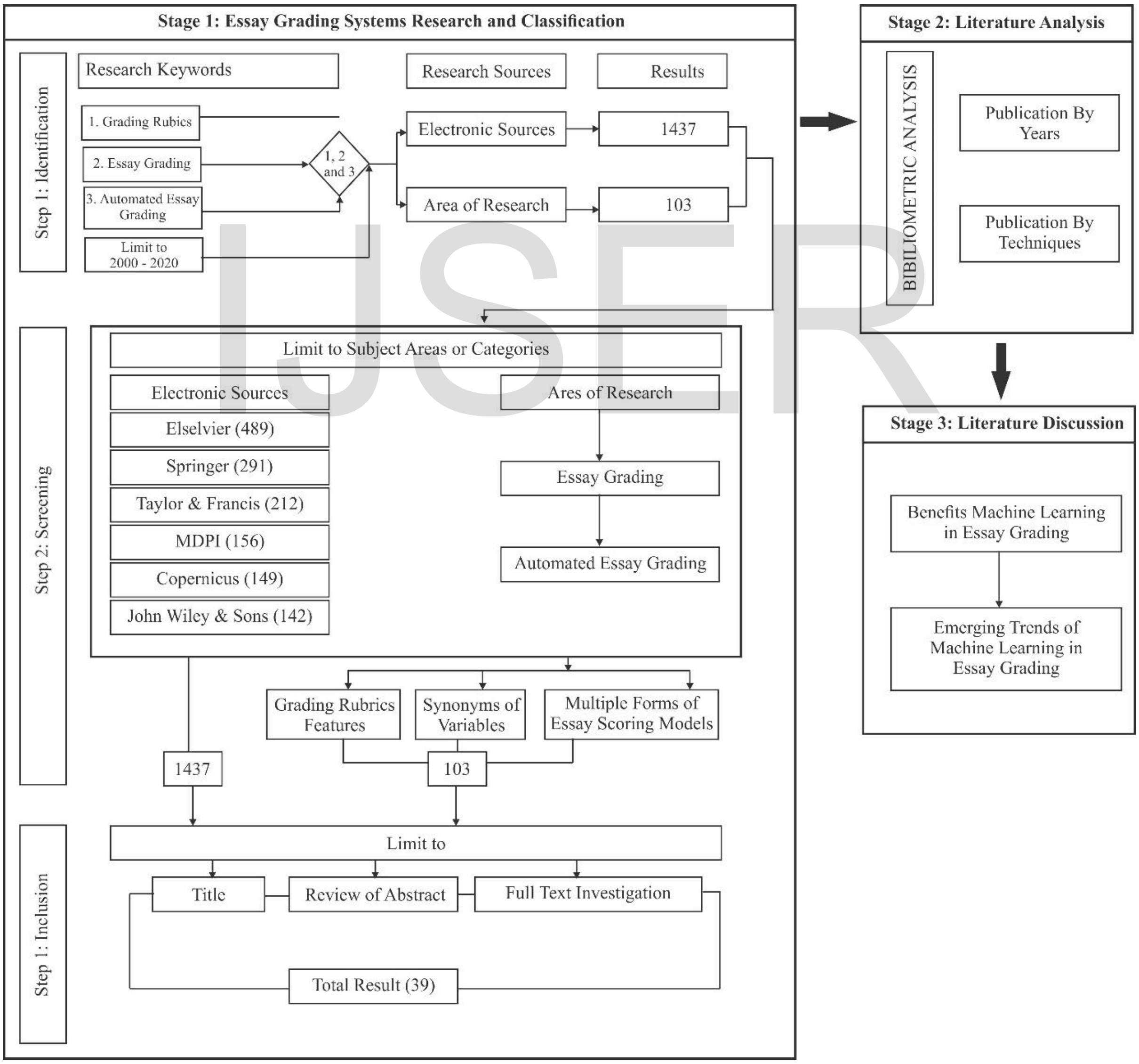

Figure 1: Process Flowchart

3.1.2 Screening (Step 2)

Trying to give an overview of the documents in the screening part. Electronic searching, hand searching, and citation tracking were carried out within Elsevier, Taylor & Francis, John Wiley & Sons, Springer, MDPI and Copernicus. Based on the area of research, the search was limited to essay grading and automated essay scoring. This approach ensured that the search was tailored to each research database which increased the efficiency and accuracy. During the search process, several terms were combined to form key terms combinations. For example, first, there is a term referring to grading rubrics such as features, the second term refers to various synonyms of variables and the third term refers to multiple forms of an essay scoring models. The combination of the three terms was used to identify the potentially relevant pieces of literature from the research databases.

3.1.3 Inclusion (Step 3)

At the end of the screening process, the inclusion step was started, which consisted in the selection of documents which bibliometric analysis were performed. The selection of documents was conducted over three stages after retrieval from the research database. This included the selection of articles based on the title, followed by the review of abstract and then a full-text investigation. An article was considered relevant if the inclusion criteria were attained, which included grading rubrics, essay grading, automated essay scoring and discussing how grading rubrics can be categorized. Articles were excluded if they did not include essay, automated scoring, features, grading rubrics and discussing how grading rubrics can be categories. The first stage (search) identified 1437 potentially useful documents. After applying the inclusion and exclusion criteria, 103 full-text articles were analysed, out of which, 64 were rejected because they did not include essay, automated scoring, features, grading rubrics and discussing how grading rubrics can be categorized. The remaining 39 were identified as being relevant and data were extracted as the final sample to be analyzed.

3.2 Stage 2: Literature Analysis

This section presents and discusses the findings of this review. Firstly, an overview of the selected studies is presented. Second, the review findings according to the characterization of essay content for content-based assessment using morphological classification technique.

3.2.1 Individual system description

A matrix was developed to provide an individual case description for each article (Table 4).

Article	Year	System	Features	Approach / Techniques	Training
The Imminence of Grading Essays by Computer. [15]	1997	Project Essay Grading (PEG)	Sentence structure (essay length, preposition count, POS, grammar, relative pronouns)	Statistical Approach (linear Regression)	Training based on previously marked essays. sample of essays from 100 to 400 essays
Automatic essay assessment. Assessment in Education: Principles, Policy & Practice. [16]	2003	Intelligent Essay Assessor	Similarity to source, coherence, plagiarism, misspelled words.	Latent Semantic Analysis	100 samples of pre-scored essay.
Free text assessment in a virtual campus. [17]	2000	Apex Assessor	most similar portion, semantic distance between sentences	Latent Semantic Analysis, NLP	set of unmarked texts for training
A study of expert scoring and intellimetric scoring accuracy for dimensional scoring of grade 11 student writing responses. [18]	2000	IntelliMetric	syntactic and grammatical structure of the essay, parts of speech, vocabulary, sentence structure, and concept expression	NLP	manually scored answers
Stumping ERater: Challenging the validity of automated essay scoring. [19]	2001	Educational Testing Service (ETS I)	Remove suffixes and stop words(manual) Metonyms (manual classification) Grammar rules	MsNLP tool, Phrasal node extraction	Domain specific, Concept specification
Towards robust computerized marking of free-text responses. [20]	2002	Automark	Punctuation and spelling, Syntactic constitutes, pattern matching	NLP	

Automated essay scoring using Bayes' theorem. [21]	2002	Bayesian Essay Test Scoring System (BETSY)	Surface features (word count, verbs count, commas count, sentence length, frequency of article) Content features (specific words, frequency of content words, occurrence of specific noun verbs)	Multivariate Bernoulli model, Bernoulli model, probability	
Automated scoring of short-answer Questions. [22]	2003	Conceptual Rater (C-Rater)	Synonyms, similar words, spelling, canonical representation of each response	Distance algorithm, pronoun resolution, filling semantic gaps	No large Collection of graded essays
Automated Essay Evaluation: The Criterion Online Writing. [23]	2004	Intelligent Essay Marking Systems (IEMS)	Pattern recognition	Pattern indexing neural networks, Index on clusterisation Algorithm	
Parameters driving effectiveness of automated essay scoring with LSA. [24]	2005	Intelligent Essay Assessor (IEA)	Bag of words	Topic modeling, Latent Semantic Analysis, Singular Value Decomposition	Student essay and class teacher essay.
An overview of automated scoring of essays. [25]	2006	Electronic Essay Rater (E-Rater)	Variety words used syntactic structure (POS), Syntax, discourse topic, Topic analysis (vocabulary usage, vector space model, weight vector)	MsNLP tool Corpus based approach, Corpora linguistic approach	Trained with set of essays scored by faculty
Automated Essay Grading using Machine learning. [26]	2012	Forward Feature Selection AES	Bag of Words, Numerical features, Parts of Speech count, Orthography, Structure and Organization.	Machine Learning	All essay set.
A ranked-based Learning Approach to Automated Essay Scoring [27]	2012	Learning to rank AES	Term usage, sentence quality, spelling errors, content fluency and richness, essay length, conjunction, grammar errors.	Machine Learning	All essay set
Automated Essay Scoring System by using Support Vector Machine. [28]	2013	Conventional Support Vector Machine AES	Term document, term complexity (essay length, number of characters, number of sentences, average sentence length, average word length, number of clauses.	Machine Learning	Feature vector extracted from essay texts and categorical score values graded by human raters.

Automated Essay Scoring by Maximizing Human-machine Agreement [29]	2013	Rank Based AES	Lexical, Syntactic, Grammar and fluency, Content and prompt specific	Listwise learning to rank	Essay with corresponding human rating.
Modelling thesis clarity in student essays [30]	2013	Regression Based AES	Lexical, category-based, syntactic and Semantic	Machine Learning	Extracted Essay set
Modelling prompt adherence in student essays. [31]	2014	Regression Based AES	Lexical, category-based, syntactic and Semantic	Machine Learning	Extracted Essay set
Flexible Domain Adaptation for Automated Essay Scoring Using Correlated Linear Regression. [32]	2015	Bayesian linear ridge regression and domain adaptation		domain adaptation technique	Extracted Essay set
Automatic text scoring using neural networks. arXiv preprint arXiv:1606.04289. [33]	2016	Score-Specific Word Embedding (SSWE) + two-layer Bidirectional Long-Short-Term Memory (LSTM)	Character and Part-of-Speech word unigrams, bigrams and trigrams; and the distribution of common nouns, prepositions, and coordinators. Additionally, we extract and use as features the rules from the phrase structure tree based on the top parse for each sentence, as well as an estimate of the error rate based on manually derived error rules	Deep neural network	Training data set and resolved score
Comparative Analysis of String Similarity and Corpus based Similarity for Automatic Essay Scoring System On E-learning Gamification. [34]	2016	Corpus Based Similarity AES	None	LSA and Cosine Similarity	All essay set.
Constrained Multi-Task Learning for Automated Essay Scoring. [35]	2016	constrained multi-task learning approach	word unigrams, bigrams, and trigrams; POS (part-of-speech) counts; essay length (as the number of unique words); GRs (grammatical relations); max-word length and min-sentence length; the presence of cohesive devices; an estimated error rate.	Machine Learning	Preference-ranking model based on a binary margin-based linear classifier
Using argument mining to assess the argumentation	2016	Regression Based AES	Length based, category based, prompt relevant,	Machine Learning	Extracted Essay set

quality of essays. [36]			syntactic, semantic and argumentation		
Automated Evaluation of School Children Essay in Arabic. [37]	2017	Hybrid AES	Spelling and Grammar mistakes, coherence, organization of the essay.	LSA and Rhetorical Structure Theory.	Entire essay set
Automated Essay Rater Using Natural Language Processing. [38]	2017	Regression Based AES	Bag of words, POS count frequency, Similarity (LSA), statistical features (word count, sentence count, average sentence length, paragraph count), orthography.	NLP and machine learning	Extracted features from essay and corresponding scores.
Automated essay rater using natural language processing. []	2017	Paperless School free-text Marking Engine (PS-ME)	correct master texts for comparison	NLP techniques	Extracted Essay features
Intelligent Auto-grading System. [39]	2018	Bi-directional Long Short-Term Memory AES system	None	NLP and Deep Learning	Whole essay dataset
ASAP++: Enriching the ASAP Automated Essay Grading Dataset with Essay Attribute Scores. [40]	2018	Random Forest Classifier AES	Length, Punctuation, Syntax, Stylistic features, Cohesion features, coherence features, language model features, n-Gram features.	Machine Learning	Essay with corresponding overall score
Ensemble Learning on Scoring Student Essay. [41]	2018	Word2vec and Ensemble method AES	Length of essay, number of misspelled words, numerical features.	Machine learning	Extracted features from the essay set.
An Overview of Schema Extraction and Matching Techniques. [42]	2018	Schema Extract Analyse and Report (SEAR)	set of common metrics, some initial calibration	NLP	subset of essays as training
A Comprehensive Study on Traditional AI And ANN Architecture. [43]	2019	Automated Text Marker (ATM).	text, and their dependencies, syntax, semantic analysis,	IE techniques, NLP	
Automated Essay Scoring with Ontology based on Text mining and NLTK tools. [44]	2018	Domain Concept Ontology Based feature extraction	Numerical features, parts of speech count, domain ontology using OntoGen, Orthography, Similarity	NLP (NLTK)	Extracted features from the essay set.
Automated Grading System using Natural Language Processing. [45]	2018	Knowledge Oriented AES	Key points and Sequence of key points, correct grammar, relevant key concepts	NLP and LSA	Extracted features from the essay set.
An empirical analysis of machine learning models for	2018	Relief and Correlation-based Feature	Word count ratio, sentence length, voice of the essay, tense of the essay,	Natural Language Processing and Machine Learning techniques	manually graded essays for training.

automated essay grading. [46]		Subset Selection (CFS)	spell check, Grammatical errors, Vocabulary, Semantic Similarity essay and semantic similarity topic essay.		
Modelling Coherence in ESOL learner Texts. [45]	2018	Ranking Based AES	Length based, category based, syntactic, semantic and discourse	Machine Learning	Extracted Essay set
Automated essay scoring with string kernels and word embeddings. [48]	2018	Regression Based AES	Word Embeddings	Machine Learning	Extracted Essay set
Automated assessment of non-native learner essays: Investigating the role of linguistic features. [49]	2018	Regression Based AES	Length Based, Lexical, Prompt relevant, syntactic and Discourse	Machine Learning	Extracted Essay set
Pairwise: Automatic Essay Evaluation using Word Mover's Distance. [50]	2018	Pair-wise Ranking AES		Word Mover's Distance using Neural word Embedding.	Entire essay set
Learning to give feedback: Modelling attributes affecting argument persuasiveness in student essays.[51]	2018	Regression (Neural) Based AES	Length based, category based, word embeddings and readability	Machine Learning	Extracted Essay set
Automated Essay Scoring System Using Multi-Model Machine Learning. [52]	2020	Feature extraction and word vector model	Word count, grammar mistakes, and part of speech count	Natural Language Processing and Machine Learning	Extracted essay features
An overview of an automated essay grading systems on content and non content based. [53]	2020	Project Essay Grading (PEG)	Sentence structure (essay length, preposition count, POS, grammar, relative pronouns)	Statistical Approach (linear Regression)	Training based on previously marked essays
AEGD: Arabic Essay Grading Dataset For Machine Learning. [54]	2021	Arabic Essay Grading System.	Number of words, Number of unique words.	Natural Language Processing and Machine Learning techniques	Supervised (Classification Algorithm)

3.2.2 Publication by Years

Consistent with what is defined in Section 3.1.1,

The first automated essay grader, Project Essay Grader, developed by Page in 1966 [55] was based on surface features disregarding the semantic side of essays. In 2002, Home measurement Inc. [56] acquired and modified PEG to analyze training essays with more than 500 features such as fluency, diction, grammar, and construction. Since 2000, many other AES such as Intelligent Essay Assessor [57], Apex Assessor [17], Intellimetric [18], Bayesian Essay Test Essay Scoring (BETSY) System [21], Intelligent Essay Marking System [37], Electronic Testing Service [19], Conceptual Rater [22] have been developed. There have been modifications and refinements on the features utilized, Lexical, syntactic and semantic, in automated essay scoring till date which have produced more reliable and objective essay scores. In general, the integrity of essay scores depends largely on the extracted features.

3.2.3 Publication by Techniques

Consistent with what is defined in Section 3.1.1,
The most common models found for AES systems are based on statistical approaches, Latent Semantic Analysis (LSA), Natural Language Processing (NLP), Machine Learning, Neural Networks and so on. Few of the reviewed papers on table (4) utilized statistical based techniques [15,53] in essay scoring without considering their semantic relationship. However, this technique was quickly replaced with the Latent Semantic analysis technique [16] and in some cases with a combination of other techniques such as Natural language Processing [24,45,17], Rhetorical Structure Theory [12], deep learning [23] and cosine similarity [34]. Natural Language Processing [44,42,25,20,19,18] was solely utilized in many instances and in some instances with combination of other techniques such as Information Extraction [45], machine learning [52,15,28,29,38,30,31] and deep learning [23]. Machine learning [29,51,41,53,49,36,47,35] was largely used as well in essay scoring. Other techniques such as phrasal node extraction [19], Bernoulli model [21], semantic gaps [59], neural networks [33], Ranking [23], domain adaptation [22] and Word embeddings [50].

3.2.4 Morphological Analysis

There were different descriptions of grading rubrics (features group), which were described and characterized into five based on essay content using morphological analysis (Table 5 and 6). Morphological analysis is very meaningful for the determination of part-of-speech structure in syntactic parsing and for the semantic analysis of a sentence. This was achievable due to the contributions and interactions of seven experts in Linguistic. These interactions gave different definition according to morphology and syntax, morphology and lexical semantics, and morphology and pragmatics. The various features employed in existing essay scoring systems have been categorized along features groups and their description in Table 5.

Table 5: Grading Rubrics Description

S/N	**Features Group**	Description
1	Word Count after Stop Word Removal	Number of words present in an essay
2	Ratio of words and Sentences	Words to sentence ratio present in an essay
3	Total Number of Characters	Number of characters present in an essay
4	Total Number of Paragraphs	Number of paragraphs present in an essay
5	Lexical	Statistics of word length, word level, unique words and spelling errors.
6	Syntactic	Statistics of sentence length, sub clauses, sentence level, mode, preposition, comma
7	Grammar and Fluency	Word bigram & trigram, POS bigram & Trigram
8	Content and Prompt Specific	Essay length, vector word similarity, semantic vector similarity, text coherence
9	Grammar and style	Number of grammar errors, spelling and style error per sentence.
10	Organization and Development	Number of discourse markers per sentence
11	Lexical Complexity	Average word length, number of tokens in an essay, number of different words in an essay.
12	Richness of essay content	Character Count, Word Count, Fourth root of word count
13	Complexity of Term usage	Average word length, Words Count >=5 and 8 char, difficult word count, long word count.
14	Orthography	Spelling error
15	Text Complexity	Unique word count, stop words count, part of speech count
16	Essay Organization	Sentence count, average sentence length, Exclamation mark count, comma count, question mark count.
17	Length	Number of Characters, words, commas, apostrophes, average word length
18	POS	Number of bad POS n-gram, Number of bad POS n-gram divided by number of bad words in the essay.
19	Prompt	Number of words in the essay that appear in the prompt, Number of words in the essay that appear in the prompt divided by the total number of words in the essay.
20	Bag of Words	Count of useful unigrams and bigrams (un-stemmed), count of stemmed and spell corrected useful unigrams and bigrams.
21	Punctuation	Count of commas, quotations, apostrophes
22	Style	Formality, word frequency, type-token ratio
23	Cohesion	Discourse connection, Entity Grid
24	Language Model	Average similarity between adjacent sentences of POS Tag Lemmas.
25	N-Gram	Word n-Grams and POS n-Grams.

From Literature, several features group with their variables have been utilized in essay scoring as depicted in Table 5. However, in order to achieve better accuracy for automated essay scoring system, irrelevant or partly relevant grading rubrics which can impact model performance should be isolated. Hence, morphological analysis was initiated based on syntax, lexical semantics, and pragmatics for content-based assessment to classify into five generic feature groups as contained in Table 6.

Table 6: characterization of essay content for content-based assessment using morphological classification technique;

S/N	Features Group	Description / Variables
1	Organization	Sentence count, total number of characters
2	Content and Coherence	Semantic similarity, Grammar
3	Lexical	Spelling errors, English and non-English words
4	Syntactic	Part of Speech Tagging,
5	N-Gram	Word N-gram

3.3 Stage 3: Literature Discussion

3.3.1 Benefits of Machine Learning in Essay Grading

With the advancement in technology, machine learning has a lot of potentials in education and Automated essay grading is a very vital machine learning application. One of the many benefits of machine learning in education is evident in automatic grading system which present the opportunity for an unbiased grading that is not influenced by teacher-student relationship. In most cases, there is bias when it comes to grading as a result of the inherent human nature which is subject to inconsistency due to fatigue, attitude, mood which often result in lack of objectivity in grading. With machine learning, students will receive grades according to their performance without any bias. It saves instructors time and presents a genuine state of a student's attainments in school. Also, the instructors will be able to know the effectiveness of teaching methods on learners [12]. If the agent can be properly trained with large set of relevant data, learning will become easier and achieve greater result.

3.3.2 Emerging trends of Machine Learning in Essay Grading

Nowadays, Machine Learning (ML) is one of the most promising application areas in the field of Information Technology where its application scope is almost unlimited. The application of machine learning in education is currently very interesting to researchers and scientists as Machine learning will assist educators to look toward the future. Through the analysis of their data in the system, emerging patterns will help identify student's areas of deficiency and provide training tailored to that area [57]. Machine learning techniques and algorithms are essential to the task of identifying features that may lead to automated essay scoring.

4. Conclusion

Writing skills are vital for success in school, career, and society. Writing development just like learning any complex skill is essentially a function of practice. Without extensive practice and regular feedback, students' writing performance will not improve. However, teachers are saddled with countless instructional responsibilities and often have to limit students' writing opportunities to make grading practicable [60]. Automated Essay Grading or Scoring (AEG or AES) involves automatically evaluating the score or grade of a written essay. AES systems are motivated by the need to develop solutions to assist teachers in grading essays in an efficient and effective manner. AES systems are also useful for students to understand issues in their writing by receiving quick feedback from a system instead of waiting for inputs from a teacher. Accurate and reliable AES systems are needed by schools, universities and testing companies to be able to manage the grading of essays by large number of students.

REFERENCES

[1] Guimaraes, A., & Tofighi, G. (2018). Detecting zones and threat on 3d body in security airports using deep learning machine. *ArXiv: 1802.00565.*

[2] Guegan, D., & Hassani, B. (2018). Regulatory Learning: how to supervise machine learning models: an application to credit scoring. *The journal of Finance and data science.*

[3] Rudner, Lawrence, M., & Liang, T. (2002). Automated essay scoring using Bayes' theorem. *The Journal of Technology, Learning and Assessment* 1(2).

[4] Rouse, M. (2017). Natural Language (NLP) *Search Business Analytics*. [Online]. Available: https://searchbusinessanalytics.techtarget.com/definition/natural-language processing.

[5] Fahim, M., & Bijani, H. (2011). The Effects of Rater Training on Raters ' Severity and Bias in Second Language Writing Assessment. *Iranian Journal of Language Testing*, *1*(1), 1–16.

[6] Leyi, Q., Yali, Z., & Yan, C. (2019). Evaluating China's Automated Essay Scoring System iWrite. *Journal of Educational Computing Research*, 1–20. https://doi.org/10.1177/0735633119881472.

[7] Siddhartha, G. & Sameen, S. F. (2010). Design of an Automated Essay Grading (AEG) System in Indian Context International *Journal of Computer Applications (0975 – 8887), 1(11)*, 60 – 65.

[8] Drolia, S., Agarwal, P., Rupani, S., & Singh, A. (2017). Automated Essay rater using Natural Language Processing. International Journal of Computer Applications. 163(10).

[9] Willet, P. (1988). Recent Trends in Hierarchical Document Clustering." Information Processing and Management, 24(5): 577-597.

[10] Guimaraes, A., & Tofighi, G. (2018). Detecting zones and threat on 3d body in security airports using deep learning machine. *ArXiv: 1802.00565.*

[11] Selvi, P. & Bnerjee, A. K. (2010). Automatic Short – Answer Grading System (ASAGS) National Institute of Technology, Tiruchirappalli, Tamilnadu, India, *InterJRI Computer Science and Networking, 2(1)*, 19 – 23.

[12] Allen Cranston (2019) How will machine Learning shape the future of education.

[13] Smeaton, A. (1992). Progress in the Application of Natural Language Processing to Information Retrieval Tasks. *The Computer Journal*, 35(3): 268-278.

[14] Hatzivassiloglou, V. (2001). SimFinder: A Flexible Clustering Tool for Summarization. NAACL Workshop on Automatic Summarization, Association for Computational Linguistics.

[15] Mubashir, M., & Zhao, X. (*2018). 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC 2018)*

[16]] Landauer, T. (2003). Automatic essay assessment. Assessment in Education: Principles, Policy & Practice 10(3):295-308. https://doi.org/10.1080/0969594032000148154.

[17] Dessus, P., Lemaire, B. & Vernier, A. (2000), Free text assessment in a virtual campus. In *Proceedings of the 3rd International Conference on Human System Learning*, 61–75.

[18] Vantage Learning Tech. (2000). A study of expert scoring and Intellimetric scoring accuracy for dimensional scoring of grade 11 student writing responses, Technical Report RB-397.

[19] Powers & Donald, E. (2001). Stumping ERater: Challenging the validity of automated essay scoring. *ETS Research Report Series* 1(1). https://doi.org/10.1016/S0747-5632(01)00052-8.

[20] Mitchell, T., Russel, T., Broomhead, P., & Aldridge, N. (2002). Towards robust computerized marking of free-text responses. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.

[21] Rudner, Lawrence, M., & Liang, T. (2002). Automated essay scoring using Bayes' theorem. *The Journal of Technology, Learning and Assessment* 1(2).

[22] Guimaraes, A., & Tofighi, G. (2018). Detecting zones and threat on 3d body in security airports using deep learning machine. *ArXiv: 1802.00565.*

[23] Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated Essay Evaluation: The Criterion Online Writing.

[24] Wild, 2005. Parameters driving effectiveness of automated essay scoring with LSA. Proceedings of the 9th CAA Conference, Lough- borough: Loughborough University.

[25] Dikli & Semire (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment* 5(1).

[26] Mahana, M., Johns, M., and Apte, A. (2012). Automated Essay Grading Using Machine Learning.

[27] Chen, H., He, B., Luo, T. and Li, B. (2012). A Ranked-Based Learning Approach to Automated Essay Scoring. https:// doi.org/10.1109/CGC.2012.41.

[28] Martinez, R., Hong, H., & Lee, D. (2013). Automated Essay Scoring System by Using Support Vector Machine. International Journal of Advancements in Computing Technology. 5(11).

[29] Cozma, M., Butnaru, A., & Ionescu, R. (2018). Automated essay scoring with string kernels and word embeddings. In Proceedings of ACL. 503–509.

[30] Persing, I., & Ng, V. (2013). Modelling thesis clarity in student essays. In Proc. of ACL, 269.

[31] Persing, I., & Ng, V. (2014). Modelling prompt adherence in student essays. In Proc. of ACL, 1534–1543.

[32] Phandi P., Chai K. & Ng H. (2015). Flexible Domain Adaptation for Automated Essay Scoring System Using Correlated Linear Regression. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 431-439.

[33] Alikaniotis, D., Yannakoudakis, H., & Rei, M. (2016). Automatic text scoring using neural networks. arXiv preprint arXiv:1606.04289.

[34] Pramukantoro, E. & Fauzi (2016). Comparative Analysis of String Similarity and Corpus-Based Similarity for Automatic Essay Scoring System On E-Learning Gamification. ICACSIS.

[35] Cummins, R., Zhang, M., and Briscoe, T. (2016). Constrained multi-task learning for automated essay scoring. Association for Computational Linguistics.

[36] Wachsmuth, H., Khatib, K., & Stein, B. (2016). Using argument mining to assess the argumentation quality of essays. In Proc. Of COLING. 1680–1691.

[37] Al-Jouie M., & Azmi A. (2017). Automated Evaluation of School Children Essays in Arabic. 3rd International Conference on Arabic Computational Linguistics. 1-22.

[38] Drolia, S., Agarwal, P., Rupani, S., & Singh, A. (2017). Automated Essay rater using Natural Language Processing. International Journal of Computer Applications. 163(10).

[39] Wang Z., Liu J., & Dong R. (2018). Intelligent Auto-grading System. Proceedings of CCIS.

[40] Mathias, S. and Bhattacharyya, P. (2018). ASAP++: Enriching the ASAP Automated Essay Grading Dataset with Essay Attribute Scores.

[41] Liu, H., Ye, Y., & Wu, M. (2018). Ensemble Learning on Scoring Student Essay. International Conference on Management and Education, Humanities and Social science. (189).

[42] Hassan, M., Mubashir, M., and Dar, K., (2018). An overview of Schema Extraction and Matching Techniques. *2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference***.** https:// doi.org/ 10.1109/IMCEC.2018.84695026.

[43] Sallauddin, M., Ramesh, D., Harshavardham, A., Pasha, S. and Shabana, (2019). A Comprehensive Study on Traditional AI and ANN Architecture. International Journal of Advanced Science and Technology. 28(19). 479-487.

[44] Contreras, J., Hilles, S., Abubakar, Z. (2018), Automated Essay Scoring with Ontology based on Text Mining and NLTK tools.

[45] Rokade, A., Patil, B., Rajani, S., Revandkar, S. and Shedge, R. (2018). Automated Grading System using Natural Language Processing. In the proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies.

[46] Madala DSV, Gangal A, Krishna S, Goyal A, Sureka A. 2018. An empirical analysis of machine learning models for automated essay grading. PeerJ Preprints 6: e3518v1 https://doi.org/10.7287/peerj.preprints.3518v1

[47] Yannakoudakis, H. and Briscoe, T. (2012). Modelling Coherence in ESOL learner texts. Proceedings of the BEA workshop. 33–43.

[48] Cozma, M., Butnaru, A., & Ionescu, R. (2018). Automated essay scoring with string kernels and word embeddings. In Proceedings of ACL. 503–509.

[49] Vajjala, S. (2018). Automated assessment of non-native learner essays: Investigating the role of linguistic features. International Journal of Artificial Intelligence in Education, 28(1):79–105, 2018.

[50] Tashu T. & Horvath T. (2018). Pairwise: Automatic Essay Evaluation using Word Mover's Distance. In Proceedings of the 10th International Conference on Computer Supported Education. (1) 59-66.

[51] Ke, Z., Carlile, W., Gurrapadi, N., & Ng, V. (2018). Learning to give feedback: Modelling attributes affecting argument persuasiveness in student essays. In Proc. of IJCAI, pages 4130–4136, 2018.

[52] Zhu, W. and Sun, Y. (2020). Automated Essay Scoring System Using Multi-Model Machine Learning. *Computer Science & Information Technology*.

[53] Dadi, R., Pasha, S., Sallauddin, M., & Sidharha, C. (2020). An overview of an automated essay grading systems on content and non-content based. *IOP Conf. Ser.: Mater. Sci. Eng.* 981 022016. https://doi.org/10.1088/1757-899X/981/2/022016.

[54] Al-Shagabi, B., Alzyadat, R., and Hama, F. (2021). AEGS: Arabic Essay Grading Dataset for Machine Learning. Journal of Theoretical and Applied Information Technology 99(6).

[55] Page, E. B, (1966). The Imminence of Grading Essays by Computer. Phi Delta Kappan,48:238.

[56] Home Measurement Incorporated. 2019. Available at http:// www.measurementinc.com/ (accessed on 25 June 2021).

[57] Lynch, M. (2018). The Benefits and Limitations of Machine Learning in Education**.**

[58] Ramalingam, V. V., Pandian, A., Prateek, C., & Himanshu, N. (2018). Automated Essay Scoring using Machine Learning Algorithm. *National Conference on Mathematical Techniques and its Applications (NCMTA 18)*, 1-7. IOP Publishing. https://doi.org/10.1088/17426596/1000/1/012030.

[59] Callan, J. P. (1994). Passage-Level Evidence in Document Retrieval. *In Proceedings of the 17th ACM SIGI International Conference on Research and Development in Information Retrieval,*302-309.

[60] Kellogg, R., & Whiteford, A. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, *44*(4), 250–266.