



Contactless Academia – The Case for Automated Essay Scoring (AES) System in COVID 19 Pandemic

Kennedy A. Osakwe^{1*}, Kunle Ola² and Pete Omotosho¹

¹RMIT University, School of Property, Construction and Project Management, 124 La Trobe St,
Melbourne VIC 3000, Australia.

²Thomas More Law School, Australian Catholic University, Level 7, 486 Albert Street, East Melbourne
Vic 3002, Australia.

Authors' contributions

This work was carried out in collaboration among all authors. Author KAO designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors KO and PO managed the analyses of the study. Authors KAO and KO managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/CJAST/2021/v40i431292

Editor(s):

- (1) Dr. Kleopatra Nikolopoulou, University of Athens, Greece.
- (2) Dr. Yahya Elshimali, Charles Drew University of Medicine And Science, USA.
- (3) Dr. Ashish Anand, GV Montgomery Veteran Affairs Medical Center, USA.

Reviewers:

- (1) Shaur Sarfaraz, Jinnah Sindh Medical University, Pakistan.
 - (2) Israel Edem Agbehadji, Central University of Technology, South Africa.
 - (3) Meseşan Schmitz Luiza, Transilvania University of Brasov, Romania.
- Complete Peer review History: <http://www.sdiarticle4.com/review-history/66677>

Systematic Review Article

Received 16 February 2021

Accepted 22 March 2021

Published 29 March 2021

ABSTRACT

Background: The 2019 SAR COV- 2 outbreak ushered and made the term 'contactless' a new normal for most businesses as a mitigation measure to risky coronavirus exposure. Similarly, there are several exposure scenarios in higher education where contact poses a threat. One of which is the handling and marking of essay scripts from assignments, task, research outputs and more. An invaluable measure worth considering is the inclusion of 'Automated Essay Scoring' (AES) system in the mitigation toolkits for higher institutions of learning.

Objectives: We conducted this scoping review to identify the suitability of AES products in higher education and examine the type of methods used to present these products.

Methods: This study was undertaken in the form of a scoping review using the Prisma flow sequence of literature search and selection from 6 databases.

*Corresponding author: E-mail: kennedy.osakwe@rmit.edu.au, kennedy.osakwe@gmail.com;

Findings: Different AES products, literatures and research designs were employed in the investigation of AES products. The outcome of reviewed literatures varied on suitability of AES in scoring essay task in Higher Institution of Learning.

Conclusion: There exist substantial case for the use of AES in most literatures amongst few opposing authors; however, in order to achieve contactless interface with human and materials in COVID 19 pandemic, AES should be used with triggers for human raters' intervention in exceptional cases.

Keywords: COVID 19; AES; contactless academia.

1. INTRODUCTION

As the realism of 'remote working' continues to make significant foothold globally due to COVID 19 outbreak, the academia will have to review its processes including student essay marking and scoring. COVID 19 is a novel contagious disease caused by a viral microorganism called SARS COV 2 [1], which has a ferocious transmission pattern. Mitigation of this contagion largely requires social distancing, wearing of mask, personal hygiene and remote working thus humanity, global industries including academia are made to operate in the virtual world [1]. One of the academic activities with potential threat is the physical handling and marking of essay scripts or materials. The global trend on essay scoring has evolved over the years from manual to automated, albeit some institution is yet to adopt the use of contactless approach such as the 'Automated Essay Scoring' (AES) system. Ellis Batten Page has been credited with the pioneering works on AES. AES is often used interchangeably with other terms like 'Automated Essay Grading' (AEG), 'Automated Writing Evaluation' (AWE) and 'Automated Essay Evaluation' (AEE) [2,3]. These terms all convey the same message of a program or system that automatically evaluate essays and provide feedback instantly, thus reads, understands, processes and provides results [3,2].

The academic technology market is replete with numerous AES products with some products presented in published literatures. Some published products includes Project Essay Grader (PEG), E-rater, Intelligent Essay Assessor (IEA), WriteToLearn, IntelliMetric, Bookette, AutoScore, Lexile, Light SIDE, Semantic Automated Grader for Essays (SAGE), Schema Extract Analyse and Report (SEAR), Paperless School Marking Engine (PS-ME), Markit, Lexile Writing Analyzer, Multi-classifier Fusion AEE, Generalized Latent Semantic Analysis (GLSA) based AES, Ranked-based

AES, Semantic tree based AEE, CRASE, Autoscore, Bayesian Essay Test Scoring System (BETSY) Lexile, OzEgrader, Markit, SAGrader and several others [2,4]. AES is an integral part of the global educational system, widely used for scoring examinations such as Pearson Test of English (PTE), Graduate Record Examination (GRE), Test of English as a Foreign Language (TOEFL), Graduate Management Admissions Test (GMAT), Cambridge Advanced English (CAE) and a host of others [5,6,4]. As adoption of AES by higher institutions becomes inevitable especially in a COVID 19 ravaged world, it is useful to gain a deeper insight of published literatures on AES. The objective of this scoping review is to identify the suitability of AES products in higher education and examine the type of methods used to present these products.

2. MATERIALS AND METHODS

This study was undertaken in the form of a scoping review using the Prisma flow sequence of literature search and selection. The processes include establishing the objectives, identifying relevant literatures, selection of literatures, charting the variables and data in selected literatures, collating, summarizing and presenting an overview of selected literatures [7]. Comprehensive literature search was undertaken in six databases including ERIC, EBSCO, Google Scholar, PROQUEST, Scopus and Web of Science. The searches were accomplished using keywords derived from an overview of title, abstracts, and related literatures. This search approach was used to identify, select, retrieve and present chosen literatures from a large volume of literature. It further presents different variables of selected literature including author, research method, findings and research process [7]. The Prisma flow instrument is a systematic and logical tool used to select and screen literature. Stages involved include initial search, removal of duplicates from initial hits, title and

abstracts screening, retrieval of full text of relevant literatures, screening against inclusion criteria, and inclusion of relevant literatures (Prisma).

2.1 Search Strategy

Study included literatures published in English language in peer reviewed journals with relevant titles on AES, abstract and full text, detailing study designs, essay scoring context, states research method and approach with key outcomes. Narratives on AES that were not published in peer review journals, product advertisement and publications with poor methodological quality were excluded. This review was conducted on published literatures searched between November 2019 and April 2020. Literatures were found by exploring Scopus, Google scholar, ERIC, EBSCO, PROQUEST and the Web of Science [2-10]. This was achieved using combined keywords such as Automated Essay Scoring, Automated Essay Grading, Automated Essay System, Automated Writing Evaluation.

2.2 Study Selection

PRISMA method approach was used in the process of selecting the articles. Initially, after retrieving all literatures and removing duplicates by the use of Endnote software; the retrieved articles were independently screened by two authors in terms of their titles and abstracts (KO, PO). The full text of articles was independently reviewed by the two authors (KO, PO); and in cases of disagreement, a third author (OR) resolves. In cases where the access to the full text of articles was not possible, the full

text of literatures was requested by sending an email to the corresponding authors. Also, the reference section of the selected articles was manually reviewed, with related studies retrieved (two were added).

2.3 Critical Appraisal of Individual Sources of Evidence

Appraisal of methodological quality of included studies was independently undertaken by two researchers (OK, PO), with differences resolved by a third reviewer (KO).

2.4 Data Extraction, Charting, and Synthesis

A data extraction form was designed and used to extract information from the selected full texts. Extracted data items included author(s) and year of publication, Intervention type, and comparator (if any), study populations, aims of the study, methods or study design, outcome measures (Technical Capabilities) and important results. Data were independently extracted by three authors and any disagreement was solved through discussions, and the extracted data were captured in the form.

3. FINDINGS

Initial outputs from all databases yielded 1,315,268 articles (see Table 1) which after removal of duplicated titles and abstracts, were reduced to 800,200 articles. Twelve articles (see Table 2) were selected for final review of methodological quality after detailed review of full texts.

Table 1. Database download analytics

Solution	ERIC	EBSCO (ERC)	Google scholar	PROQUEST	Scopus	Web of science
Search date	19/12/2019	19/12/2019				
Automated Essay Scoring	130	77	29,200	1159	313	204
Automated Essay Grading	21	14	25,300	935	103	64
Automated Essay Systems	104	26	87,700	2942	389	212
Automated Writing Evaluation	170	65	1,160,000	5513	396	231
Sum	425	182	1,302,200	10549	1201	711
Total	1,315,268					

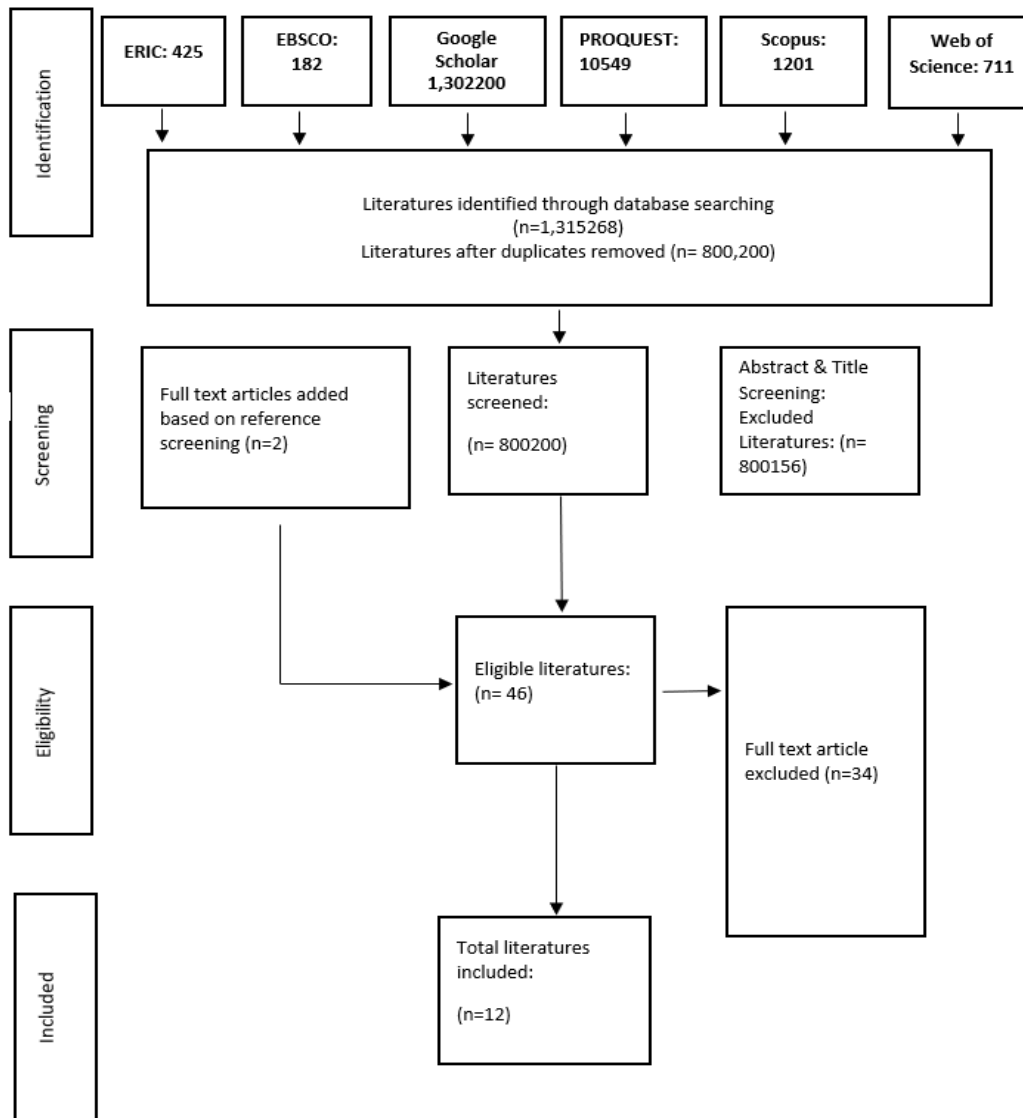


Fig. 1. PRISMA flow diagram of identified literature

Exclusion of 800,188 articles was due to poor methodological quality, product advertisement and lack of peer review. Further details are presented on PRISMA flowchart (see Fig. 1) and database download analytics (see Table 1).

4. DISCUSSION

4.1 Summary of Findings

AES products were analysed and presented using Qualitative, Descriptive, Comparative Cross-Sectional Analysis, Literature review and mixed method research approach [2,8]. While

Zupanc & Bosnic adopted several approaches; (Wang, Stallone & Lewis) & (Hussein & Hassan) used single methods to analyse and present AES products [11]. They presented AES products includes WriteToLearn, IntelliMetric, WritePlacer *Plus* test, Automark, Automatic Essay Assessor (AEA), Paperless School Marking Engine (PS-ME), Project Essay Grade (PEG), Bayesian Essay Test Scoring system (BETSY), Intelligent Essay Assessor (IEA), Graphical Interface for Knowledge Structure (GIKS), SA Grader, Bayesian Essay Test Scoring system and E-rater [11]. While these AES systems displayed strengths in reducing labour-intensive marking activities, ensuring a

consistent application of marking criteria and facilitating equity in scoring [11]; students' experiences with AES in some reviewed literatures were mixed thus not supported by some school of thoughts.

4.2 Interpretation of Findings

While different research design could be used to explore the AES concept, a comparative research design in which the functionality, capability, efficiency and effectiveness of AES product is compared with human essay raters is most desired as this would verify the replaceability of human element as academia ventures into a virtual era. Rudner and Gagner compared three AES products (PEG, IEA, and e-rater) and found that PEG has the advantage of being conceptually simpler and less taxing on computer resources [2]. They further discovered that IEA and e-rater are superior choices for grading content, as PEG relies on writing quality to determine grades. Zimmerman examined Graphic Interface for Knowledge Structure (GIKS) and found that it was able to capture and visually represent the structure of semantic KS inherent in students' writings as a network graph that features key concepts and relations [9]. Giles investigated the impact of an automated grading system (SAGrader) on student's performance. He observed that it automatically grades students' essay, can parse answers, which could be several paragraphs long, can extract meaning from text, can analyse essays using a rubric and can provide feedback in less than a minute [11]. Rudner and Liang proposed that computers and artificial intelligence can be used as tools to facilitate the evaluation of students using Bayesian Essay Test Scoring system [5]. They discovered that BETSY uses multinomial or Bernoulli Naïve Bayes models to classify texts into different classes (e.g. pass/fail, scores A-F) based on content and style attributes such as word unigrams and bigrams, sentence length, number of verbs, noun-verb pairs etc and is capable of auto analysis. Additionally, they observed that it can provide feedback but is often limited to terms such as extensive, essential, partial and unsatisfactory [5]. WriteToLearn was more consistent but highly stringent when compared to the four trained human raters in scoring essays however failed to score seven essays. Results of the test indicated that the mean score assigned by the AES tool Intellimetric was significantly higher than the faculty human raters' mean score on Writer Placer *Plus* test. While the AES systems have

many strengths in reducing labour-intensive marking activities, ensuring a consistent application of marking criteria and facilitating equity in scoring; students' experiences with some products were mixed. Perez's outcome is based on empirical data and indicated that the best technique between LSA, PLSA and LDA is LSA with an outstanding 75% correlation between the human graders and the automatic grades for the same texts. Tom Mitchell, Terry Russell, Peter Broomhead and Nicola Aldridge advised that Automark employs the techniques of Information Extraction to provide computerised marking of short free-text responses. The system incorporates several processing modules explicitly intended at providing robust marking irrespective of errors in spelling, typing, syntax, and semantics. Rudner, Garcia and Welch observed that results for the IntelliMetric system are compared to individual human raters, through Bayesian system using simple word counts, and a weighted probability model. Also, an evaluation based on system reliability depend exclusively on comparisons to scores calculated by human graders, therefore mutual agreement is considered the main determinant of performance of the IntelliMetric system during the test. Paperless School Marking Engine (PS-ME) uses numeric plus some formative feedback to the student in different areas within the subject. The PS-ME is still a work in progress. So, no result indicators are available yet.

4.3 Strength and Weaknesses

The strength of this studies lies in the systematic and rigorous literature selection process using a universal research tool called the Prisma flow literature selection method, which affords this study external validity. As this review involved three academics with authors carrying out different roles such as method critique and review of opinion with a third reviewer resolving differences in opinion, afforded the study credible internal validity. The limitation however lies in the paucity of literatures fulfilling all inclusion criteria as evidenced by the reduction of initial literature from 1,302,200 to 12 literature.

4.4 Further Work

Further work would be required to identify essential baseline capabilities needed in an AES and safeguards to the use of AES.

Table 2. Analysis of methodological quality

Author(s)	Intervention type, and comparator	Study populations	Aims of study	Methodology	Outcome measures	Results
1 Zupanc, K., & Bosnic, Z. [3]	Used 'WriteToLearn' AES on Chinese undergraduate English majors' essays to assess scoring ability and the accuracy of its error feedback. The performance of 'WriteToLearn' was compared with human raters	The study participants were 163 undergraduate English Language learners (male = 9; 5.5%; female = 154; 94.5%) enrolled in English Education in a major university in the Sichuan province of the People's Republic of China.	To systematically evaluate WriteToLearn's performance in scoring essays written and its accuracy in providing error feedback when compared to trained human raters.	Descriptive and comparative cross-sectional analysis.	Technical capabilities of product include: Part of speech tagging (POS), Syntactic parsing, Sentence fragmentation, Discourse segmentation, named entity recognition, content vector analysis (CVA), Automated summarization, Sentiment analysis, Grammatical error detection, Language analyser carry out various types of language analyse, including analyses of syntax (sentence structure), morphology (word structure) and semantics (meaning).	i. More consistent but highly stringent when compared to the four trained human raters in scoring essays; and ii. Failed to score seven essays. iii. Had an overall precision and recall of 49% and 18.7% respectively. iv. Had difficulty in identifying errors made by Chinese undergraduate English majors in the use of articles, prepositions, word choice and expression.
2 Wang, J., & Brown, M. [8]	Use of AES products by students. Compares the automated essay scoring (AES) group (Intellimetric) with the human rater's group	A sample of 107 developmental writing students with 52% male and 48% female.	The purpose of the study was to investigate the validity and usefulness of automated essay scoring for large-scale placement tests by comparing the performance of AES with that of human raters in assigning group.	Comparative study design.	Results of the test indicated that the mean score assigned by the AES tool Intellimetric was significantly higher than the faculty human raters' mean score on WriterPlacer <i>Plus</i> test.	This finding did not corroborate previous studies that reported non-significant mean score differences between AES and human scoring. This implies an unacceptable generalisability.
3 Lewis [9]	AES products used by students	A business law class made of twenty-seven students. Twenty-one students participated in	Students were surveyed on their use, satisfaction, perceptions and technical issues utilizing the Write	A mixed methodology approaches. Likert scale questions were quantitatively analyzed	The students' experiences using the Write Experience software were mixed.	The AES system not advised for small cohort of students.

Author(s)	Intervention type, and comparator	Study populations	Aims of study	Methodology	Outcome measures	Results	
		the online survey.	experience automated essay scoring (AES) software	while open ended questions with free text data were qualitatively analysed.			
4	Hussein, M.A., & Hassan, H.A [12]	Review of different AES products	Database search and review of relevant literatures	The purpose of this paper is to review literatures on AES systems used for grading the essay questions.	Literature review using Google Scholar, EBSCO and ERIC databases.	AES systems has many strengths in reducing labour-intensive marking activities, ensuring a consistent application of marking criteria and facilitating equity in scoring. However, they lack the sense of the rater as a person, they can be tricked into assigning a lower or higher score to an essay than it deserved or not.	AES models have been found to utilize a broad range of manually-tuned shallow and deep linguistic features. 2016).
5	Zupanc, K., & Bosnic, Z. [2]	Used 'WriteToLearn' AES on Chinese undergraduate English majors' essays to assess for scoring ability and the accuracy of its error feedback. The performance of 'WriteToLearn' was compared with human raters (comparator)	The study participants were 163 undergraduate English Language learners (male = 9; 5.5%; female = 154; 94.5%) enrolled in English Education in a major university in the Sichuan province of the People's Republic of China. Their age ranges between 19 and 23 years and overall, they had been learning English for approximately 9.5 years at the time of the study.	The aim of the study was to systematically evaluate WriteToLearn's performance in scoring essays written by Chinese undergraduate English majors and its accuracy in providing error feedback when compared to trained human raters.	Qualitative, Descriptive, Comparative Cross-Sectional Analysis, Literature review. Many-facet Rasch measurement (MFRM) was conducted to calibrate WriteToLearn's rating performance in scoring the whole set of essays against those of four trained human raters. In addition, the accuracy of WriteToLearn's	Part of speech tagging (POS), Syntactic parsing, Sentence fragmentation, Discourse segmentation, named entity recognition, Content vector analysis (CVA), Automated summarization, Sentiment analysis, Grammatical error detection, Language analyser carry out various types of language analyse, including analyses of syntax (sentence structure), morphology (word	The two main findings related to scoring were that: i. WriteToLearn was more consistent but highly stringent when compared to the four trained human raters in scoring essays; and ii. WriteToLearn failed to score seven essays. In terms of error feedback, WriteToLearn had an overall precision and recall of 49% and 18.7% respectively. These figures did not meet the

Author(s)	Intervention type, and comparator	Study populations	Aims of study	Methodology	Outcome measures	Results	
				feedback on 60 randomly selected essays was compared with the feedback provided by human raters	structure) and semantics (meaning).	minimum threshold of 90% precision (set by Burstein, Chodorow, and Leacock, 2003) for it to be considered a reliable error detecting tool. Furthermore, it had difficulty in identifying errors made by Chinese undergraduate English majors in the use of articles, prepositions, word choice and expression.	
6	Rudner & Gagner, [2]; Page [4]	Authors undertook an overview of three approaches to scoring written essays by computers and used Project Essay Grade (PEG), Intelligent Essay Assessor (IEA) and E-rater as the basis of the overview	The overview focused on the assessment and evaluation of student essays. Human raters graded many Essays (100-400) and determine values for up to 30 proxies. In one study, Page (1994) analysed samples of 495 and 599 senior essays from the 1998 and 1990 National Assessment of Educational Progress. With 20 variables, PEG reached multiple Rs as high as .87, close to the apparent reliability of the targeted judge groups.	To describe the three most prominent approaches to essay scoring	Comparative analysis, Literature review, qualitative and descriptive	Scores essays Measures trins and proxies. Can identify intrinsic higher-level variable, such as punctuation, fluency, diction, grammar etc., It uses regression analysis to score new essays based on a training set of 100 to 400 essays.	<p>i. States and other programs to consider automated scoring services.</p> <p>ii. Institutions should embrace AES as a validation tool with each essay scored by one human and by the computer.</p> <p>iii. Use retired essay prompts as instructional tools. The retired essays and grades can be used to calibrate a scoring system.</p>
7	Zimmerman et. Al., [9]; Kim [10]; Tawfik et. al [9]	Graphical Interface for Knowledge Structure (GIKS)	Short essay prompts were developed for use with the (GIKS) in 2	To construct a visual representation of an individual's knowledge	Comparative analysis, Analytical, literature review, empirical	Able to capture and visually represent the structure of semantic KS	Students' scores should increase from the pre- to post-test if the essay

Author(s)	Intervention type, and comparator	Study populations	Aims of study	Methodology	Outcome measures	Results
		<p>studies. For study I, participants were students in two online sections of an undergraduate-level introductory statistics course. Of the 72 enrolled students, 53 completed the pre-test and gave permission for their data to be used for research purposes. 38 students also completed the post-test. 6 students gave permission for their data to be used for research purposes but only completed the post-test. For study II, there were 109 enrolled students. 95 completed the pre-test and gave permission for their data to be used for research purposes. 75 completed the post-test. One student who completed the pre and post-test did not complete the final exam. Therefore, data was on only 74 students. All who took the post test took the pre-test due to the introduction of a new course</p>	<p>structure.</p>		<p>inherent in students' writings as a network graph that features key concepts and relations</p>	<p>prompts are measuring content that was taught in the course. As evidence of this curricular validity, scores did increase from pre- to post-test on all measures for both prompts.</p>

Author(s)	Intervention type, and comparator	Study populations	Aims of study	Methodology	Outcome measures	Results	
8	Zupanc & Bosnic, [2]	SA Grader	management system. The study participants were 1000 students in a First-year biology class. The students answered questions using SAGrader.	To determine the impact of an automated grading system (SAGrader) on student's performance.	Comparative, empirical and analytical	It automatically grades students' essay. It can parse answers, which could be several paragraphs long. It can extract meaning from text, can analyse essays using a rubric and can provide feedback in less than a minute.	1. Students like SAGrader because it provides feedback on their work in less than a minute. 2. Students can resubmit their work, having taken the feedback into account, and may gain a higher score previous years.
9	Rudner & Gagner, [2]; Zupanc & Bosnic [3]	Bayesian Essay Test Scoring system, A non-proprietary windows-based program using the Bayesian Linear Ridge Regression. Two Bayesian models for text classification from the information science field was used.	Both models were calibrated using 462 essays with two score points. The calibrated systems were applied to 80 new, pre-scored essays with 40 essays in each score group. Manipulated variables included the two models; the use of words, phrases and arguments; two approaches to trimming; stemming; and the use of stop words.	The proposal of computers and artificial intelligence as tools to facilitate the evaluation of students	Literature review, analytical, comparative	BETSY uses multinomial or Bernoulli Naïve Bayes models to classify texts into different classes (e.g. pass/fail, scores A-F) based on content and style attributes such as word unigrams and bigrams, sentence length, number of verbs, noun-verb pairs etc and is capable of auto analysis. Able to provide feedback but is often limited to terms such as extensive, essential, partial, unsatisfactory	While the text classification literature suggests the need to calibrate on thousands of cases per score group, accuracy of over 80% was achieved with the sparse dataset used in this study.
10	Perez [2, 4]	Automatic Assessor (AEA) Essay	To evaluate the performance of the system, an experiment was carried out using three essays types put together from courses on education, marketing	To assess essays written in Finnish by comparing the student's essay with a set of assignment-specific texts corpus. Secondly, to decrease the	Statistical analysis using a tokenizer and a sentence splitter.	Uses empirical data, learning materials and relatively few teacher-graded essays for calibrating the scoring mechanism	The representation to which compare the human graded essays and determine the threshold similarity values for each grade category.

Author(s)	Intervention type, and comparator	Study populations	Aims of study	Methodology	Outcome measures	Results
		and software engineering up to a total of 100-150 essays.	time in which students get feedback for their writings, and to reduce the costs of grading			The student's essay is measured by machine-human correlation in grading; the percentage of the grades given by the human assessor and the system.
11	Mitchell et al. [2,6]	Automark	Test items are drawn from archive material of science for pupils at age 11 was the domain explored. For the purposes of this study, four items of varying degrees of open-endedness were selected from the 1999 papers.	To provide robust marking in the face of errors in spelling, typing, syntax, and semantics.	A quantitative and qualitative approach with a view of the potential for further development in the future.	AutoMark employs the techniques of Information Extraction to provide computerised marking of short free-text responses. The study provided feedback to the system developers on improvement opportunities about the robustness of the system for real-world applications. It also provided insights to the test developers on the accommodations that might be needed to enhance the practicability the system.
12	Rudner, Garcia & Welch [2].	IntelliMetric	The IntelliMetric system performance is first compared to that of individual human raters using more than 750 responses to each of six prompts. The second, larger evaluation compares the IntelliMetric system ratings to those of human raters using approximately 500 responses to each of 101 prompts.	The evaluation examines the performance of the IntelliMetric system by comparing it to individual human raters. The paper also evaluated the IntelliMetric system reliability.	Uses three comparative approaches: i. True automated scoring system ii. The Bayesian approach which employs only simple word counts in building a model. iii. The probability approach provides a comparison with chance.	Results for the IntelliMetric system are compared to individual human raters, through Bayesian system using simple word counts, and a weighted probability model. Numeric using a number scale i. The IntelliMetric automated scoring system replicates the scores provided by human raters and produces superior perfect and adjacent agreement statistics. ii. It can identify "copied" essays. iii. Very few essays would need to be adjudicated if the IntelliMetric system were to be used to verify human ratings.

Author(s)	Intervention type, and comparator	Study populations	Aims of study	Methodology	Outcome measures	Results
						iv. Using 100 texts for training, they achieved 98% adjacent agreement. v. It assessed essays that are not written in English language, such as Hebrew attaining 84% correlation

5. CONCLUSION

While this study looked at the research methods used to present information on suitable AES; authors observed a substantial case for the use of AES in higher education as against few opposing authors. It identified different types of AES products, literatures and research designs employed in the investigation of AES products. Although, we did not observe any link between research methods and findings; all the papers however adopted literature review as a primer with majority using comparative analysis in addition. The outcome of reviewed literatures varied on suitability of AES in scoring essay task in Higher Institution of Learning. Since the case for and against the use of AES varied in some literatures, the authors of this study therefore posit that AES should be used with established triggers for human raters' intervention in exceptional cases. In order to achieve contactless interface between academics and student handled papers, the deployment and use of suitable AES could prevent exposure of human markers to paper borne SAR COV 2. The AES system is a software platform assessible on institutions intranet environment. Academic and markers should use suitable AES to assess and score assignments, written tasks or essay, dissertations, thesis or any academic written output with exceptional cases assessed manual as might be triggered by the AES.

ETHICAL APPROVAL

The authors declare that due diligence and compliance with high ethical standards was observed in the course of this study.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Osakwe K.A, Menkiti O.C and Ukaegbu K. Three-month Epidemiological Trend of Covid-19 in Nigeria -Rapid Review CJAST. 2020;39(26):9-18.
2. Rudner LM, Gagne P. An overview of three approaches to scoring written essays by computer. Practical Assessment, Research, and Evaluation. 2000;7(26):1-5.
3. Zupanc K, Bosnic Z. Advances in the field of automated essay evaluation. Informatica. 2015;39:383–395.
4. Page EB. New computer grading of student prose, using modern concepts and software. Journal of Experimental Education. 1994;62(2):127-142.
5. Rudner LM, Liang T. Automated essay scoring using Bayes' theorem. The Journal of Technology, Learning and Assessment. 2002;1(2):3–21.
6. Hearst M. The debate on automated essay grading. Intelligence Systems. 2000;1-37.
7. Arksey D, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology: Theory and Practice. 2005; 8(1):19-32.
8. Wang J, Brown MS. Automated essay scoring versus human scoring: A comparative study. Journal of Technology, Learning, and Assessment. 2007;6(2).
9. Zimmerman WA, Kang HB, Kim K, Gao M, Johnson G, Clariana R, Zhang F. Computer-automated approach for scoring short essays in an introductory statistics course. Journal of Statistics Education. 2018;26(1):40-47.
10. Rudner LM, Garcia V, Welch C. An Evaluation of the IntelliMetric essay scoring system. The Journal of Technology, Learning and Assessment. 2006;4(4):3–20.
11. Giles J. AI makes the grade. New Scientist. 2011;211(2828):22.
12. Hussein MA, Hassan HA, Nassef M. Automated language essay scoring systems: A literature review. Peer J Comput. Sci. 2019;5(208).

© 2021 Osakwe et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
<http://www.sdiarticle4.com/review-history/66677>