European Journal of Computer Science and Information Technology
Vol.11, No.1, pp.9-21, 2023
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)

Grading Multiple Choice Questions Based On Similarity Measure

Okure U. Obot[1], Gregory O. Onwodi[2], *Kingsley F. Attai[3], Anietie E. John[3] and Etiese Wilson[1]

[1] TETFUND Centre of Excellence on Computational Intelligence, University of Uyo, Uyo, Nigeria,
[2]Africa Centre of Excellence on Technology Enhanced Learning, National Open University of Nigeria, Nigeria,
[3]Ritman University, Ikot Ekpene, Nigeria

DOI : https://doi.org/10.37745/ejcsit.2013/vol11n1921 Published Feburary 12, 2023

Citation: Obot O.U., Onwodi G.O., Attai K.F., John A.E.and Wilson E. (2023) Grading Multiple Choice Questions Based On Similarity Measure, *European Journal of Computer Science and Information Technology*, Vol.11, No.1, pp.9-21

ABSTRACT: *Recent findings show the poor performances of students in Multiple Choice Questions (MCQ) examinations. This could be attributed to the administration of questions on application of knowledge rather than on fundamental of knowledge. Candidates doing these examinations at times choose options that are close to the actual answer but get zero (0) as the reward. The objective of developing a system that rewards a candidate based on the approximation of an option chosen not on the exactness of the option to the answer was therefore formulated in this study. 500 MCQs with their model answers, students' answers and scores obtained from two universities in Nigeria were collected. Jaro similarity measure was used to compute the degree of similarity between the model answers and the student answers. Results of the experiment show an average deviation of 13.3 marks. The adoption of this method of grading would be very beneficial in evaluating learners on e-learning platforms. The result is encouraging but could be improved upon with semantic similarity measure hybridized with string similarity.*

KEYWORDS: MCQ, Fuzzy logic, approximation, exactness, similarity measure, Jaro, semantic, examinations, average deviation, e-learning.

INTRODUCTION

Multiple choice question (MCQ) is a set of questions with a range of answers to choose from but requires a specific answer or answers from the choice options. MCQs are the most prevalent form of questions for numerous levels of assessment in electronic examinations and are extensively accepted for large-scale assessment in numerous domains and applications (Ch & Saha, 2020). MCQ are sometimes called Objective Response Question or Objective Test Question. MCQ are predominantly used in market research, customer reviews, elections, educational testing etc. and they appear in several different forms, depending on what it is used for or what it is intended to accomplish. MCQ consists of the stem and the answers. The stem is the problem to be solved or the question that is presented to the student or respondent which can appear in a form of an incomplete statement to be completed (filling in blank spaces). The answers consist of the correct responses or answers, known as *keys* and the incorrect responses or answers known as *distractors*, which are included to augment the choice options. Distractors are placed to confuse respondents from the correct answer due to their similarity to the correct answer (Ch & Saha, 2020; Susanti et al., 2017).

According to (Shah et al., 2017) distractors are generated by using context similarity derived with paradigmatic relation discovery on the self-made corpus and dictionary. However, points are awarded for correct responses and nothing is earned for incorrect responses even when the 'incorrect' option is very similar to the correct answer. In some cases, respondents are punished through fractional point deduction for incorrect responses and fractional points can be awarded for unanswered questions just to deter respondents from guesswork.

Human reasoning is based on approximate rather than on exact (Boolean logic) reasoning. Even on critical missions there are some elements of approximation as can be found in clinical surgery, launching of missiles and construction of high rising buildings. The study of fuzzy logic technology lends support to approximate reasoning such that most intelligent systems are today designed on the basis of fuzzy logic rather than on classical Boolean logic. In administering Multiple Choice Questions (MCQ), a student is expected to exhibit the traditional logic system of either it is correct or it is false. This has affected the performance of candidates in MCQ examinations especially those on electronic learning (e-learning) platforms such that some of them engage in guesswork to get an answer to a question. In order to check guesswork, some examiners have resorted to deducting fractional points for any wrong answer, leaving a candidate to either choose a correct answer or skip an answer to score zero instead of apply guesswork, which often result in point deduction. This study is intended to employ string similarity measure to evaluate the similarity of the model answer with the student's answer and award a mark based on the closeness of the student's answer to the model answer. This will eliminate the current trend of 'winner takes all' even when one chooses an option that is very close to the model answer but the examiner gives the candidate the same zero mark as the other person that chose an option that has no relationship to the model answer.

English words are ambiguous and their meanings can only be understood based on the context used. They are related to one another and the degree of their relationship is a measure of similarity between their pairs. The same is applicable to phrases, clauses, sentences, paragraphs and documents. Words could be similar lexically, if they have similar character sequence. They can be similar semantically, if they are of the same meaning, opposite in meaning to each other or used in the same context (Gomaa and Fahmy, 2013). Measuring similarity between texts can be categorized into topological; statistical similarity; semantic based; vector space model; word alignment based and machine learning based (Majumder et al., 2016). Some extensively used word level semantic similarity measures include; Jiang similarity, Resnik similarity, Lin similarity, Leacock similarity, and Wu similarity (Leacock & Chodorow, 1998; Resnik, 1995; Schoknecht et al., 2017; Lin, 1998) and the best results can be attained by aggregating a number of semantic similarity measures(Martinez-Gil, 2016 and Do & Rahm, 2002).

This study applies the Jaro similarity measure, a character-based measure, to compare the model answers provided for 500 MCQs in 5 subject areas taken by 250 students in 2 universities in Nigeria with the answers provided by the students. The remainder of the paper consists of related literature presented in Section 2, in Section 3, samples of the data collected along with the experiment conducted on Jaro similarity measure are presented. The analysis and discussion of results of the experiment are presented in Section 4 while the conclusion and recommendations of the study are presented in Section 5.

RELATED LITERATURE

Although several techniques have been adopted over the years for analyzing and evaluating short answers (de Assis Zampirolli et al., 2016; Dumal et al., 2017; Rasiq et al., 2019; Tavana et al., 2016), the importance of improving the grading of MCQs cannot be overemphasized. Ramachandran et al. (2015), presented an approach using word-order graphs to identify important patterns from human provided title texts and top-scoring student answers. The method utilized semantic metrics to determine groups of related words, for the representation of alternative answers. Basak et al. (2019) proposed a rule-based method, which relied on recognizing entailment relation between dependency structures of the two answers to assign grades. Le and Mikolov (2013) presented a supervised learning approach for automatic scoring of short answers based on the method of generating numeric fixed-length vector representations for variable length pieces of texts, known as paragraph embedding or Paragraph vectors. In Magooda et al. (2016), two word vector representations Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) were used for grading short answers. Adams et al. (2016) proposed an unsupervised approach for determining text similarity using one-to-many alignment of word vectors and the proposed technique often outperformed other compositional distributional semantics approaches as well as vector space methods such as latent semantic analysis. An unsupervised automatic short answer grading (ASAG) technique requiring sequential pattern mining and an intuitive scoring process was also proposed by Roy et al. (2016). This technique effectively exploited wisdom of students to deliver better performance than prior ASAG techniques as well as distributional semantics-based approaches that require heavy training with a large corpus. Bash (1999) review different marking schemes for MCQ in order to avoid gaining marks through guesswork. He suggested an alternative where marking schemes could be relaxed to reward candidates whose selections are nearly correct. One of such rewards is termed order of preference and confidence assessment where a candidates can be assigned confidence level to the choice they make in the alternative answers. Anyabolu and Okoye (2017) report the findings of using negative marking schemes in MCQ in health institutions in Nigeria. The schemes include informed negative marking but no negative marking (scheme A), informed no negative marking and no negative marking (scheme B) and informed negative and negative marking (scheme C). Results obtained show an average score of 46.2%, 57.7% and 39.9% respectively. Mckenna (2018) compared MCQ examination results with that of constructed response questions (CRQ) results obtained by the same set of students. There was a better performance in MCQ than in CRQ. Based on this, the study concluded that MCQ overestimated students understanding and recommended that MCQs have a role in informative assessment and should not be used for summative assessment

In Sultan et al. (2016), measures of text similarity are combined with grading specific constructs to produce top results on multiple benchmarks by computing a real-valued score for a student response, (Mohler et al., 2011) and assigning an annotation to the response to show the appropriateness of the answer to the question (Dzikovska et al., 2015). Chaturvedi & Basak (2021) proposed a method with three different models, using corpus-based semantic similarity measurements with each model scoring the students' responses individually. Mohler et al, (2011) combined several graph alignment features with lexical semantic similarity measures using machine-learning techniques to show that student answers can be more accurately graded than when semantic measures were used in isolation. Zhan et al. (2019), developed an automatic semi-open-ended short-answer grading model that integrates domain-general and domain-specific information, utilizing a long-short-term-memory recurrent neural network which learn the representation in the classifier and also considers word sequence information.

Sadr and Nazari (2019) proposed a method for improving the performance of short answer grading systems based on semantic relatedness and similarity measures that leverages students' answers with the highest score as feedback. Students' answers when used as feedback can improve the precision of semantic relatedness and similarity measures in the automatic assessment of examinations with short answers. In Suzen et al. (2020), an automatic short answer grading is undertaken by applying standard data mining techniques to the corpus of student answers to measure the similarity between the student answers and the model answer. The model also predicts marks based on the similarities between the student answers and the model answer. Obot et al (2021) employed Jaccard, Jaro, Dice and Cosine similarity measures to test the similarity between short answers provided by students and model answers given by examiners in 647 questions in 7 examinations. Results show that Jaro measure ranked closest to the mean score of the 3 examiners with variance absolute error of 0.62% and covaried strongly by 97% with 0.001 significant level. Olowolayemo et al (2018) used the Cosine and Levenshtein distance (LD) in assessing 240 students' short (2-5 words) answers of 10 questions for each student. Results obtained show an accuracy of 92% and 94% for for LD and Cosine respectively.

Semantic similarity, also known as semantic nearness, semantic proximity or semantic closeness is the ability to determine the similarity between various terms such as words, sentences, documents, concepts or instances. According to (Martinez-Gil, 2016), semantic similarity measurement aims to determine the likeness between two text expressions that use different lexicographies for representing the same real object or idea. Semantic similarity measure has vast significance in many fields such as information retrieval, educational system, text summarization and natural language processing (Ali et al., 2018).

Data Collection and the Experiment

Five subject areas with 100 questions each were collected from the general studies directorate in two institutions of higher learning in Nigeria. In addition, answers and scores to the answers from 250 students were also collected from the directorates. A sample of the questions and scores are shown in Appendix A. The answers of individual student were subjected to comparison with that of the model answers provided by the examiners with a view to finding the degree of similarities between the option chosen by the student and the model answer provided by the examiner. Jaro similarity measure was used in finding the similarity due to its high degree of similarity with manual scoring shown in existing literature (Obot et al., 2020; Obot et al., 2021)

Jaro measure is given as:

$$\text{Jaro} = \frac{1}{3}\left(\frac{c}{3}+\frac{c}{t}+\frac{x-c}{c}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where:

s = model answer to a question
t = student's answer to a question
c = common character between s and t
x = the number of transpositions.
Jaro (1989).

First, it computes the length of the two strings |*s*| and |*t*|. Second, it finds the common characters (c) between the two strings; two characters match if they are the same and located no farther than [*max* (|*s*|, |*t*|)/2] – 1 in the string. Third, it finds the number of transpositions ($x = m/2$), which is the number of matching characters (m), but in reverse order (*a*/*u*, *u*/*a*) (Gali et al., 2016).

ANALYSIS AND DISCUSSION OF RESULTS

For each subject, there are 100 questions and each question carries 1 mark giving a total of 100 marks. The results of what each student scores for each subject is presented in Appendix B. The presentation shows what a student scored when the conventional (Boolean) marking method was used and when the text similarity method is used. It also shows the average score for each of the two methods and the average deviation of the scores of the conventional from the test similarity methods. The average deviation is computed as:

$$Avedev=\left(\frac{x}{y}\right)/(n) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where:

Avedev = average deviation
x = the total scores for the text similarity measure's method
y = the total scores for the conventional method
n = the total number of students that have the scores

The mean of the average deviation is 13.3marks. This represents the average mark that a student will gain if the proposed method is adopted. In the proposed method, an answer that is 0.70 similarity and above is added to the candidate whereas such is lost in the conventional method. The 0.70 and above score shows that the option taken by the candidate is similar to the exact answer by 70%. This is a deviation from the conventional method where a candidate has to score an exact answer before he is considered to have scored a mark otherwise he scores a zero (0) mark. The threshold of 0.70 is considered after going through the questions, options and answers and ascertaining that the values within the threshold. This compares favorably with human judgement on the similarity of the model answers and options taken by the students and scores computed by Jaro similarity measure though there are few outliers. Examples of these are as found in question 21 on ICT, the model answer is “Bit”, a candidate who chose bits scored 1 mark, another who chose byte scored 0.70 marks the one who chose beat scored 0.39. Question 26 Model answer is Transistor technology, a student who chose “Transistor” scored 0.72 mark whereas in the conventional method the student scored 0. In English Language; Question number 76, the model answer is “Visits”, the candidate that chose “Visit” scored 0.75 but in the conventional, the student scored 0. In Agricultural Science; question number 26, the model answer is “Forage crop” the student who chose “fibre crop” scored 0.6 but in the conventional method the candidate scored zero (0).

CONCLUSION

In this research, five hundred multiple-choice questions, their model answers and answers and scores of individual students were collected from two universities. The model answers and the answers of individual students were subjected to Jaro, a character-based similarity measure for finding the degree of similarity between the two. Results show different degree of similarities ranging from zero (0) to one (1), i.e. [1,0]. The results depict the nearness of the candidate’s answer to the model answer and in effect the closeness of the student’s intelligence to the exact answer. Conventionally, MCQ is measured by the exactness of the answers or otherwise resulting in poor performance of students in examinations that employ this approach to setting and administering examinations. Recent study shows that this approach is becoming prevalent due to its simplicity in conducting and marking the examinations through computerization especially in an era of electronic learning and technology enhance learning.

Results of the experiment conducted with Jaro similarity measure shows that there is an average deviation of 12.4, 13.4, 14.9, 13.8, and 16.7 for ICT, Use of English, Library studies, Agricultural science and Nigerian culture, philosophy and logic respectively. This gives an overall average deviation of 13.3 marks. This means that if Jaro similarity measure is used in marking MCQs every candidate’s score will be improved by 13.3 marks on the average. While text similarity could be based on strings, corpus and knowledge, according to Gomaa and Fahmy (2013), a corpus based similarity is a semantic similarity measure that determines the nearest neighbor between words, phrases and sentences according to the information gained from large corpora while a string based measure operates on string sequences and character compositions. WordNet uses a knowledge based similarity measure that is based on identifying the degree of similarity between words using information derived from semantic network. (Mihalcea et al, 2006). String or character based measures at times could be misleading to a wrong answer, example, the word ‘bit’ and ‘bite’ have nothing in common in terms of their meaning but could be ranked very close to each other using a string/character based measure such as Jaro, Jaccard, Cosine, Dice.

In order to have a better result that enhances grading of MCQ, further research is recommended where corpora of terminologies of different subject areas are used to perform semantic similarities and hybridized with Jaro measure to produce a human-like cognitive ability. With the popularity of MCQs and the currency of electronics examinations, educational managers and evaluators should give deep consideration on how to improve on the administration of

MCQs examinations and grading of results to give the candidates due advantage without compromising standard. The implementation of this study has helped to bring into fore the application of similarity measures and its promising features to develop a human cognitive-like system especially in the area of education evaluation.

References

Adams, O., Roy, S., and Krishnapuram, R. (2016, December). Distributed Vector Representations for Unsupervised Automatic Short Answer Grading. In *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)* (pp. 20-29).

Ali, A., Alfayez, F., and Alquhayz, H. (2018). *Semantic Similarity Measures between Words: A Brief Survey*. Sci. Int. (Lahore),30(6), 907-914, 2018 ISSN 1013-5316;CODEN: SINTE 8

Anyabolu E. N. and Okoye I. C (2017). Modified Negative Marking Schemes in Multiple Choice Questions in a Health Institution in Nigeria. Clinical Medicine and Diagnosis 7(4): 101-105

Basak, R., Naskar, S. K., & Gelbukh, A. (2019). Short-answer grading using textual entailment. *Journal of Intelligent & Fuzzy Systems*, *36*(5), 4909–4919. https://doi.org/10.3233/JIFS-179038

Bash M.(1999). Alternative Marking Schemes for On-line Multiple Choice Question Tests. In 7th Annual Conference on the Teaching of Computing. Belfast. 101-103.

Ch, D. R., and Saha, S. K. (2020). Automatic Multiple Choice Question Generation From Text: A Survey. *IEEE Transactions on Learning Technologies*, *13*(1), 14–25. https://doi.org/10.1109/TLT.2018.2889100

Chaturvedi, B.,and Basak, R. (2021). Automatic Short Answer Grading Using Corpus-Based Semantic Similarity Measurements. In C. R. Panigrahi, B. Pati, P. Mohapatra, R. Buyya, & K.-C. Li (Eds.), *Progress in Advanced Computing and Intelligent Engineering* (Vol. 1199, pp. 266–281). Springer Singapore. https://doi.org/10.1007/978-981-15-6353-9_24

de Assis Zampirolli, F., Batista, V. R., & Quilici-Gonzalez, J. A. (2016). An automatic generator and corrector of multiple choice tests with random answer keys. *2016 IEEE Frontiers in Education Conference (FIE)*, 1–8. https://doi.org/10.1109/FIE.2016.7757467

Dumal, P. A. A., Shanika, W. K. D., Pathinayake, S. A. D., & Sandanayake, T. C. (2017). Adaptive and automated online assessment evaluation system. *2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 1–8. https://doi.org/10.1109/SKIMA.2017.8294135

Do, H. H., & Rahm, E. (2002). COMA—a system for flexible combination of schema matching approaches. In *VLDB'02: Proceedings of the 28th International Conference on Very Large Databases* (pp. 610-621). Morgan Kaufmann.

Dzikovska M., Nielsen R., and Leacock C. (2015). The Joint student Response Analysis and recognising textual entailment Challenge: making sense of student responses in educational applications. Language Resources and Evaluation 50(1), DOI: 10.1007/s10579-015-9313-8

Gali, N., Mariescu-Istodor, R. and Fränti, P. (2016) Similarity measures for title matching, *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancún Center,

Gomaa, W.H. and Fahmy, A.A. (2013) 'A survey of text similarity approaches', International Journal of Computer Applications, Vol. 68, No. 13, pp.75–87.

Jaro, M.A. (1989) Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, *Florida. Journal of the American Statistical Association*, 4(406),.414–420.

Le, Q. and Mikolov, T.(2014).Distributed representations of sentences and documents," in In Proceedings of the 31st International Conference on Machine Learning (ICML 2014), Beijing, China, 2014.

Leacock, C., Chodorow, M.(1998). Combining local context and WordNet similarity for word sense identification. In: WordNet: An electronic lexical database, MIT Press, (1998).

Lin, D.(1998). An information-theoretic definition of similarity. In: Proceedings of the 15th International Conference on Machine Learning, pp. 296-304, (1998).

Magooda, A. E., Zahran, M., Rashwan, M., Raafat, H., & Fayek, M. (2016). Vector based techniques for short answer grading. In The twenty-ninth international flairs conference.

Majumder, G., Pakray, P., Gelbukh, A., & Pinto, D. (2016). Semantic Textual Similarity Methods, Tools, and Applications: A Survey. *Computación y Sistemas*, *20*(4). https://doi.org/10.13053/cys-20-4-2506

Martinez-Gil, J. (2016). CoTO: A novel approach for fuzzy aggregation of semantic similarity measures. *Cognitive Systems Research*, *40*, 8–17. https://doi.org/10.1016/j.cogsys.2016.01.001

Mckenna P (2018). Multiple Choice Questions: Answering correctly and knowing the Answer. In international conference on e-learning. ISBN: 978-989-8533-78-4

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. (2013). Efficient estimation of word representations in vector space. In proceeding of the International Conference on Learning Representations Workshop track, Arizona, USA.

Mihalcea, R., Corley, C. and Strapparava, C. (2006) 'Corpus based and knowledge-based measures of text semantic similarity', Proceedings of the American Association for Artificial Intelligence, Boston, MA, Vol. 8, No. 7, pp.19–38.

Mohler, M. A. G., Bunescu, R., & Mihalcea, R. (2011). *Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments*. Proceedings of the 49th Annual meetings of the Association for Computational Linguistics, Portland, oregon, July, 19-24, 2011.

Obot, O.U., Udoh, S.S. and Attai, K.F. (2021) The suitability of similarity measures to the grading of short answers in examination, *Int. J. Quantitative Research in Education*, 5(3), 207–222.

Obot, O. U, Uzoka, F. M. E, John, A. and Udoh, S. S (2020) Soft-Computing Method for settling Land Disputes Cases Based on Text Similarity Measures. International Journal of Business information Systems, 1(1) 1-12

Olowolayemo, A., Nawi, S.D. and Mantoro, T. (2018) 'Short answer scoring in English grammar using text similarity measurement', *2018 International Conference on Computing,Engineering, and Design (ICCED)*, pp.131–136, IEEE [online] https://ieeexplore.ieee.org/ abstract/document/8691068/ (accessed 24 April 2020).Cancún, México, 4–8 December, 978-1-5090-4846-5/16/$31.00 ©2016.

Ramachandran L., Cheng J and Foltz, P. (2015). Identifying Patterns for Short Answer Scoring using Graph-based Lexico-Semantic Text matching. Proceedings of the Tenth Workshop on Innovative use of NLP for Building Educational Applications held in Denver Colorado, June 4,2015, Association for computational Linguistics , 97-106.

Rasiq, G. M. R. I., Al Sefat, A., & Hasnain, M. M. F. (2019). Mobile based MCQ Answer Sheet Analysis and Evaluation Application. *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 144–147. https://doi.org/10.1109/SMART46866.2019.9117468

Resnik, P.(1995). Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 448-453. Quebec, Canada..

Roy, S., Dandapat, S., Nagesh, A., & Narahari, Y. (2016). Wisdom of students: A consistent automatic short answer grading technique. In Proceedings of the 13th International Conference on Natural Language Processing (pp. 178-187).

Sadr, H., & Nazari Solimandarabi, M. (2019). Presentation of an efficient automatic short answer grading model based on combination of pseudo relevance feedback and semantic relatedness measures. *Journal of Advances in Computer Research*, *10*(2), 17-30.

Schoknecht, A., Thaler, T., Fettke, P., Obserweis, A., Laue, R.(2017) Similarity of Business Process models – A State-of-the-Art Analysis. ACM Computing Survey, 50(4), pp. 1-33, (2017).

Shah, R., Shah, D., Kurup, L. (2017). Automatic Question Generation for Intelligent Tutoring Systems. In: the 2nd International Conference on Communication Systems, Computing and it Applications (CSCITA), pp. 127–132. https://doi.org/10.1109/CSCITA.2017.8066538.

Sultan, M. A., Salazar, C., & Sumner, T. (2016). Fast and Easy Short Answer Grading with High Accuracy. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1070–1075. https://doi.org/10.18653/v1/N16-1123

Susanti, Y., Tokunaga, T., Nishikawa, H., & Obari, H. (2017). Evaluation of automatically generated English vocabulary questions. *Research and Practice in Technology Enhanced Learning*, *12*(1), 11. https://doi.org/10.1186/s41039-017-0051-y

Süzen, N., Gorban, A. N., Levesley, J., & Mirkes, E. M. (2020). Automatic short answer grading and feedback using text mining methods. *Procedia Computer Science*, *169*, 726-743.

Tavana, A. M., Abbasi, M., & Yousefi, A. (2016). Optimizing the correction of MCQ test answer sheets using digital image processing. *2016 Eighth International Conference on Information and Knowledge Technology (IKT)*, 139–143. https://doi.org/10.1109/IKT.2016.7777754

Zhan, Y., Lei, W., & Guan, Y. (2019). A QoE Evaluation Method for RT-HDMV Based on Multipath Relay Service. *Symmetry*, *11*(9), 1127. https://doi.org/10.3390/sym11091127

European Journal of Computer Science and Information Technology

Vol.11, No.1, pp.9-21, 2023

Print ISSN: 2054-0957 (Print),

Online ISSN: 2054-0965 (Online)

Appendix A: Sample MCQs

ICT 21. _ is the smallest unit of measuring data in the computer. a) Byte (b) bite (c.) beat **(d) bits** 24. Which of the following is a component of the CPU? _ a.) RAM b.) Memory Unit c.) ALU d.) Hard disk 66._ defines a set of rules and signals that computers on the network use during communication (a) Software (b) Ethernet c) Protocol d.) Communication 67. _ converts analog signals to digital signals and vice versa a) MODEM b) BROWSER c) Signal Converter d.) Demodulator 75. How many characters can a file of 2 kilobytes holds _ (a) 200 characters (b) 2,000 characters (c) 2048 characters (d) 2024 characters 91. Which of these devices is used in the supermarket at the point of sales _ (a) Optical Character Reader (b) Magnetic Character Reader (c) Light Pen (d) Graphic Tablet
Use of English 1 _ is the part of the sentence, which performs the action of the verb. (a) object (b) Verb (c) subject (d) adjuncts 9. _ indicates possession or ownership. (a) Coherence (b) rules (c) apostrophe (d) period. 19. A normal essay comprises of _ parts. (a) 4 (b) 3 (c) 2 (d) 5. 38. Words that are coined and used for dramatic effects are _a). Parts of speech b). Figure of speech c). Speech d). Language 47._compares two opposite ideas for contrast and emotional effect. a). Antithesis b). Irony c). Simile d). euphemism 58.What type of essay would you use in writing “my first day in Ritman”_ (a)Argumentative (b)Narrative (c)Descriptive (d)Expository
Library studies 10. Which among these is not a study skill? _ (A) Note-taking (B) Chatting (C) Reading Comprehension (D) Study Groups 14. The following are names for virtual library except one _ (A) E-library (B) Electronic library (C) Mini library (D) Library without walls 19. The legal right granted the owner of intellectual property is called _ (A) Plagiarism (B) Financial incentive (C) Copyright (D) All of the above 25. Database resources are found in _ (A) Virtual library (B) Circulation section (C) Reference section (D) Serials section 50. Questionnaire is a _ (A) Research method (B) Measurement technique (C) Tool for data collection (D) Data analysis technique
Agricultural Science 3. Instrument used for measuring wind velocity is known as _(A) Hydrometer (B) Pyranometer (C) Altimeter (D) Anemometer 7. Agricultural land can be acquired through the following EXCEPT _ (A). Communal land tenure (B). Rent tenancy (C). Pass land (D). sacred land 19. Which is the highest digestible protein nonleguminous crop among the following _(A) Napier(B) Maize silage(C) Maize(D) Iowar 31.The irrigation efficiency of sprinklers depend upon the of water application_ (A) Degree of uniformity(B) Quality(C) Size(D) Quantity
Nigerian Culture, philosophy and Logic 6.The tending of cattle, sheep or goats or a combination of all, is referred to as _ ?(a)Fallow(b)Animal husbandry/Pastoralism(c)Plant husbandry/Pastoralism(d)Agronomy

Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)

9.The leader of the king makers was referred to as the _ ?(a)Boshorun (b)Bashorun (c)Bushorun (d)Vashorun
15.Social inequality was for the Igbo pre-colonial society, what social stratification was for _ ?pre-colonial society?(a)Ibibio (b)Benin (c)Yoruba (d)Hausa/Fulani
25.The inescapable end to man's sojourn here on earth is known as _ ? (a)Peace (b)Paradise (c)Life (d)Death
27. One of the following is not a characteristic of culture_? a) shared, b) learned, c) earned, d) integrated

Appendix B: Results

ICT			Use of English		
S/n	**Enhanced**	**Conventional**	**S/N**	**Enhanced**	**Conventional**
1	59.9	51	1	40.1	28
2	44.4	32	2	45.7	33
3	44.5	31	3	35.5	21
4	53.5	40	4	38.4	24
5	48.2	36	5	31.9	19
6	48.2	36	6	40.1	29
7	50.6	39	7	37.4	24
8	52.6	41	8	43.2	29
9	44.1	32	9	35.7	24
10	47.0	34	10	43.9	31
11	45.4	33	11	43.6	30
12	50.6	39	12	43.2	28
13	51.0	40	13	35.7	23
14	44.6	32	14	39.9	24
15	42.2	30	15	44.6	31
16	45.3	31	16	38.3	23
17	53.0	44	17	30.6	18
18	47.3	35	18	42.6	30
19	46.3	32	19	33.8	22
20	56.3	41	20	34.2	20
21	48.7	38	21	33.9	19
*22	49.7	38	22	33.4	23
23	52.6	40	23	39.1	23
24	43.7	33	24	39.9	28
25	47.9	35	25	41.9	26
26	44.6	31	26	37.6	22
27	55.7	36	27	40.3	26
28	47.3	35	28	42.5	31
29	44.6	31	29	42.9	31
30	51.8	39	30	46.3	32
31	48.7	39	31	36.2	22
32	50.6	38	32	47.1	36
33	52.6	41	33	47.0	31

@ECRTD-UK: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

34	41.1	29	34	37.7	22
35	48.0	35	35	35.3	22
36	46.9	34	36	47.3	29
37	44.5	34	37	41.9	26
38	51.7	41	38	46.1	34
39	50.7	39	39	42.4	28
40	45.6	34	40	36.6	24
41	53.6	41	41	40.8	29
42	46.3	32	42	42.7	30
43	48.3	36	43	28.7	17
44	54.0	45	44	38.5	25
45	44.3	30	45	33.0	21
46	44.3	30	46	40.2	30
47	53.9	44	47	37.7	25
48	48.3	36	48	42.9	29
49	43.6	30	49	34.7	22
50	53.9	42	50	43.8	31
Average	48.7	36.3	**Average**	39.5	26.0
AVEDEV	12.4		**AVEDEV**	13.4	

Agricultural Science			**Nigeria Culture, Philosophy and Logic**		
S/N	**Enhanced**	**Conventional**	**S/N**	**Enhanced**	**Conventional**
1	56.6	46	1	78.2	59
2	44.3	34	2	73.5	55
3	38.5	23	3	71.7	51
4	56.3	48	4	61.0	46
5	45.4	31	5	72.4	61
6	54.7	43	6	64.8	46
7	49.4	39	7	61.3	46
8	54.7	44	8	59.0	37
9	54.7	41	9	65.9	47
10	39.5	23	10	62.3	47
11	47.2	30	11	72.5	58
12	53.5	43	12	70.8	56
13	34.1	16	13	76.4	62
14	21.4	13	14	65.4	47
15	38.4	18	15	69.2	55
16	40.4	23	16	80.8	73
17	36.8	17	17	79.7	70
18	44.9	22	18	83.1	68
19	45.9	23	19	77.9	66

20	38.1	19
21	56.7	45
22	41.5	22
23	52.7	41
24	51.4	38
25	48.5	33
26	51.6	37
27	51.6	37
28	53.5	38
29	43.3	29
30	39.8	29
31	41.8	28
32	52.5	37
33	55.0	44
34	46.8	36
35	57.1	47
36	58.3	49
37	52.0	38
38	60.1	51
39	55.9	42
40	64.3	50
41	54.4	40
42	61.0	46
43	56.7	45
44	56.2	46
45	50.2	37
46	53.8	42
47	57.1	43
48	56.2	44
49	53.9	42
50	51.8	38
Average	49.5	35.8
AVEDEV	13.8	

20	81.1	70
21	60.2	40
22	68.2	47
23	70.4	50
24	70.9	52
25	72.0	54
26	53.2	35
27	55.1	38
28	58.3	41
29	59.2	41
30	69.5	46
31	57.0	39
32	55.6	36
33	56.5	37
34	55.5	35
35	54.3	35
36	54.2	34
37	62.8	50
38	62.2	53
39	52.6	38
40	54.7	34
41	62.2	51
42	77.2	66
43	74.7	58
44	59.8	47
45	57.5	45
46	61.4	42
47	70.8	54
48	74.1	55
49	66.1	45
50	68.2	47
Average	49.3	66.0
AVEDEV	16.7	

Acknowledgement
We acknowledge the TETFUND centre of excellence on computational intelligence and the Africa centre of excellence for technology enhanced learning (ACETEL) for funding this research work.