Journal of Information Systems Engineering and Management
2025, 10(27s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Automated Multiple-Choice Test Checking System

Prof. John B. Lacea[1], Dr. Thelma D. Palaoag[2]
[1] *Assistant Professor, Department of Information Technology, College of Information Sciences, Benguet State University, Benguet, Philippines.*
[2] *Associate Professor, Department of UC Innovation and Technology Transfer, College of Information Technology and Computer Science, University of the Cordilleras, Baguio City, Philippines.*
Email: [1] jblacea2007@gmail.com, [2] tdpalaoag@uc-bcf.edu.ph
Orchid Id number: [1]0009-0007-0937-8236, [2]0000-0002-5474-7260
Corresponding Author: John B. Lacea.*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 15 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

This research study was conducted mainly to address the problems of checking test papers manually by teachers. The researchers developed a software application that offers fast and convenient way to check multiple-choice test. It applied the OCR as alternative solution for detecting the answer marks. The research design implemented the quantitative research with statistical analysis. The rectified answers detection algorithm is an improvised design solution that compensate the imperfection of the OCR. The algorithm rectifies the misinterpreted character and maps the matching letter as proper answer. There is important differentiation among automated and manual checking of test papers on efficiency. There is greatly important differentiation among color of pen with complete shading and color of pen with X symbol answer marks on the precision of OCR.

Keywords: rectified answers detection algorithm; optical character recognition; automated checking; multiple-choice test

I. INTRODUCTION

Teaching is a self-sacrificing career because it has great responsibilities to achieve for the future of our students and also for the progress of our country. Teachers are always carrying lots of work from school to home. Besides teaching, checking quiz and/or exam papers is a demanding job in order to evaluate the performance of students. Ise observes that high volume of exam papers to examine is laborious, wearisome, takes lengthy time and susceptible to mistakes [1]. Donna points out that too much worked and pressured through subject overload, creating lessons, making assessments and checking exam papers cause mental and physical exhaustion that affected the health life of every teacher [2].

The computerized system would carefully perform the task to lessen the immense consequence of inaccurately calculated student grades in student academic performance in the university [1]. Limpuasan et al. perceive that it waited one week to collect the scores of examinations at the Sucat Elementary School, Muntinlupa City, Philippines. Teachers make errors when registering each student exam scores which some students accidentally received low grade [3].

In South Korea, the company BL SOFT Co., Ltd. sold Optical Mark Reader (OMR) for automated checking of exam papers to schools. In the past, teachers have long time to check and analyze the students' exam results. Nevertheless, the work productivity greatly improved when they applied the automated checking system [4].

The antiquated manual answer script correction is arduous and lethargic. Human efforts are utilized and squandered [5]. According to Tanwar, there are two(2) types of examinations which are objective and subjective. Modest exams are multiple-choice test and run online with great number of students take the test [6]. Rosebrock made a bubble sheet multiple-choice checker using Python, OpenCV and OMR. The Optical Mark Recognition (OMR) is the method of identifying human-marked data and translating its meaning [7].

Shrestha et al. show that automatic answer checking using Natural Language Processing(NLP), the answers based on keywords are equally distributed marks which are input data minimize the time and manpower in checking [8]. A similar study conducted by Bansal et al. assess descriptive answers using NLP. The NLP algorithm uses word patterns from the teacher's answer in checking the student's answer. The AI assigns score in which as good as score set by the teacher [9].

Zampirolli et al. use mathematical morphology to detect answer marks captured from digital camera and attained great precision in the results using MATLAB commands [10]. Tavana et al. write that multiple-choice test was one of the general procedures to assess the student's performance. A very costly tool is crucial to check the students' answer sheets. They recommend using mathematical morphology, neighborhood and thresholding to enhance the checking method that boost the detection for unfinished answer marking will attain high precision [11]. Likewise, China et al. develop Android application using OpenCV library that corrects multiple-choice test answer sheets and offered outstanding outcomes. It immediately shows the accurate grade from the test and has statistics to explore the scores. The application implemented the image segmentation from Mathematical Morphology that showed how inadequacy resulting from images captured in imperfect conditions can be avoided, providing good results compared to experiments managed in perfect conditions [12].

Cuerdo et al. concluded that there was substantial difference between alternative OMR and manual processes in the productivity. The difference was very high undeniably in prefer of alternative OMR [13]. Nagy et al. stated that OCR has poor accuracy even though it is commonly used. The results given by the OCR need to be corrected manually and some portions of the documents are discarded in approval of worker account [14].

The proponent tried and tested the Optical Character Recognition(OCR) technology as another way in developing the software application that automatically checks multiple-choice tests. The software application manages multiple-choice test only. The student chooses one letter as answer by completely shading the circle using a black color of pen on the answer sheet. The answer sheets are scanned by a computer scanner device. The software application detects the completely shaded answers as letters and then computes the scores. The minimum items is 15 points and the maximum items is 80 points. The paper size of the answer sheet is either A4 or Letter size.

Research Objectives

The core purpose of this research study is to devise and make a software application that automatically checks multiple-choice tests.

Specifically, the research aimed to:

1. find the level of problems in manual checking of test papers;
2. devise and make a software application that checks completely shaded answers implementing the OCR technology;
3. find the important differentiation among automated and manual checking of test papers on efficiency; and
4. find the important differentiation among color of pen with complete shading and color of pen with X symbol answer marks on the precision of OCR.

II. METHODOLOGY

Research Design

The research study implemented the quantitative research and statistical analysis to explain the level of problems in manual checking of test papers and assess the efficiency among automated and manual checking of test papers and also on the precision of OCR among color of pen with complete shading and color of pen with X symbol answer marks.

Sources of Data

The proponent collected information on problems in manual checking of test papers using survey questionnaires. There were 170 respondents composing of instructors and teachers at Benguet State University in the academic school year 2018-2019 and August to September of the year 2022.

The proponent conducted multiple-choice test exam and carried out comparisons between automated and manual checking of test papers on efficiency and the precision of OCR using the developed multiple-choice test software application. The test was 40 items and 85 students took the exam.

Data Collection Instruments

The tools used to collect and process information were Likert Scale, p-value and survey questionnaires. The face validity was implemented to check the survey questionnaires by 3 experts. The proponent tried and tested 4 different colors of pen and 2 answer marks as shown on Figure 1.

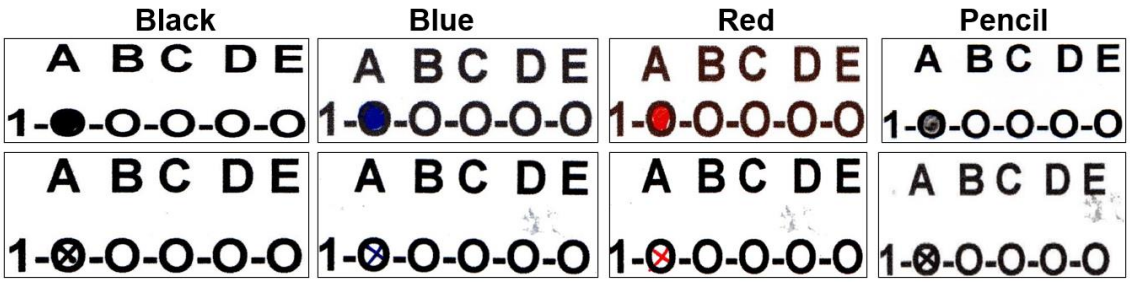

Figure 1 Pen Colors and Answer Marks

Treatment of Data

The treatment of data applied the statistical analysis such as the repeated measures t-Test and average or mean. To interpret the level of problems in manual checking of test papers, the average was calculated from the list of problem statements encountered. The computed averages were translated applying the range scale for level of problems on Table 1.

Table 1 Level of Problems Range Scale

Range Scale	Interpretation
4.21 to 5.00	Severe Problem
3.41 to 4.20	Serious Problem
2.61 to 3.40	Moderate Problem
1.81 to 2.60	Minor Problem
1.00 to 1.80	No Problem

The checking of each test paper was timed on both manual and automated. The averages of their time were calculated among automated and manual checking of test papers. The computed averages were translated by applying the range scale for efficiency on Table 2. The repeated measures t-Test was applied in order to find the important differentiation on efficiency.

Table 2 Efficiency Range Scale

Range Scale	Interpretation
1.00 to 1.80	Very Efficient (≤ 1 min)
1.81 to 2.60	Efficient (2 mins ≥)
2.61 to 3.40	Moderately Efficient (3 mins ≥)
3.41 to 4.20	Somewhat Inefficient (4 mins ≥)
4.21 to 5.00	Inefficient (5 mins ≥)

The 4 different colors of pen with complete shading and X symbol were tested. The averages of their OCR precision were calculated. The repeated measures t-Test was applied in order to find the important differentiation among color of pen with complete shading and color of pen with X symbol answer marks on the precision of OCR.

III. RESULTS AND DISCUSSION

Find the level of problems in manual checking of test papers

The results in finding the level of problems in manual checking of test papers are shown on Table 3. The findings show that having large quantities of test papers to check is a severe problem. The majority of teachers have more than 7 subjects with greater than 40 students per subject to teach every semester. The lack of free and available time to concentrate in checking test papers is a severe problem. There is no academic break after the exam and teachers are passionate to teach the subsequent topics. The rest days, Saturday and Sunday are not adequate to complete the checking. Teachers are workaholic that they have hectic obligations and responsibilities to work on. They experienced mental and physical exhaustion that aggravate the deferment, procrastination and correctness of checking. Additional problems encountered stated such as unchecked, assignments, quizzes and exercises increased the quantities to check and affected by illness that hinders the easiness and comfort while checking.

Table 3 Level of Problems in Manual Checking of Test Papers

Problems Encountered	Average	Interpretation
1. You have large quantities of test papers to check.	4.25	Severe Problem
2. You are too slow to check.	3.5	Serious Problem
3. You have errors in totaling the correct scores.	2.75	Moderate Problem
4. You have errors in checking for example you corrected a wrong answer or you mistaken the right answer.	3	Moderate Problem
5. You are affected by monotony, dullness and tediousness when you checked in extended period of time.	3.75	Serious Problem
6. You have other hectic obligations and responsibilities to work on.	4	Serious Problem
7. You lack free and available time to concentrate in checking test papers.	4.25	Severe Problem
8. You have delayed schedule of exam that aggravates the backlog of test papers.	2.5	Minor Problem
9. You have mental and physical exhaustion in your daily teaching.	3.5	Serious Problem
10. You need lots of time to take a break for your good health.	4	Serious Problem
11. Additional problems that you encountered, please state • here are unchecked assignments, quizzes and exercises that increase the quantities to check. • ffected by illness that hinders the easiness and comfort while checking.	3 4	Moderate Problem Serious Problem

Devise and make a software application that checks completely shaded answers implementing the OCR technology

The appropriate tangible components on the device it encompasses is known as physical layer [15]. The physical layer in Figure 2, shows the students answer the exam. Using a printer scanner device, the answer sheets are converted into images and stored in the computer. The developed software application read the image and checks the completed shaded answers.

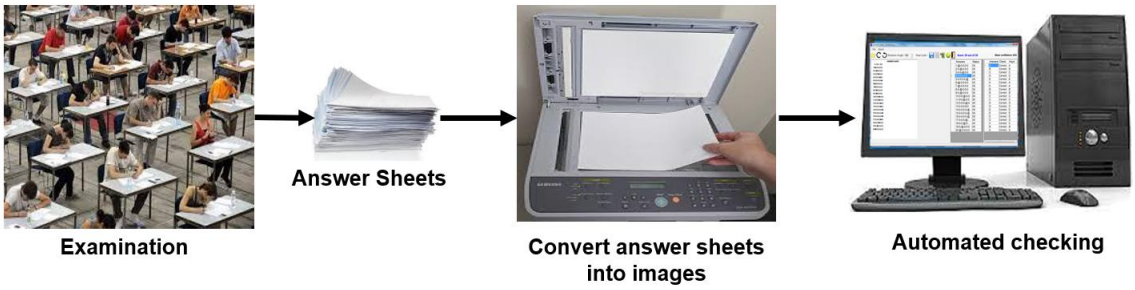

Figure 2 Physical Layer

The functions, features and requirements of the software is known as logical layer [15]. The logical layer in Figure 3, shows the OCR detects the completed shaded answers from the image which is the answer sheet. The checker compares the detected answers to the key answers and counts the right and wrong answers.

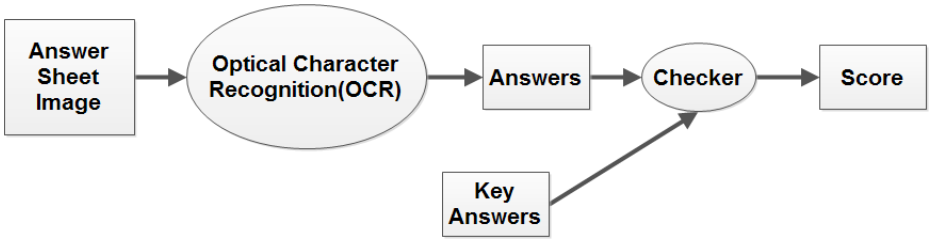

Figure 3 Logical Layer

Rectified Answers Detection Algorithm

The OCR technology has imperfection to detect complete shaded answers on the answer sheet. The OCR interprets the complete shaded circle as character such as "©", "8", "®9", "@", "®", etc. The OCR interprets the unfilled circle as characters such as “oO”, "o", “Oo”, "O". At other times, the OCR does not detect it. To solve the flaw of the OCR, an invented solution was devised to amend the inappropriate interpretations. The rectified answers detection algorithm has logical sequence of steps to rectify the misinterpreted character and to map the matching letter as proper answer. The algorithm validates the correct number of shaded answers as shown on Table 4.

Table 4 Rectified Answers Detection Algorithm

Algorithm: Rectified Answers Detection
Input: The $S = \{x_i\}_{i=1}^{N}$ is the set S that contains the N scanned characters from the answer sheet by the OCR. The set T is an array of detected answers. The set V is an array of the status of the detected answers. Output: Set T
Step 1. Get s_i character from the set S and then analyze and rectify "oO" or "Oo" or "oO" to zero ‘o’ or letter ‘O’.
Step 2. Get s_i character from the set S and confirm if it has zero 'o' or letter 'O'. Count the number of characters, if the total is 4 then it is correct. Otherwise, it has greater than 1 answer or there is no answer.
Step 3. If there is 1 answer then set the status V_i to "OK" which is valid answer. Otherwise, set the status V_i to "?" which is unrecognized answer.
Step 4. If there is a valid answer then map the T_i to the matching letter answer such as 'A', 'B', 'C', 'D', and 'E'.
Step 5. Perform the steps 1 to 4 repeatedly up to all N items. The set T will contain the valid interpreted and rectified answers.

Find the important differentiation among automated and manual checking of test papers on efficiency

The Table 5 shows the results that there is important differentiation($t(84) = -2.39$, $p = 0.031$) among automated and manual checking of test papers on efficiency. There is a moderately efficient in automated checking and somewhat

inefficient in manual checking. The automated checking(Mean = 2.65) is faster to check the answer sheet compared to the manual checking(Mean = 3.55). This is medium size effect(Cohen's d = -0.62) on the comparisons.

Table 5 Differentiation among the Automated and Manual Checking

Method	Average	Interpretation	P-Value	Cohen's d
Automated	2.65	Moderately Efficient	0.031	-0.62
Manual	3.55	Somewhat Inefficient		

Find the important differentiation among color of pen with complete shading and color of pen with X symbol answer marks on the precision of OCR

The Table 6 shows the results that there is greatly important differentiation($t(29) = 9.91$, $p = 0.000003852$) among color of pen with complete shading and color of pen with X symbol answer marks on the precision of OCR. The color of pen with complete shading(Mean = 3.52) has better precision compared to color of pen with X symbol(Mean = 2.43). This is large size effect(Cohen's d = 1.81) on the comparisons.

Table 6 Differentiation among the color of pen with complete shading and color of pen with X symbol

Answer Marks	Average	P-Value	Cohen's d
Color of Pen with Complete Shading	3.52	0.000003852	1.81
Color of Pen with X Symbol	2.43		

The Table 7 shows the findings of experiments, the black color is the best color that the OCR detects. The complete shading is better recognized by the OCR than X symbol. The black color of pen with complete shading is the best method of marking the answer.

Table 7 Experiments on methods of marking the answer

Methods of marking the answer	Accuracy of OCR Detection
Black color of pen with Complete Shading	94%
Blue color of pen with Complete Shading	70%
Red color of pen with Complete Shading	50%
Pencil with Complete Shading	5%
Black color of pen with X Mark	70%
Blue color of pen with X Mark	65%
Red color of pen with X Mark	25%
Pencil with X Mark	5%

Automated checking Versus Manual checking

Exam Type: Multiple-Choice Test

Total Test Papers: 85 Total Score: 40 points

Table 8 Results of Automated checking Versus Manual checking

	Manual checking	Automated checking
Finished Time:	3 hrs. 4 min. 9 sec.	2 hrs. 2 min. 41 sec.
Accuracy of Checking:	95%	100%
Average Accuracy of OCR Detection:	n/a	94.47%

The Table 8 shows the results of comparison, the automated checking finished first the checking and quicker than manual checking. The accuracy of checking is when you match the answers to the key answers and computes the total score. The automated checking has 100% accuracy of checking compared to the manual checking. The manual checking has 95% accuracy of checking with 5% error margin. 100% accuracy of checking is attained in manual checking if we check slowly and carefully but it will take more time to finish. The developed software application that automates the checking is imperfect that sometimes it cannot recognize the complete shaded answer. The user fills out the unrecognized answer to compensate.

Results of Software Design

The results of the design, development and coding of software application that automates the checking of multiple-choice test are listed in the following.

1. The OCR is imperfect to detect the answer marks on the answer sheet.
2. The black color is the best color that the OCR detects compared to blue, red and pencil.
3. The complete shading is better recognized by the OCR than X symbol.
4. A large size of font is better recognized by the OCR than small size. The answer sheet prepared has font size of 20 pts.
5. If you rotate the answer sheet anti-clockwise then it decreases the precision of the OCR.
6. If you update and save the answer sheet then it decreases the precision of the OCR.

Benefits and Disadvantages of Using the Developed Software Application

The features of implementing the developed software application are shown on Table 9.

Table 9 Features of Developed Software Application

Features	Disadvantages
1. he software application applied the Optical Character Recognition(OCR) to interpret the complete shaded answers as alternative solution for detecting the answer marks.	1. he software application has imperfection to interpret the complete shaded answers on the answer sheet.
2. t has 94% precision of OCR to detect complete shaded answers.	2. he user requires to fill in the unrecognized answer.
3. t has 100% accuracy of checking by matching the answers with key answers and computing the total score.	
4. he automated checking is quicker and more convenient compared to manual checking.	
5. t has feature that stores and manages the records of students.	
6. he software application can check a minimum of 15 points and maximum of 80 points of either letter or A4 paper size of answer sheets.	

IV. CONCLUSION

The logical and physical layers are the design and development of the software application. The rectified answers detection algorithm is an improvised design solution that compensate the imperfection of the OCR. The algorithm rectifies the misinterpreted character and maps the matching letter as proper answer. There is important differentiation among automated and manual checking of test papers on efficiency. There is greatly important differentiation among color of pen with complete shading and color of pen with X symbol answer marks on the precision of OCR. The black color of pen with complete shading is the best method of marking the answer.

There are great problems encountered by teachers in checking test papers manually such as high volume of exam papers, slow to check, some errors committed in checking and not enough vacant time to check. Thus, teachers should be granted sufficient time to check the test papers. Moreover, teachers who are 55 years old and above or affected by illness or injury should handle not greater than 30 students per class.

Acknowledgments

We are grateful to all teachers at Benguet State University for their sacrifice time and effort in providing us the needed data in our research.

REFERENCES

[1] Ise O. A. A Novel Framework for Student Result Computation as a Cloud Computing Service. American Journal of Systems and Software. January 2015, vol. 3, no.1 p.1, doi:10.12691.
[2] Jacobson Donna. Causes and Effects of Teacher Burnout. Walden University ScholarWorks. 2016.
[3] Limpuasan Arden Joy, Dela Cruz Angelu, Panelo Shiena May. Development of an Automated Exam Checker System with Data Visualization for Sucat Elementary School. 2019.[Online]. Available: https://www.academia.edu/40419048/Automated_exam_checker_system.
[4] BL SOFT Co., Ltd. Automated Scoring System for Exams Greatly Improved Work Efficiency at Schools.2018.[Online].Available:https://panasonic.net/cns/office/products/scanner/case_studies/pdf/CS_ed_BL_SOFT.pdf.
[5] Yerramilli Nithin Sameer, Johnson Nobin Jaision, Omsri Saindha Reddy, Monika Polaki, S. Prajwal. Automatic Exam Answer Checker using Optical Character Recognition and Sentence Embedding. International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON). 2021.
[6] Tanwar Vishwas. Machine Learning based Automatic Answer Checker Imitating Human Way of Answer Checking. International Journal of Engineering Research & Technology (IJERT). vol.10, no.12,p.1, December 2021 issn:2278-0181.
[7] Rosebrock Adrian. Bubble sheet multiple choice scanner and test grader using OMR, Python, and OpenCV-PyImageSearch.2019. [Online]. Available: https://pyimagesearch.com/2016/10/03/bubble-sheet-multiple-choice-scanner-and-test-grader-using-omr-python-and-opencv/detectportal.firefox.com/canonical.html.
[8] Shrestha Ronika, Gupta Raj, Kumari Priya.Automatic AnswerSheet Checker.2022.[Online].Available: https://easychair.org/publications/preprint_open/X18R.
[9] Bansal Vasu, Sharma M.L., Tripathi Krishna Chandra.Automated Answer-Checker. International Journal for Modern Trends in Science and Technology.2020.vol.6, no.12, pp.1-2,doi:10.46501/IJMTST061229.
[10] Zampirolli Francisco de Assis, Quiltici-Gonzalez José Arthur, Oliveira Neves Rogério Perino de. Automatic Correction of Multiple-Choice Tests using Digital Cameras and Image Processing.ResearchGate.April 2013, pp. 5-6.
[11] Tavana Abuzar Mehrabi, Abbasi Mahdi, Yousefi Ali.Optimizing the correction of MCQ test answer sheets using digital image processing. Eight International Conference on Information and Knowledge Technology ResearchGate.September 2016.pp. 1-2.
[12] China Rodrigo Teiske, Zampirolli Franciso de Assis, Quiltici-Gonzalez José Artur, Oliveira Neves Rogério Perino de.An Application for Automatic Multiple-Choice Test Grading on Android. January 2016 Conference Paper ResearchGate.January 2016,pp. 1 & 20.
[13] Cuerdo Ronnel, Ison Michael Jomar, Onate Christian Diols.Effectiveness of Automation in Evaluating Test Results Using EvalBee as an Alternative Optical Mark Recognition(OMR). International Journal of Theory and Application in Elementary and Secondary School Education (IJTAESE).2015.vol. 3, no.2 p.1, doi:10.31098.
[14] Nagy George, Nartker Thomas, Rice Stephen.Optical character recognition: an illustrated guide to the frontier. December 1999 Proceedings of The International Society for Optical Engineering(SPIE). December 1999 pp.1-2,doi:10.1117/12.373511.
[15] Hanmer Robert.Patterm-Oriented Software Architecture. Chichester, West Sussex, England: John Wiley & Sons, Ltd., 2013.

Disclosure

Funding: This research received materials funding from the BSU Office of Research Services.

Software Patronage: One of the main outputs of this research is a software application that checks multiple-choice exam papers using Optical Character Recognition(OCR) technology. The Lacea Quiz Checker software application can be obtained for free at this URL:

https://drive.google.com/file/d/1e37bL0JMTh3QtlUVbG27Kv3EXwD2rSVK/view?usp=sharing.Also, you can watch the demo at this URL: https://youtu.be/JPdiO8_Q3ds

Conflicts of Interest: None.