

Reducing Cheating in Online Exams Through the Proctor Test Model

The Case of Indonesian Learners of Arabic from the Book of Silsilat Al-Lisan

Yusring Sanusi Baso^{1*}, Nurul Murtadho², Syihabuddin³, Hikmah Maulani⁴

Andi Agussalim⁵, Haeruddin⁶, Ahmad Fadlan⁷, Ilham Ramadhan⁸

Department of Arabic Studies-Faculty of Cultural Science, Hasanuddin University, Makassar, Indonesia 90245¹

Department of Arabic Studies, Malang State University, Indonesia²

Department of Arabic Studies, Indonesia University of Education, Indonesia^{3, 4}

Department of Arabic Studies-Faculty of Cultural Science, Hasanuddin University, Makassar, Indonesia 90245^{5, 6, 7, 8}

Abstract—The World Health Organization (WHO) officially declared coronavirus (COVID-19) a pandemic on March 11, 2020. Educational institutions must change most face-to-face learning activities in class to online. This situation forces academic institutions to change the format of assessing student learning outcomes. Online exams surveillance applications utilizing cameras and other blocking browsers (proctors) are becoming popular. However, the appearance of the proctor model supervised exam system also raises controversy. The main discussion regarding this proctor system is the integrity of assessment and the capacity of students to adapt to this new method of supervision. The main question is whether students feel comfortable using the proctor system in exams and whether this system affects students' scores. To answer this question, we have analyzed the scores obtained from a trial of 98 scores of students learning Arabic at Hasanuddin University, Makassar, Indonesia. The experiment involved three exam models: online format from home using the Sikola Learning Management System (Modality 1), online directly using the Proctor System in the Sikola Learning Management System (Modality 2), and a paper exam format in person under the supervision of a lecturer (Modality 3). The results show that students prefer Modality 1 (online at home with the Sikola LMS system). There is a statistical difference between the scores obtained by students from the three modalities analyzed. Student scores with modality 1 are higher than the other two modalities. On the other hand, there was no difference in scores between modalities 2 and 3. The online exam system (modality 2) can be applied to online exams in higher education institutions because it can reduce or even keep students from cheating.

Keywords—Reducing cheating; online exam; proctor test models; Indonesian learners of Arabic; Silsilat Al-Lisan

I. INTRODUCTION

The World Health Organization (WHO) declared the new coronavirus (COVID-19) a pandemic on March 11, 2020 [1] officially. COVID-19 has spread rapidly due to the high transmission capacity of the virus and the routes of transmission (especially via aerosols when coughing and sneezing). In addition, about 30% of patients have various life-threatening symptoms [2]. Higher education institutions must change their learning process activities based on the spread and the symptoms it causes. The face-to-face learning process is suspended. Universities are trying to adapt to this situation which has become a challenge worldwide. At the study program

level, the teaching team must adjust their teaching and assessment systems [3]. On the other hand, lecturers and students have received online learning activities well, both in learning activities and in the assessment system.

An online assessment or exam system with strict monitoring is starting to be widely used. This proctor system has generated significant controversy, especially in science education [4]–[7]. Among the concerns of educators and students are psychological disturbances, privacy, and various environmental factors [5]. Dragan et al. (2020) concluded that online exams with remote supervision proved to be a timely solution, and also, the emotional needs of students who might feel stressed by these rigorous proctor exams should be considered [8].

Two of the most significant challenges in the online exam proctor system are the assessment standards and students' ability to adapt to this new exam modality. Several studies have been carried out to analyze student dishonesty in the online exam proctor system and various ways to prevent it. A study by Guangul et al. (2020) concluded that combining various assessment methods has helped minimize academic dishonesty [4]. Li (2021) developed an anti-collusion approach based on optimizing remote online testing [8]. Recently, Pettit et al. (2021) have analyzed studies conducted so far that provide recommendations for improving student authentication and preventing fraud [7]. Baso [2022] also wrote an online exam model with a reliable proctor system to reduce academic cheating that might occur [9].

The integrity of the evaluation in online exams can be monitored in various ways, including taking the exam in person or using a real-time supervisor system [5]. One of the real-time online monitoring systems is the website Uji.sikoola.com which provides real-time online monitoring services using a webcam and a browser lock [10]. Students are connected to the online exam page sikoola.ujian.com which monitors students during the online exam. Before the exam starts, students must check their internet speed. The sikoola.ujian.com system will analyze the internet speed used by students in real-time and provide good, moderate, or bad status for student internet speed. Furthermore, students must check the webcam device they are using. If the webcam is not working, students cannot enter the online exam waiting room.

*Corresponding Author.

The main objective of this study was to compare the attitudes and scores obtained by students in the Arabic study program who took the Arabic language skills exam learned from the Silsilat al-Lisan book [11] through the online exam proctor system on the sikoola.ujian.com page.

In relation to the purpose of this research, the questions to be answered are:

- What is the students' attitude towards the three modalities of the exam in measuring the Arabic language skills they have learned from the book Silsilat al-Lisan?
- Are there differences in the scores obtained by students from the three exam modalities given?

II. MATERIALS AND METHODS

A. Population and Sample

The research was conducted in the Arabic study program at Hasanuddin University, in two batches of students, namely the 2022 and 2021 batches (ages between 20-21 years). The 2022 batch consists of 52 students, and the 2021 batch of 46 students. This group of students sequentially took Arabic 1 and Arabic 3 courses. The questions tested were sourced from the book Silsilat al-Lisan.

The exam is carried out three times at an interval of two weeks. The score range of each test is 0-100. The model of the three exam questions is the Multiple-Choice Question (MCQ). The first exam was conducted online via <https://sikola.unhas.ac.id>. Students take exams simultaneously where they can access this page from anywhere (modality 1). The second exam is also conducted online on the <https://sikoola.ujian.com> page which has a strict monitoring feature (modality 2). The third exam was conducted in class with supervision from a team of lecturers supporting Arabic 1 and Arabic 2 courses (modality 3).

B. Study Design

Students must take all three exams (modalities 1, 2, and 3). The reason is that students can feel the difference when taking exams with different modalities. In the modality one exam, they take the exam from anywhere. It's just that the time for carrying out the exam has been determined when it starts to be accessed and when the questions will be closed automatically. In other words, the first modality test is carried out simultaneously at the same time.

Examination with modality two is also carried out simultaneously at the same time. Exam venues may vary. It's just that there are mandatory requirements that must be met, namely, the device used must have a functioning webcam. In addition, the internet speed used is at least 40 Mbps upload. This speed would be considered moderate. If the internet speed test result equals or exceeds 50 Mbps, the system will categorize it as good. However, if the internet speed test results are below 40 Mbps, the sikoola.ujian.com system will assess the network used by this student as bad. Bad speed test results will prevent students from accessing exam questions on the sikoola.ujian.com page. If this condition occurs, students must

move locations to find an access point with a minimum internet speed test of 40 Mbps.

In addition, in this online exam with modality 2, the exam supervisor can observe the behavior of the examinees. The faces of each examinee will be displayed on the monitor screen on the sikoola.ujian.com page. If there are examinees who move a lot by looking left and right, the exam supervisor can send a warning message to these students. In addition, if an examinee opens a new page other than the exam page from the browser used, the sikoola.ujian.com application will block it. Thus, the examinee's monitor screen will be locked. Likewise, if the examinee opens another browser, the screen will be locked again. In the same way, if the examinee presses a key on the keyboard, the monitor screen will also be locked. Lockdown time can be set as needed, for example, 20 seconds or 50 seconds, or even an hour.

Examination with modality three is conducted in class. Examinees receive test papers. The exam supervisor will supervise them during the test. After students took the three MCQ tests with three modalities, they filled out a survey that included questions about their attitudes toward the three test modalities.

C. Statistical Analysis

The research team used the statistical software SPSS 26 to analyze the data. The research team conducted a normality test on standardized residual values to calculate whether the three test score variables were normal or not. If the results are normal, then the repeated measure ANOVA test is used, and vice versa. If it is not normal, the researcher will use the Friedman (non-parametric statistics) test to analyze the data.

The basis for the decision of the normality test used in this study is:

- If the Sig value > 0.05, then the data is normally distributed.
- If the Sig value < 0.05, the data are not normally distributed.

The interpretation is as follows:

- The Within-Subjects Factors output table shows three score variables from three exams (modality 1, modality 2, and modality 3).
- The output table of Mauchly's Test of Sphericity is used to see the similarity of the assumption of variance (Sphericity Assumed) of the research data provided that the Sig value is <0.05. If these conditions are not met, the researcher will use the Greenhouse-Geisser as a condition for testing the hypothesis.
- The Tests of Within-Subjects Effects output table will be used by researchers to:
 - 1) Make a hypothesis formulation.
 - 2) Know the basis for decision making.
 - 3) Conclusion.

- The formulation of the hypothesis is:
 - H0: there is no difference in the average score of the three test modalities given to students
 - Ha: there is a difference in the average score of the three test modalities given to students.
- The basis for the decision to be used is:

1) If the value of Greenhouse-Geisser Sig > 0.05, then H0 is accepted and Ha is rejected.

2) If the Greenhouse-Geisser Sig value < 0.05, then H0 is rejected and Ha is accepted.

III. RESULTS

A descriptive analysis of this study can be seen in Table I:

TABLE I. DESCRIPTIVE

			Statistic	Std. Error
Standardized Residual for MODALITY1	Mean		0,0000	0,10102
	95% Confidence Interval for Mean	Lower Bound	-0,2005	
		Upper Bound	0,2005	
	5% Trimmed Mean		-0,0022	
	Median		-0,2304	
	Variance		1,000	
	Std. Deviation		1,00000	
	Minimum		-1,61	
	Maximum		1,70	
	Range		3,30	
	Interquartile Range		1,65	
	Skewness		-0,019	0,244
	Kurtosis		-1,291	0,483
Standardized Residual for MODALITY2	Mean		0,0000	0,10102
	95% Confidence Interval for Mean	Lower Bound	-0,2005	
		Upper Bound	0,2005	
	5% Trimmed Mean		-0,0007	
	Median		0,0324	
	Variance		1,000	
	Std. Deviation		1,00000	
	Minimum		-1,69	
	Maximum		1,75	
	Range		3,44	
	Interquartile Range		1,72	
	Skewness		-0,034	0,244
	Kurtosis		-1,186	0,483
Standardized Residual for MODALITY3	Mean		0,0000	0,10102
	95% Confidence Interval for Mean	Lower Bound	-0,2005	
		Upper Bound	0,2005	

	5% Trimmed Mean	-0,0058	
	Median	-0,1034	
	Variance	1,000	
	Std. Deviation	1,00000	
	Minimum	-1,53	
	Maximum	1,68	
	Range	3,22	
	Interquartile Range	1,67	
	Skewness	0,196	0,244
	Kurtosis	-1,178	0,483

A. Research Question One

The data is obtained in the following Table II to answer the first question of this study about students' attitudes towards the results of the MCQ exam scores with three modalities.

TABLE II. ATTITUDES TOWARDS MODALITIES

NUMBER OF STUDENTS	ATTITUDE TOWARDS MODALITIES					
	MODALITY 1		MODALITY 2		MODALITY 3	
	Trust	Untrust	Trust	Untrust	Trust	Untrust
Batch 2021	16	30	46	0	46	0
Batch 2022	33	35	52	0	52	0

B. Research Question Two

Table III shows three variables to be tested, namely, MCQ score modality 1, MCQ score modality 2, and MCQ score modality 3, as shown in Table IV.

As for Sphericity, Assumed with the condition that Sig. < 0.05 and then in the following Table V, it is obtained Sig 0.010 > 0.05. Thus, Sphericity Assumed is not fulfilled. Therefore, Greenhouse-Greiser will be used as the basis for testing the hypothesis.

TABLE III. TESTS OF NORMALITY

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual for MODALITY1	0,173	98	0,000	0,934	98	0,000
Standardized Residual for MODALITY2	0,105	98	0,010	0,951	98	0,001
Standardized Residual for MODALITY3	0,092	98	0,039	0,937	98	0,000
a. Lilliefors Significance Correction						

TABLE IV. WITHIN-SUBJECTS FACTORS

Measure:	MCQSCORE
MODALITY	Dependent Variable
1	MODALITY1
2	MODALITY2
3	MODALITY3

TABLE V. MAUCHLY'S TEST OF SPHERICITY

Mauchly's Test of Sphericity ^a							
Measure:	MCQSCORE						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
MODALITY	0,908	9,283	2	0,010	0,916	0,932	0,500
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.							
a. Design: Intercept Within Subjects Design: MODALITY							

To get conclusions from this statistical data, it can be seen in Table VI. This output table, known as Tests of Within-Subjects Effects, will display the Greenhouse-Geisser as can be seen in Table V below:

TABLE VI. TESTS OF WITHIN-SUBJECTS EFFECTS

Measure:	MCQSCORE					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
MODALITY	Sphericity Assumed	17880,660	2	8940,330	193,272	0,000
	Greenhouse-Geisser	17880,660	1,831	9764,337	193,272	0,000
	Huynh-Feldt	17880,660	1,865	9589,111	193,272	0,000
	Lower-bound	17880,660	1,000	17880,660	193,272	0,000
Error(MODALITY)	Sphericity Assumed	8974,007	194	46,258		
	Greenhouse-Geisser	8974,007	177,628	50,521		
	Huynh-Feldt	8974,007	180,874	49,615		
	Lower-bound	8974,007	97,000	92,516		

Based on the hypothesis with the Greenhouse-Geisser Sig. < 0.05, then $0.000 < 0.05$ is obtained, meaning H_0 is rejected, and H_a is accepted, or there is a difference in the average score of the three test modalities given to students.

IV. DISCUSSION

The COVID-19 pandemic has forced lecturers and students to change academic activities, including activities to measure student learning outcomes. Several research results show acceptance of the learning achievement measurement method with proctor applications [12]–[14], including in Indonesia [10]. In this study, to answer the first research question, the results obtained can be seen in Table I previously, where it can be seen that students believe in the MCQ exam model

modalities 2 and 3. However, in the 2021 and 2022 batches, some students do not believe in modalities 1.

The research team conducted interviews to explore the causes of the distrust of some students towards the MCQ exam model with modality 1. In general, it can be concluded that the cause of this distrust is the opportunity for students to commit fraud during the exam. Opportunities for students to commit fraud include searching for answers via Google (Googling), asking other friends via mobile (chatting via WhatsApp, Telegram, etc.) during exams, or opening books. Opportunities for fraud are open because there is no direct supervision. However, some students still have a reliable academic attitude by not cheating during exams.

On the other hand, the research team found that the basis for the 100% confidence of the two groups of students in the MCQ modality 2 and 3 exam models was the feelings of students monitored both offline and online. This strict supervision makes students not think of committing fraud. This condition instead makes students focus on answering exam questions.

Table I above (descriptive statistics) shows that the test scores in modality 1 are higher than in modalities 2 and 3. This data can be accepted logically because it is possible that during the exam, some students cheated. They are looking for answers to questions they have difficulty answering via googling or asking other friends by chatting. On the other hand, the test scores on modalities 2 and 3 are lower because, logically, it can also be understood that there are no opportunities for students to commit fraud. Examination modalities 2 and 3 close opportunities for cheating during exams, especially for students who have cheated on exams with modality 1.

The assumptions of the research team are based on Fig. 1:

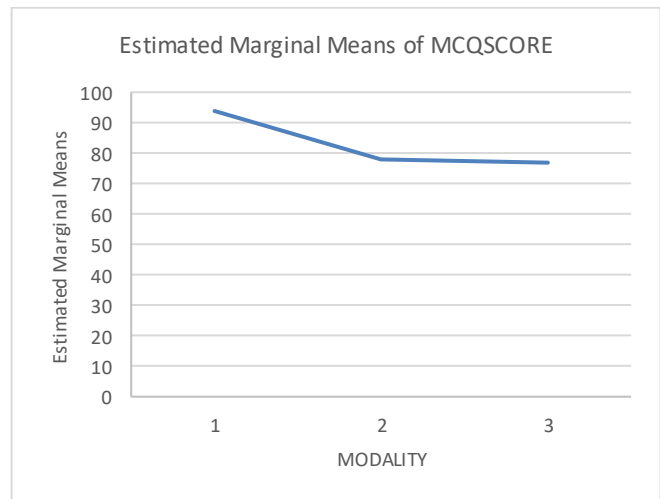


Fig. 1. The test scores with modality.

Fig. 1 shows that the results of the test scores with modality 1 are higher than modalities 2 and 3. However, the results of the test scores for modalities 2 and 3 do not appear to have a significant difference. On the other hand, modality 2 and modality 3 have different test channels; modality 2 is carried out online while modality 3 is carried out offline.

Logically the score results are relatively the same or not significantly different even though the exam channels are different because exam students are under strict supervision. In modality 2, supervision uses the proctor application system. As for modality 3, supervision is carried out directly by F2C in the exam room.

Thus it can be said that modalities 2 and 3 have in common close supervision during the exam. This strict supervision puts the exam atmosphere in a safe condition where students think they have no opportunity to cheat in the exam.

V. CONCLUSIONS

In this study, it was found that the attitude of students lacked confidence in the score of the exam results with modality 1. The cause of this distrust was caused by the opportunity for students to commit fraud during the exam. In contrast, students believe in the score of the exam results with modalities 2 and 3. Student trust in the score of the exam results with modalities 2 and 3 is due to the strict supervision conditions during the exam even though the channels of modality 2 and 3 are different, where modality 2 is online channel whereas modality 3 is offline channel.

With these results, it can be claimed that measuring learning outcomes through MCQ with modalities 2 and 3 is reliable. If the exam is conducted online, modality 2 can be used as a solution; if offline, modality 3 is used as an alternative.

ACKNOWLEDGMENTS

We are grateful to the all students of Arabic Study program of Universitas Hasanuddin who involved in this research.

AUTHOR CONTRIBUTIONS

Conceptualization, Yusring Sanusi Baso; methodology, Yusring Sanusi Baso, Hikmah Maulani, dan Andi Agussalim; formal analysis, Yusring Sanusi Baso and Haeruddin; investigation, Fadlan Ahmad; data curation, Ilham Ramadhan; writing-original draft preparation, Yusring Sanusi Baso; writing-review and editing, Syihabuddin and Nurul Murtadho. All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," *Acta Biomed.*, vol. 91, no. 1, pp. 157–160, 2020.
- [2] P. J. Marín García, A. Amau, and L. Llobat, "Preferences and scores of different types of exams during COVID-19 pandemic in faculty of veterinary medicine in Spain: A cross-sectional study of paper and E-exams," *Educ. Sci.*, vol. 11, no. 8, pp. 0–5, 2021.
- [3] A. S. Milone, A. M. Cortese, R. L. Balestrieri, and A. L. Pittenger, "The impact of proctored online exams on the educational experience," *Curr. Pharm. Teach. Learn.*, vol. 9, no. 1, pp. 108–114, 2017.
- [4] F. M. Guangul, A. H. Suhail, M. I. Khalit, and B. A. Khidhir, "Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College," *Educ. Assessment, Eval. Account.*, vol. 32, no. 4, pp. 519–535, 2020.
- [5] F. F. Kharbat and A. S. Abu Daabes, "E-proctored exams during the COVID-19 pandemic: A close understanding," *Educ. Inf. Technol.*, vol. 26, no. 6, pp. 6589–6605, 2021.
- [6] P. R. Lockman, J. A. Gaasch, K. Borges, A. Ehlo, and Q. R. Smith, "Using WebCt to implement a basic science competency education course," *Am. J. Pharm. Educ.*, vol. 72, no. 2, p. 39, 2008.
- [7] M. Pettit, S. Shukla, J. Zhang, K. H. Sunil Kumar, and V. Khanduja, "Virtual exams: has COVID-19 provided the impetus to change assessment methods in medicine?," *Bone Jt. Open*, vol. 2, no. 2, pp. 111–118, 2021.
- [8] I. F. Dragan, L. F. Yildiz, K. Dunn, and A. Ramesh, "Integrating remote proctoring in dental education: Problem, solution, and results," *J. Dent. Educ.*, vol. 85, no. S1, pp. 1071–1073, 2021.
- [9] M. Li, "Optimized collusion prevention for online exams during social distancing," *npj Sci. Learn.*, vol. 6, no. 1, 2021.
- [10] Y. S. Baso, "Proctoring and Non-proctoring Systems: A Comparative Study of Online Exam Scores for an Arabic Translating Course," vol. 13, no. 6, pp. 75–82, 2022.
- [11] D. M. A. and M. B. Dr. Bakri Sheigh Ameen, Ali Hamadallah, Dr. Fakhruddin Qabawah, Dr. Mazen Almubarak, "Kitab Silsilat Al-Lisan," Abu Dhabi, 2021. [Online]. Available: <https://mothertongue.ae/arabic-curricula/silsilat-al-lisan/>. [Accessed: 06-Jun-2022].
- [12] M. S. Medina and A. N. Castleberry, "Proctoring strategies for computer-based and paperbased tests," *Am. J. Heal. Pharm.*, vol. 73, no. 5, pp. 274–277, 2016.
- [13] H. Meishar-Tal and A. Levenberg, "In times of trouble: Higher education lecturers' emotional reaction to online instruction during COVID-19 outbreak," *Educ. Inf. Technol.*, vol. 26, no. 6, pp. 7145–7161, 2021.
- [14] A. Sangwan, A. Sangwan, and P. Punia, "Development and Validation of an Attitude Scale towards Online Teaching and Learning for Higher Education Teachers," *TechTrends*, vol. 65, no. 2, pp. 187–195, 2021.