

APPLICATION OF PERFORMANCE ASSESSMENT WITH METHOD DIFFICULTY LEVEL, DISTINGUISHING POWER, DISTRACTOR EFFECTIVENESS, AND ANOVA (Case Study: Paper Based Test (PBT) Period 2 Questions 2020–2021 Islamic University of Indonesia)

Safira Khoirunnisa^{1*}, Edy Widodo²

^{1,2} Statistics Study Program, Faculty of Mathematics and Natural Sciences, Islamic University of Indonesia
Kaliurang St., KM 14.5, Sleman, Yogyakarta, 55584, Indonesia

Corresponding author's e-mail: ^{1*} safirakhoirunnisa16@gmail.com

Abstract. The quality of human resources can be improved through education. The Islamic University of Indonesia provides various selection patterns for prospective new students, one of which is PBT. Questions in PBT need to be analyzed to get quality students. The length of time to analyze the performance of questions that include the level of difficulty, distinguishing power, and distractor effectiveness in PBT Period 2 2020-2021 UII is one of the reasons for making computer programs using Microsoft Excel 2013. The methods used are simple random sampling and One-Way ANOVA analysis. The results of this study can make it easier to process data and analyze the performance of the questions with the help of the OASIS application which can be used as material for future evaluation. Analysis of the performance of the questions is classified as very difficult and difficult at the level of difficulty, the category is bad on the distinguishing power, and the effectiveness of the distractor is functioning properly, and there is one of the scores obtained by the test takers from three types of questions with the choice of the Faculty of Medicine and not choosing the Faculty of Medicine of UII which has the average is different from the others.

Keywords: difficulty level, distinguishing power, distractor effectiveness, Microsoft Excel 2013, one-way ANOVA.

Article info:

Submitted: 16th February 2022

Accepted: 25th April 2022

How to cite this article:

S. Khoirunnisa, and E. Widodo, "APPLICATION OF PERFORMANCE ASSESSMENT WITH METHOD DIFFICULTY LEVEL, DISTINGUISHING POWER, DISTRACTOR EFFECTIVENESS, AND ANOVA (CASE STUDY: PAPER BASED TEST (PBT) PERIOD 2 QUESTIONS 2020–2021 ISLAMIC UNIVERSITY OF INDONESIA)", *BAREKENG: J. Il. Mat. & Ter.*, vol. 16, iss. 2, pp. 539-550, June, 2022.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).
Copyright © 2022 Safira Khoirunnisa, Edy Widodo.

1. INTRODUCTION

Human resources are very important in building a nation. Many aspects can be used to improve the quality of human resources. According to [40] these aspects include aspects of attitude, mental, behavior, ability, intelligence, religion, law, health, and so on. Realizing the ideals of the Indonesian nation, namely educating the nation's life which is written in the opening text of the 1945 Constitution, and being able to improve the quality of human resources can be achieved through education [13],[38]. According to [14] humans need the education to be able to manage the nature provided by Allah SWT with the knowledge they have. In Article 5 Paragraph 1 of the Republic of Indonesia Law No. 20 of 2003 chapter four, the first part on the Rights and Obligations of Citizens regarding the National Education System that "Every citizen has the same right to obtain a quality education"[32]. In Indonesia, there are public universities and private universities. One of the private universities in Indonesia is the Indonesian Islamic University. From the various PMB selection patterns carried out to register as UII students, it is necessary to evaluate to determine the success or failure of the test process [2] which is carried out to select the right students according to the chosen study program.

One of the PMB selection patterns offered to get quality students is the PBT selection pattern [33]. This item analysis aims to determine whether the items are good, need to be revised, or some items need to be discarded. In PBT, the questions selected are in the form of multiple choices which provide alternative answers so that the test takers only choose one alternative that is considered correct. According to [11] 3 important things in conducting item analysis are the level of difficulty, distinguishing power, and the effectiveness of the distractors for each question.

The length of time used to conduct item analysis is one of the backgrounds for making computer programs for item analysis. Science and technology that are increasingly developing can be utilized to help perform item analysis easily and quickly. According to [24] in Legal News that many application programs for item analysis. However, some of these applications are in foreign languages, complicated to use, difficult to understand, less practical and applicable, and have a variety of information. Meanwhile, applications are using Indonesian, namely the AnBuSo application created by Ali Muhson using Microsoft Excel [39] and the ANATES application developed by Drs. Karno, M.Pd. and Yusuf Wibisono, ST. [42]. Researchers are also interested in making a performance analysis program for multiple choice questions by optimizing the functions of Microsoft Excel 2013, a case study of PBT Period 2 for 2020–2021 at the Islamic University of Indonesia with an attractive appearance, easy to understand, and providing data visualization in the form of graphs that are summarized in the dashboard so that it can make decisions quickly and more effectively to use. After that, further analysis was carried out to determine the difference in the average scores of the test takers from the sample used using the Rstudio software.

Based on the problems above, the objectives to be achieved in this study are as follows: (1) Obtaining an application with optimized Microsoft Excel 2013 functions in helping to analyze the performance of multiple choice questions on university entrance exams. (2) Knowing the performance of the Paper Based Test (PBT) for Period 2 of 2020-2021 at the Islamic University of Indonesia when viewed from the Level of Difficulty, Distinguishing Power, and Effectiveness of Distractors. (3) Knowing the difference in the average scores obtained by test takers from the three types of Paper Based Test (PBT) questions for the period 2 years 2020–2021 on test-takers with the choice of the Faculty of Medicine and not choosing the Faculty of Medicine, Islamic University of Indonesia.

2. RESEARCH METHODS

The population in this study were all test participants for prospective new students at the Islamic University of Indonesia through the Paper Based Test. The sample in this study was the test participants for prospective new students at the Islamic University of Indonesia through the Paper Based Test (PBT) period 2 in 2020-2021. The data used in the research is secondary data regarding the answers of the participants of the Paper Based Test period 2 of the Islamic University of Indonesia in 2020–2021, both participants of the Paper Based Test by choosing the Faculty of Medicine and those who did not choose the Faculty of Medicine, along with the answer key for each multiple choice item, with The total data used was 472 PBT participants for the period 2 years 2020-2021, Islamic University of Indonesia. The variables used include the level of difficulty, discriminatory power, the effectiveness of distractors, grades, and types of questions.

The research sample was taken using simple random sampling or a simple random technique as many as 249 Paper Based Test participants with the choice of the Faculty of Medicine and 223 Paper Based Test participants who did not choose the Faculty of Medicine.

2.1. Research Stages

The stages of making the application carried out by researchers are as follows:

1. Analysis of application requirements or the initiation stage is the stage where the needs made can be understood by users to achieve the desired goals.
2. Application design is designed based on the results of the stages required to achieve the objectives. If the design has been formed, the researcher will make an application.
3. Making applications which include: coding for application logins with excel macros and making formulas according to the formulas to analyze the performance of questions.
4. The application trial or testing phase is carried out to determine whether the application can be used properly or not. If it cannot be used properly, it is necessary to check the application creation and make improvements to the application.
5. If the application trial is successful, it can collect and identify the data needed to perform a test performance analysis.
6. Login with username and password.
7. Input data in the application.
8. The final score is obtained in the form of the total value of each participant and the results of the performance analysis of multiple choice questions in the form of the level of difficulty, distinguishing power, and distractor effectiveness from the participants of the Paper Based Test period 2 years 2020-2021 Indonesian Islamic University and displayed in the dashboard.
9. Perform simple descriptive statistics to provide an overview of the participants of the Paper Based Test for the period 2 years 2020–2021 Indonesian Islamic University.
10. Conduct further analysis with R on One-Way ANOVA to find out the difference in the average scores of the three types of questions in the choice of the Faculty of Medicine or not choosing the Faculty of Medicine.
11. Checking assumptions can be investigated through residual examination including independence with the Durbin-Watson test, normality with the Kolmogorov-Smirnov test, and homogeneity with the Bartlett test.
12. If the independent assumption is not met, you can return to point 5, if normality is not met, data transformation can be carried out, and if the data is not homogeneous, you can use an alternative test, namely the Welch test.
13. Make conclusions on the results of the analysis obtained, so that they can provide information about the final score of each participant and the results of the performance analysis of multiple choice questions in the form of difficulty level, distinguishing power, and distractor effectiveness. In addition, you can find out the difference in the average scores of participants for the Paper Based Test Period 2 for 2020-2021 at the Islamic University of Indonesia and as an evaluation material for the following year.

2.2. Application

The software can be categorized into two parts, namely application programs, and operating systems. According to [12] the application is a computer program to process input into output which is composed of several instructions that are run in a computer.

2.3. Microsoft Office Excel

Microsoft Excel is a spreadsheet program created and distributed by Microsoft Corporation that can run on Microsoft Windows and Mac OS. Microsoft Excel has many features in mathematical operations, solving mathematical models, processing data, visualizing data, and others [15].

2.4. Excel Macros

According to [43] macro excel is a visual basic program development in Microsoft Excel. Excel macros can be active if the excel file is saved in xlsx format, if it is saved in xlsm format, the excel macro does not work. There are many advantages of using excel macros, namely completing work quickly, the process is automatic so that it saves more energy, and reduces errors in completing work.

2.5. Evaluation

Evaluation includes two activities, namely measuring and assessing. According to [6] measuring is an activity to compare something with a size or is quantitative. While assessing is an activity to decide something with a qualitative assessment measure or a good or bad measure. Evaluation activities must be carried out regularly and continuously covering the process and the results obtained. According to Nurawati [30] evaluation has several objectives, including achieving the initial goal, knowing what test takers know, motivating learning, being able to find out the results of learning scores, making evaluation results for improvement in the next question.

2.6. Item Analysis

Item analysis is a process of collecting, summarizing information derived from answers to make a decision [35] by identifying the questions and obtaining information to improve the quality of the questions and as a guide in making further questions [31]. [23] describes that in item analysis there are two measurement theories, firstly the classical measurement theory (Classical Measurement Theory) which includes the level of difficulty, discriminatory power, and distractor effectiveness, the second item is response theory (Item Response Theory).

Difficulty level describes how easy or difficult the item is given. According to [34] the level of difficulty is a comparison of the number of test participants able to answer the questions correctly with the number of respondents. The more participants who answer correctly, the item has a low level of difficulty. In compiling the items, the difficulty must be balanced, namely 25% difficult, 50% moderate, and 25% easy [37]. The difficulty index is symbolized by P which means proportion, and the formula used is written in equation 1 as follows [36]:

$$P = \frac{B}{JS} \quad (1)$$

The index of difficulty level can be seen in Table 1 as follows:

Table 1. Difficulty Index

P	Category	Decision
0.00	Very Difficult	Rejected/Revised
0.00 – 0.30	Hard	Rejected/Revised
0.30 – 0.70	Currently	Accepted
0.70 – 1.00	Easy	Rejected/Revised
1.00	Very Easy	Rejected/Revised

Distinguishing Power explains the difference between the upper and lower groups or an index that shows how much power the items have to distinguish the abilities of prospective new students. The higher the index on the item, the more capable the item is to distinguish the test takers who are in the upper and lower groups [41]. According to [28] large groups (more than 30) and small groups (less than 30). The large group can be distinguished by the way the large group is taken by both poles, namely 27% of the highest score as the upper group and 27% of the lowest score as the lower group. As for the small group, the whole group is divided into two so that 50% of the upper group and 50% of the lower group are obtained. The formula used to find the discrimination index in equation 2 is as follows [36]:

$$D = P_A - P_B = \frac{B_A}{J_A} - \frac{B_B}{J_B} \quad (2)$$

The Differentiating Power Index according to [26] can be seen in Table 2 as follows:

Table 2. Differing Power Index

D	Category	Decision
0.00 – 0.20	Bad	Rejected/Revised
0.21 – 0.40	Enough	Rejected/Revised
0.41 – 0.70	Well	Accepted
0.71 – 1.00	Very Well	Accepted
$D \leq 0.00$	Very Ugly	Rejected

Distractor effectiveness is the ability to outwit questions. Good items, distractors will be chosen evenly by students who answer incorrectly. On the other hand, items that are not good at distracting will be chosen unequally. According to [4] the formula used is written in equation 3 as follows:

$$\text{Distractor Effectiveness} = \frac{\text{The number of test-takers who chose the option}}{\text{Total number of test-takers}} \quad (3)$$

The classification of distractor effectiveness according to [36] can be seen in Table 3 as follows:

Table 3. Distractor Effectiveness Classification

Turnouts	Decision
$\geq 5\%$	Option Accepted
$< 5\%$	Need to be revised
0%	Option Denied

2.7. Descriptive Statistics

Informative, concise, and clear visualization to understand the results of the research can be displayed in the dashboard. One of the tools used in creating dashboards is Microsoft Excel. Dashboards present important information with the help of graphs [3]. Graphs that are commonly used to visualize data are bar charts and pie charts.

2.8. One-Way ANOVA

ANOVA was developed by R.A Fisher [18]. The method to check whether there is a relationship between two or more groups of data is called ANOVA or Analysis of Variance. For example, in a group of k , taken from a population with an average of $\mu_1, \mu_2, \dots, \mu_k$, then the test hypothesis is obtained as follows:

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$ (there is no effect of the independent variable on the dependent variable. That the mean in the first population is the same as the mean in the second population, and so on from the same population.)

$H_1: \exists \mu_i \neq \mu_j$ (there is at least 1 value pair ($i: j$). In other words, if there is one mean in the population that is different from the others, then the sample comes from a different population.)

By using the decision that $F_{count} > F_{table}$, then reject H_0 . The following Table 4 describes the One-Way ANOVA table according to [17]:

Table 4. One-Way ANOVA

Source of Variance	Sum of Squares	Degrees of freedom	Mean Square	F_{count}	F_{table}
Treatment	JKP	$k - 1$	$RKP = \frac{JKP}{k - 1}$	$\frac{RKP}{RKG}$	$F_{(k-1),(n-k);\alpha}$
Error	JKG	$n - k$	$RKG = \frac{JKG}{n - k}$		
Total	JKT	$n - 1$			

Where Equations 4 to 7 to solve One-Way ANOVA can be calculated by the following equation.

$$\text{Correction Factor} = \frac{y_{..}^2}{\sum_{i=1}^t r_i} \quad (4)$$

$$JKT = \sum_{i=1}^t \sum_{j=1}^r (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - FK \quad (5)$$

$$JKP = \sum_{i=1}^t \sum_{j=1}^r (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^t \frac{y_{i.}^2}{r_i} - FK \quad (6)$$

$$JKG = \sum_{i=1}^t \sum_{j=1}^r (y_{ij} - \bar{y}_{i.})^2 = JKT - JKP \quad (7)$$

Calculations performed for $t =$ the treatment $-i$ and $r =$ the test $-j$ can be done by observing the tabulation of the data in Table 5.

Table 5. Data Tabulation

Test (Name of Participant)	Treatment (Type of Question)			Total
	1	2	3	
1	y_{11}	y_{12}	y_{13}	
⋮	⋮	⋮	⋮	
a	y_{a1}	y_{a2}	y_{a3}	
Total Treatment	$y_{1.}$	$y_{2.}$	$y_{3.}$	$y_{..}$
Average	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$	$\bar{y}_{..}$

Source: [10]

If the decision obtained is to reject H_0 , then there is an average difference, it can be known by posthoc test. According to [25] a multiple comparison test that can be used if the number of samples for each group is different is using the Scheffe test. The hypothesis used is as follows [1]:

$H_0 : \mu_i = \mu_j$ (There is no significant difference between variables)

$H_1 : \mu_i \neq \mu_j$ (There is a significant difference between variables)

With $i, j = 1, 2, \dots, t$ dan $i \neq j$

The F test statistics used are as follows in equation 8 [7]:

$$F_{i-j} = \frac{(\bar{X}_i - \bar{X}_j)^2}{RKG \left[\frac{1}{n_i} + \frac{1}{n_j} \right]} \quad (8)$$

With a critical region, $DK = \{F | F > (k - 1)F_{\alpha; k-1, N-k}\}$

According to [8] in conducting the ANOVA test, it is necessary to fulfill the assumption that the residuals are normally distributed, each observation is independent of each other (independent), and has a homogeneous variance.

1. Independence

According to [29] independent or independent if the sample is taken at random. According to [5] the Durbin-Watson test can be carried out to determine whether the data is random or not. The test hypothesis used is

$H_0 =$ Residual random data

$H_1 =$ Residual data is not random

If $DW < dL$, reject H_0 , with the test statistic in equation 9 [21]:

$$DW = \frac{\sum_{i=2}^N (e_i - e_{i-1})^2}{\sum_{i=2}^N e_i^2} \quad (9)$$

2. Normality

If the data in the sample is normally distributed, it can be said that the population is normally distributed. If the data is not normal, the sample can be generalized to the population [20]. According to [20] both methods to determine whether a sample is normal can be used, if the number of samples is equal to or more than 50, then use the Kolmogorov-Smirnov method, if the number of samples is less

than 50 using the Shapiro Wilk method. The Kolmogorov-Smirnov test formula is obtained which is written in equation 10 as follows [19]:

$$D = \max |F_n(x_i) - F_0(x_i)| \quad i = 1, 2, \dots, n \quad (10)$$

Hypothesis Testing

H_0 = Residual data is normally distributed

H_1 = Residual data is not normally distributed

3. Homogeneity

According to [16] the homogeneity test is a test to determine the value of the same variance in the sample data or to determine the characteristics of the same sample data or not consisting of two or more data groups. The Bartlett test is used with the number of each group member being equal or not the same and according to [22] the Bartlett test is used if the data is normally distributed, with the hypothesis:

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ (Residual variance of each homogeneous group)

$H_1: \sigma_i^2 \neq \sigma_j^2$ (The residual variance of each group is not homogeneous)

Determine the Bartlett test, using the formula in equation 11:

$$B = (\log S^2) \sum (n_i - 1) \quad (11)$$

According to [27] if the data is not homogeneous, then an alternative test for the homogeneity test on One-Way ANOVA with the Welch test is carried out. The resulting decision is $F_{count} > F_{table}$, then reject H_0 or the variance is not homogeneous, using the formula written in equation 12 [9]:

$$F = \frac{\frac{1}{k-1} \sum_{j=1}^k w_j (\bar{x}_j - \bar{x}')^2}{1 + \frac{2(k-2)}{k^2-1} \sum_{j=1}^k \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{w} \right)^2} \quad (12)$$

3. RESULTS AND DISCUSSION

3.1. Multiple Choice Question Performance Application Results

The design is designed to make it easier for users to analyze the performance of questions with the type of questions in the form of multiple choice designed with excel macros and several formulas. This application can be used with the help of Microsoft Excel. An overview of this application is as follows:

1. Login menu designed using excel macros functions to enter the application. This menu consists of a username, password, login button, a message field if the username or password is incorrect, and a column to create a new account by contacting the admin. The login menu is displayed in Figure 1.

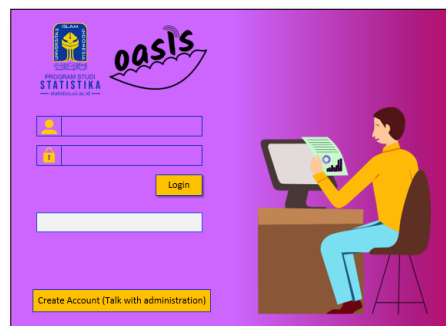


Figure 1. Initial View Login Menu

- Home is used to select the menu used. Consists of a dashboard menu, question identity, input data, reports, analysis, and final data, with the displayed in Figure 2.

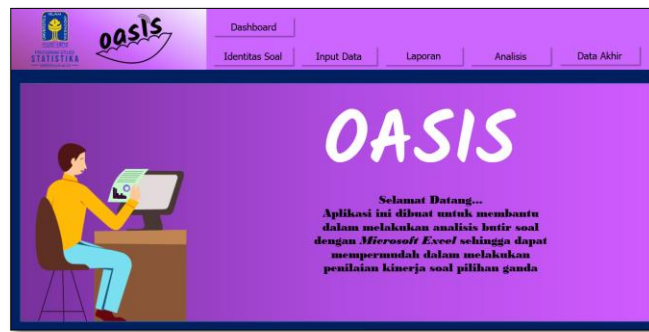


Figure 2. Home Menu

- The dashboard menu is used to display data visualization regarding the results of the test takers' final scores, the total value of each component, and the performance of the questions including the level of difficulty and discriminating power.
- Input data on the question identity menu in the form of the test date, number of questions, number of test takers for each type of question, answer key for each type.
- The data input menu consists of inputting participant numbers, participant IDs, and test takers' answers based on the types of questions.
- The report menu consists of the distribution of answers and the final score of the test takers.
- The performance of the questions is analyzed in the analysis menu with previously inputted data consisting of the level of difficulty, discriminating power, and distractor effectiveness.
- The final score of the test takers, the value of each component, and the quality of the questions are contained in the final data menu.

3.2. Descriptive Analysis of Question Performance

The percentage of answers from the Paper Based Test participants for the 2 the year 2020-2021 period is as follows:

- In the choice of the Faculty of Medicine, the percentage of answers for each question in type X, the choice of answer A is 21.12%, answer choice B is 21.77%, answer choice C is 22.87%. The answer choices D and E are 19.67% and 13.97%, respectively. The percentage of answers type Y, answer choice A is 20.02%, answer choice B is 21.45%, answer choice C is 23.30%. In the answer choices D and E as much as 20.29% and 13.90%. While in Type Z, the test participants chose 19.41% of A's answers, 21.83% of B's answer choices, 23.36% of C's answer choices. In the answer choices D and E as much as 19.74% and 14.01%.
- In the option of not choosing the Faculty of Medicine, the percentage of answers for each question in type X, the choice of answer A is 20.14%, answer choice B is 19.70%, answer choice C is 21.82%. The answer choices D and E were 21.64% and 14.96%, respectively. The percentage of answers type Y, answer choice A was 20.32%, answer choice B was 19.97%, answer choice C was 22.36%. In the answer choices D and E as much as 20.75% and 15.68%. While in Type Z, the test participants chose answer A as much as 15.55%, answer choice B as much as 14.75%, answer choice C as much as 16.70%. The answer choices D and E were 15.53% and 11.76%, respectively.

The Paper Based Test for the period of 2 years 2020-2021 has several components to be tested. Consisting of 25 questions on Islam, 25 questions in English, 25 questions in Logic and Analytical Reasoning, 25 questions in Mathematics or Arithmetic, and 25 questions on Medicine. The following is the result of the average value of each component of the question in Figure 3 and Figure 4.

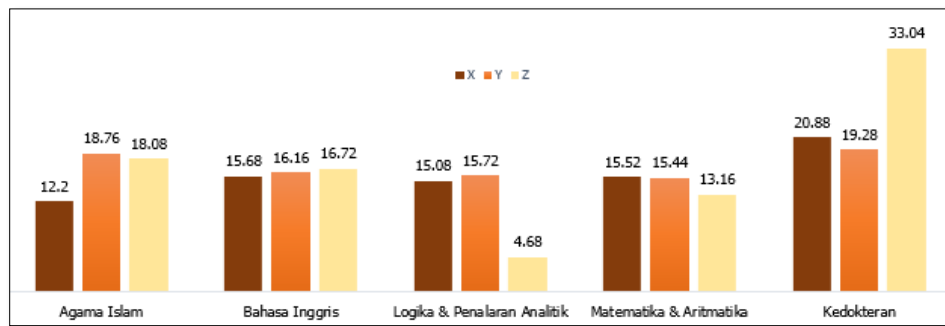


Figure 3. The Average Value of Each Component of the Three Types of Choice Questions in the Faculty of Medicine

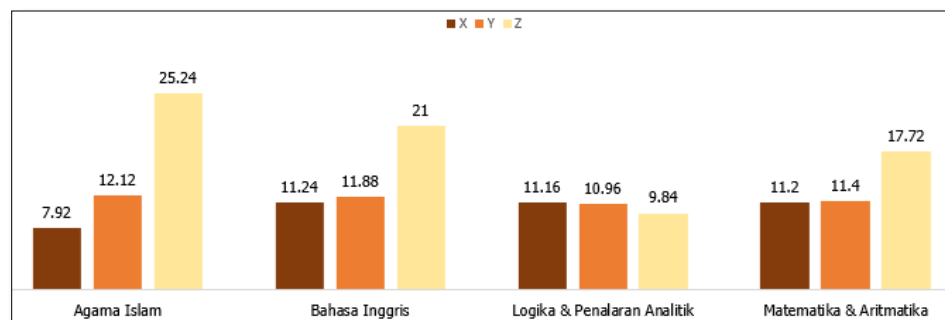


Figure 4. Average Value of Each Component of Three Types of Questions Not Selecting the Faculty of Medicine

Difficulty Level

Questions that have a medium category are considered good, according to [36] an item is said to be good if it has a moderate level of difficulty or is in the range of 0.30–0.70. However, the usefulness of the difficulty level is related to the purpose of the test. The items carried out by the Paper Based Test participants for the 2020-2021 period at the choice of the Faculty of Medicine have a very easy category of 1%, an easy category of 4%, a difficult category of 50%, and a very difficult category of 31%. While the questions that are said to be good or are included in the medium category are only 14%. The choice of not choosing the Faculty of Medicine has a very easy category of 1%, an easy category of 3%, a difficult category of 65%, and a very difficult category of 23%. While the questions that are said to be good or are included in the medium category are only 8%. Follow-up that must be done according to [26] on the level of difficulty is as follows: (1) If the item is in the good category with moderate difficulty, it can be stored in a question bank so that it can be reused in the next test. (2) If the item is categorized as very difficult, difficult, very easy, and easy, it can be revised, discarded, or used in a strict test for questions that are classified as difficult and used on tests that are not tight/loose for questions that are classified as easy.

Distinguishing Power

The items carried out by the Paper Based Test participants for the 2020-2021 period at the choice of the Faculty of Medicine were classified as very good at 0%, in the good category by 1%, in the sufficient category by 10%, in the bad category by 78%, and the rest classified as very bad. In the choice of not choosing the Faculty of Medicine, it was classified as very good at 0%, good category by 1%, sufficient category by 9%, bad category by 79%, and the rest classified as very bad. The follow-up to be done according to [26] on discriminatory power is as follows: (1) If the items are included in the good and very good category, they can be stored in a question bank so that they can be reused in the next test. (2) If the item is classified as low discriminatory, it includes the bad category and it is sufficient that it can be revised or discarded. (3) If the item has a negative discrimination index or 0.00, then it is discarded because it is included in the very bad category.

Distractor Effectiveness

The items carried out by the Paper Based Test participants for the 2020-2021 period at the choice of the Faculty of Medicine, in general, the distractor effectiveness of the 4 options available in the questions

has functioned well because as many as 69% of the available options have been selected at least 5% of the total number of participants. test. In the choice of not choosing the Faculty of Medicine, in general, the distractor effectiveness of the 4 available options in the question has functioned well, because 75% of the available options have been selected at least 5% of the number of test-takers. According to [36], the follow-up to be done on the effectiveness of the distractor is as follows: (1) If the question option is accepted, then the option is classified as good. (2) If the option needs to be revised, then the option is not good. (3) If the option is rejected, then the option is not good.

One-Way ANOVA at Final Grade

This analysis is used to find out whether there is a difference in the average value of each type of question given on the Paper Based Test period 2 years 2020-2021 with the choice of the Faculty of Medicine and not choosing the Faculty of Medicine. The hypothesis of One-Way ANOVA is as follows:

$H_0: \mu_1 = \mu_2 = \mu_3$ (The average value in the population of type X is the same as the average value of the population of questions of type Y is the same as the average value of the population of questions of type Z)

$H_1: \exists \mu_i \neq \mu_j$ (There is one value in the population of each type of question that has a different average from the others)

For the choice of Faculty of Medicine, $F_{\text{count}} = 3.5578 > F_{\text{table}} = 3.0325$, while not choosing Faculty of Medicine, $F_{\text{count}} = 10.158 > F_{\text{table}} = 3.037$. The two results give the decision that rejects H_0 , which means by using a 95% confidence level, it is obtained that there is one value in the population of each type of question that has a different average from the others. These results mean that there is one value obtained by prospective new student test participants from three types of Paper Based Test questions for the period 2 years 2020-2021 in test participants with the choice of the Faculty of Medicine and not choosing the Faculty of Medicine, the Islamic University of Indonesia, which has a different average from the one chosen. other. The average is different from the others based on the results of the one-way ANOVA test, which can be determined by performing a multiple comparison test using the Scheffe test. The results obtained are the average value of the types of questions A and B and the types of questions B and C do not have a significant difference. While in the types of questions A and C there are significant differences. The difference between the types of questions A and C can be seen in the average grades of types A and C. The average value of type A is 24,1951 which is smaller than the average value of type C of 25.8072, so type C has a higher average than type A for the choice of the Faculty of Medicine and the average value of the types of questions A and B and the types of questions A and C did not have a significant difference. While in the types of questions B and C there are significant differences. The difference between the types of questions B and C can be seen in the average grades of types B and C. The average value of type B is 19.6441 which is smaller than the average value of type C which is 17.0833, so type B has a higher average than type C to not choose the Faculty of Medicine.

The independence test was carried out using the Durbin Watson test with the following hypothesis:

$H_0 =$ Residual data is random

$H_1 =$ Residual data is not random

The results obtained give a decision for each type of question to fail to reject H_0 , which means by using a 95% confidence level, it is obtained that the residual data is random in the data chosen by the Faculty of Medicine and does not choose the Faculty of Medicine.

Normality test using Kolmogorov-Smirnov, with the following hypothesis:

$H_0 =$ Residual data is normally distributed

$H_1 =$ Residual data is not normally distributed

The results obtained give a decision for each type of question to fail to reject H_0 , which means by using a 95% confidence level, it is obtained that the residual data is normally distributed in the selected data from the Faculty of Medicine and does not choose the Faculty of Medicine.

Homogeneity test was carried out using the Bartlett test with the following hypothesis:

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ (Residual variance of each homogeneous group)

$H_1: \sigma_i^2 \neq \sigma_j^2$ (The residual variance of each group is not homogeneous)

The results obtained give a decision to reject H_0 , which means by using a 95% confidence level, it is obtained that the residual variance of each group is homogeneous in the selected data of the Faculty of Medicine. While not choosing the Faculty of Medicine, reject H_0 , which means by using a 95% confidence level, it is obtained that the residual variance of each group is not homogeneous, so that the Welch test was carried out and it was found that it failed to reject H_0 , which means by using a 95% confidence level, it was obtained that the residual variance in each group homogeneous.

4. CONCLUSIONS

Based on the results of the research that has been done, it can be concluded as follows:

1. OASIS application (Problem Analysis Data Processing) for performance analysis of multiple choice questions by optimizing Microsoft Excel 2013 functions can facilitate data processing in the form of test taker's answer choices and performance analysis of questions.
2. The performance of the Paper Based Test at the Islamic University of Indonesia with the choice of the Faculty of Medicine and not choosing the Faculty of Medicine shows that: (1) The level of difficulty in the Paper Based Test for the period 2 of 2020-2021 is said to be not good because it is still dominated by the very difficult category of 31% and difficult by 50% in the choice of the Faculty of Medicine. Meanwhile, they did not choose the Faculty of Medicine in the very difficult category by 23% and the difficult category by 65%. (2) The discriminatory power on Paper Based Test questions for the period 2 of 2020-2021 is said to be not good because it is still dominated by the bad category of 78% in the choice of the Faculty of Medicine. Meanwhile, 79% did not choose the Faculty of Medicine. (3) The effectiveness of the distractor on the Paper Based Test for the period of 2 years 2020-2021 it is said that from the 4 options available in the question it has functioned well in the choice of the Faculty of Medicine and did not choose the Faculty of Medicine.
3. There is one value obtained by prospective new student test participants from three types of Paper Based Test questions for the period 2 years 2020-2021 on test participants with the choice of the Faculty of Medicine and not choosing the Faculty of Medicine, the Islamic University of Indonesia which has a different average from the others.

REFERENCES

- [1] A. H. Primandari, "Multiple Comparison Test" [PowerPoint slides], 2018, [Online]. Available: https://arumprimandari.files.wordpress.com/2018/10/course-5_uji-perbandingan-ganda.pdf. [Accessed 12 December 2020].
- [2] A. L. F. Tilaar, V. Sulangi dan A. Madalise, "Analysis of the Quality of Test Items Made by a Class X Mathematics Teacher at SMA Negeri 1 Airmadidi Knows the Teachings of 2019/2020". *JSME (Journal of Science, Mathematics & Education)*, vol. 8, no. 2, pp. 190–196, 2020.
- [3] A. P. Utomo dan M. P. Damayanti, "Design of a Dashboard for an Information System for Monitoring and Evaluation of High School Student Achievement Based on the 2013 Curriculum". *E-BISNIS*, vol. 9, no. 1, pp. 24-31, 2016.
- [4] A. Sudijono, *Introduction to Educational Evaluation*. Jakarta: Raja Grafindo Persada, 2006, 411.
- [5] A. Wibowo, Improvement of Autocorrelation Assumptions Using the Cochran-Orcutt Method (Study of Is Oil and Gas A Blessing Or Disaster". *JMPM: Journal of Mathematics and Mathematics Education*, vol. 5, no. 1, pp. 23-30, 2020.
- [6] A. Z. Izza, M. Falah, dan Susilawati, "Literature Studies: Problems of Learning Evaluation in Achieving Educational Goals in the Era of Free Learning". *Scientific Conference on Education*, vol. 1, no. 1, pp. 10-15, 2020.
- [7] Budiyono, *Statistics for Research*. Surakarta: Sebelas Maret University, 2004, 214-215.
- [8] B. F. Nugraha, "Forecasting Indonesian Inflation Based on CPI on Exports, Imports and Bi Rates," Doctoral dissertation, Institut Teknolohi Sepuluh November, 2018.
- [9] C. Zaiontz, "Real Statistics Using Excel Welch's ANOVA Test", 2015, [Online]. Available: <https://www.real-statistics.com/one-way-analysis-of-variance-anova/welchs-procedure/> [Accessed 9 November 2020].
- [10] D. C. Montgomery, *Design and Analysis of Experiments* (8thed.). Arizona: Arizona State University, 2013.
- [11] D. Rahmasari dan Ismiyati, "Analysis of Items in Introduction to Office Administration". *Economic Education Analysis Journal*, vol. 5, no. 1, pp. 317–330, 2016.
- [12] D. Setiadi, Febriansyah, T. Susanti, "Socialization and Workshop on Application of Processing Values for Middle School/MTS Exams in Pagar Alam City by Using Microsoft Excel 2016". *NGABDIMAS*, vol. 04, no. 01, 2021
- [13] E. Dwimawati, F. Beliansyah, dan S. A. Zulfa, "Community Empowerment through Utilization of Technology in Order to Improve the Quality of Human Resources in Gunung Menyan Village". *Abdi Dosen: Journal of Community Service*, vol. 3, no. 1, 2019.
- [14] Haderani, "Philosophical Overview of the Function of Education in Human Life". *Jurnal Tarbiyah:Scientific Journal of Education*, vol. 7, no. 1, pp. 41–49, 2018.
- [15] Harmastuti dan D. Setyowati, "Utilization of Microsoft Excel for Learning Mathematics and Graphs". *Dharma Bakti Journal*, vol. 1, no. 1, pp 57–66, 2018.

- [16] H. D. Lestari dan D. P. Parmiti, "Development of Science E-Modules with Online Tests to Improve Learning Outcomes". *Journal of Education Technology*, vol. 4, no. 1, pp. 73-79, 2020.
- [17] J. Harlan, *Analysis of Varians*(1sted.). Depok: Gunadarma, 2018.
- [18] J. L. Marpaung, A. Sutrisno, dan R. Lumintang, "Application of ANOVA Method for Analysis of Mechanical Properties of Coconut Fiber Composites". *Online Journal of Mechanical Engineering Axis*, vol. 6, no. 2, pp 151–162, 2016.
- [19] J. R. Lazante, "Testing for Differences Between Two Distributions in the Presence of Serial Correlation Using the Kolmogorv-Smirnov and Kuiper's Tests". *International Journal of Climatology*, Vol. 41, no. 14, pp. 6314-6323, 2021.
- [20] L. Herawati, *Health Data Normality Test Using SPSS* (1sted.). Yogyakarta: Poltekkes Jogja Press, 2016.
- [21] L. Nurdin, Sugiman, dan Sunarmi, "Application of the Combination of Ridge Regression (RR) Method and Generalized Least Square (GLS) Method to Overcome Multicollinearity and Autocorrelation Problems". *Mipa Journal*, vol. 41, no. 1, pp. 58–68, 2018.
- [22] M. B. Brown dan A. B. Forsythe, "Robust Test for The Equality of Variances". *Journal of The American Statistical Association*, vol. 69, pp. 364-367, 1974.
- [23] M. Ilyas, "Analysis of the Final Examination Questions for the Arabic Language Semester of MTs Class VII Output from the Ministry of Religion Banyumas at MTs MWI Kebarongan, Kemranjen, Banyumas for the 2017/2018 Academic Year," Doctoral dissertation, IAIN, 2018.
- [24] M. Puspitarini, "Measuring the Effectiveness of Questions with AnBuso", 2012, [Online]. Available: <https://news.okezone.com/read/2012/08/07/373/674542/ukur-tingkat-efektivitas-soal-dengan-anbso> [Accessed 7 September 2020].
- [25] M. Z. Ashfiya, "Comparison of Engagement Rates between Instagram Photo and Video Posts," Thesis. Yogyakarta State University, 2018.
- [26] N. Agustin, "Analysis of Odd Semester Final Examination Questions for Islamic Religious Education Subjects From the Level of Difficulty and Distinguishing Power in Class VIII IN Bina Insani Metro Integrated Islamic Junior High School," Thesis. State Islamic Institute (IAIN), 2019.
- [27] N. C. Karissanata, "Analysis of the Effect of the January Effect and Moday Effect on the Return of Companies Listed as LQ-45 Index on the Indonesia Stock Exchange for the 2015–2018 Period," Thesis. Islamic University of Indonesian, 2019.
- [28] N. K. Erawati, "Analysis of Competency Achievement Assessment Tests for Midwifery Students". *Penjajora Journal*, vol. 5, no. 2, pp. 111-120, 2019.
- [29] Nurma, "Analysis of Variance (ANOVA)" [PowerPoint slides], 2013, [Online]. Available: https://nurma.staff.uns.ac.id/wp-content/blogs.dir/224/files/2013/03/Chap_1_ANAVA1-2012_1-2013.pdf [Accessed: 7 September 2020].
- [30] Nurmawati, *Evaluation of Islamic Education*. Bandung: Cipta Pustaka Media, 2016, 44-45.
- [31] N. Rumbia, "Analysis of the Quality of Even Semester Exam Questions in the PAI Study Sector class XI at the Islamic High School Integrated Wadah Islaminya Makassar for the 2016/2017 Academic Year," Thesis. Alauddin State Islamic University Makassar, 2018.
- [32] Government of Indonesia, Law no. 20 of 2003 which regulates the National Education System. State Gazette of the Republic of Indonesia Year 2003, No. 4301. Secretariat of State Jakarta, 2003.
- [33] PMB UII, "History and Development", 2020, [Online]. Available: <https://pmb.uii.ac.id/tentang-universitas-islam-indonesia/> [Accessed: 7 September 2020].
- [34] R. Moeriyandani dan B. Yulianto, "Analysis of Dokkai Test Items for Class XI Students Specializing in Japanese at SMA Negeri Sumenep". *Journal of Education and Development*, vol. 9, no. 3, pp. 461-466, 2021.
- [35] R. T. Jurnal, "Quantitative Method with Classical Approach in Application of Item Analysis as Evaluation Media for Quality Question Determination". *Kilat*, vol. 7, no. 1, pp. 15-23, 2018.
- [36] S. Arikunto, *Fundamentals of Educational Evaluation*. Jakarta: Bumi Aksara, 2012, 223-234.
- [37] S. H. Pambayu, "Analysis of the Quality of Test Items at the End of Odd Semester Final Test for Physical Education Subject Class VII SMP Negeri 1 Jati District, Kudus Regency, 2018/2019 Academic Year," Thesis. Yogyakarta State University, 2019.
- [38] Suratini, "The Influence of Education in Improving the Quality of Human Resources in Indonesia". *Journal of Management and Accounting*, vol. 5, no. 1, pp. 68–84, 2017.
- [39] UNY, "Staff Site Yogyakarta State University", 2022, [Online]. Available: <http://staffnew.uny.ac.id/staff/132232818> [Accessed 25 Maret 2022].
- [40] Wulan, "Analysis of Multiple Choice Questions for the Final Assessment of Odd Semester Subjects for Physical Education, Sport and Health Class V at SD Negeri Jetis, Bantul Regency, 2018/2019 Academic Year," Thesis. Yogyakarta State University, 2019.
- [41] W. Walidin, "Directions of Human Resource Development in the Dimensions of Islamic Education". *Journal of Education*, vol. 2, no. 2, pp. 147–163, 2016.
- [42] Y. D. Prastika, "The Effect of Reliability Validity and Difficulty Level on the Quality of Economic Questions Using Anates Software at SMKN 3 Bangkalan," Doctoral dissertation, STKIP PGRI BANGKALAN, 2021.
- [43] Z. H. Siregar, "Implementation of Macro Excel (VBA) as a Registration Information System and Monitoring of Hajj Bailout at PT. Bank Mandiri Cab. Ternate". *IESM Journal*, vol. 1, no. 1, pp. 48-67, 2019.