SAKO: Journal of Islamic Studies, 2(1) - January 2025 12-22

E - ISSN XXXX-XXXX
P - ISSN XXXX-XXXX

Analysis of Multiple Choice Questions Using Classical Measurement Theory in the End of Year Assessment (PAT) of PAI and BP Class VIII Subjects at SMPN 2 Kuantan Mudik

Umi Nasyrah [1], Bustanur [2], Zulhaini [3]
[1] *Universitas Islam Kuantan Singingi, Indonesia*
[2] *Universitas Islam Kuantan Singingi, Indonesia*
[3] *Universitas Islam Kuantan Singingi, Indonesia*

Corresponding Author: Name, Umi Nasyrah E-mail; umi.nasyrah01@gmail.com

Article Information:
Received January 17, 2025
Revised January 28, 2025
Accepted January 30, 2025

ABSTRACT

This study is motivated by the fact that 50% of eighth-grade students at SMPN 2 Kuantan Mudik did not meet the Minimum Competency Criteria (KKM) in the final assessment of the PAI and BP subjects. The aim of this research is to evaluate the quality and follow-up actions for the test items, analyzed using classical test theory. This research employs a mixed-method approach, combining both qualitative and quantitative methods. The sample consists of 44 students using a total sampling technique. Data collection was carried out through documentation, and data analysis used techniques applied to the final assessment (PAT) instrument. The results of this study show that, based on content validity, 94.6% of the test items met the material aspect criteria, 97.6% met the construction aspect criteria, and 100% met the language aspect criteria. An analysis of the cognitive level of the questions shows that 18 questions were categorized as C1 (Remembering), 23 questions as C2 (Understanding), 3 questions as C3 (Applying), and 6 questions as C4 (Analyzing). Furthermore, based on classical test theory, it was found that the questions had an easy difficulty level, poor discriminating power, and low distractor effectiveness. Consequently, follow-up actions recommended include revising 34 questions and discarding and replacing 16 questions

Keywords: *question items classic measurements, multiple choice*

Journal Homepage http://journal.denisyasmartconsulting.co.id/index.php/sako/

How to cite: Nasyrah, U., Bustanur., & Zulhaini (2025). Analysis of Multiple Choice Questions Using Classical Measurement Theory in the End of Year Assessment (PAT) of PAI and BP Class VIII Subjects at SMPN 2 Kuantan Mudik. *SAKO: Journal of Islamic Studies*, 2(1). https://doi.org/10.55849/sako.v1i1.1
Published by: CV. Denisya Smart Consulting

INTRODUCTION

A test is an assessment tool in the form of a task consisting of questions or commands that must be answered by all students, where this test must be structured well so that it can be assessed. In this context, the tests given by teachers are in the form of multiple choice questions in school exams. Where the student's test results will be compared with the minimum completeness criteria (Mania, et.all : 2020). One of the procedures that must be carried out by teachers in order to produce adequate quality questions is through item analysis. In this regard, from the results of interviews during pre-research at SMP N 2 Kuantan Mudik it was discovered that it was true that the teacher made his own multiple choice test for the End of Year Assessment (PAT). However, this teacher has never carried out an analysis of the question items, both in terms of content validity, including aspects of material, construction, language, as well as studying the proportions of cognitive levels, or in terms of quantitative analysis using classical measurement theory. (Guspan &Linovita : 2023)

From the interviews it was also known that the students' scores in the End of Year Assessment (PAT), namely that there were 22 out of 44 students or around 50% who had not reached the Minimum Completeness Criteria (KKM). This indicates that question item analysis should be carried out to produce complete information regarding the problem of students' low scores and to determine the quality of the question items in order to make improvements and refinement of the question items. By seeing this, it is very important to carry out analysis of multiple choice questions, especially using classical measurement theory, in order to know the validity, reliability, level of difficulty, distinguishing power and effectiveness of distracting questions with the aim of being a teacher's guide in determining whether the questions will stored in the question bank, revised or discarded. It is hoped that this will improve the quality of questions and student learning outcomes (Sudijono : 2015)

The formulation of the problem in this research is what is the quality of the questions and how to follow up on the results of the analysis of PAT questions in the PAI and BP class VIII subjects at SMPN 2 Kuantan Mudik using classical measurement theory. So from the formulation of this problem, the aim of this research is to find out the quality of the questions and the follow-up to each question item that has been analyzed. For this reason, the researcher wants to conduct research related to 'Analysis of Multiple Choice Questions Using Classical Measurement Theory in End of Year Assessment (PAT) (Class VIII PAI and BP Subjects at SMP N 2 Kuantan Mudik)'.

RESEARCH METHODOLOGY

This research is included in the type of mixed research (Mixed Method). Where this research uses a qualitative and quantitative approach (Sugiyono : 2019). This research is included in the type of mixed research (Mixed Method). Where this research uses a qualitative and quantitative approach.

The data collection technique used is using documentation techniques. Documentation is a data collection technique for data that is ready, has passed or is secondary data. Documentation is a way of collecting data by taking data from notes,

appropriate administrative documents or the problem being studied. Documentation techniques in this research were used to obtain documents and archives in the form of grid sheets, question sheets, answer key sheets, answer sheets, Learning Objective Flow Mapping sheets, teaching module sheets, as well as students' original score sheets in the End of Year Assessment (PAT) and as a guide in reviewing questions in terms of level proportions, language, material and question construction, analysis of levels of thinking in terms of cognitive level proportions, validity and reliability tests and analysis using classical measurement theory in PAI and BP class VIII subjects at SMP N 2 Kuantan Homecoming (Saat & Mania : 2019).

The data analysis technique used is the analysis technique in the End of Year Assessment (PAT) tool. Where the questions are reviewed in terms of material, language and question construction. Next, we review the analysis of levels of thinking in terms of the proportion of cognitive levels, as well as analyzing the items using classical measurement theory (validity, reliability, level of difficulty, distinguishing power, and effectiveness of distracting questions). Qualitative analysis includes reviewing the questions in terms of material, language and question construction as well as analyzing levels of thinking in terms of the proportion of cognitive levels, namely by carrying out source triangulation techniques through panelists or reviewers and consulting with experts. Meanwhile, quantitative analysis is carried out in the following way.

1. Validity

   To find out validity, look for the biserial correlation coefficient using the following formula :

   $$\gamma_{pbi} = \frac{M_p - M_I}{St} \sqrt{\frac{p}{q}}$$

   Information:

   Ypbi= Biserial Correlation Coefficient

   Mp= Average Score of Correct Answers to Searched Items

   Mt= Mean Total Score

   St= Standard Deviation of Total Score

   P= Proportion of Students Answering Correctly

   q= Proportion of Students Answering Wrong

   The validity of this question item is determined by comparing the coefficient results. If the biserial correlation coefficient is greater than the r-table correlation coefficient, then the question item is empirically valid. The r-table of research conducted with a sample (N=44) at a significance level of 5% (α= 0.05) is 0.2973

2. Reliability

   The reliability test in this research was carried out through an internal reliability test, namely the criteria came from the instrument data itself. The data was obtained from multiple choice questions using the Alpha Cronbach code (R11) formula, which is as follows (Arikunto : 2018):

$$r_{11} = \left[\frac{k}{(k-1)}\right]\left[1 - \frac{\sum \sigma^2 b}{\sigma^2 t}\right]$$

r11 = alpha reliability coefficient
k = number of question items
$\sum \sigma^2 b$ = number of item variants

The interpretation of the results of Cronbach's Alpha reliability calculation (r11) is that when the results of r11 ≥ 0.60 then the question can be declared reliable. On the other hand, when the calculation result of r11 ≤ 0.60 then the question is declared unreliable (Regeta, et.all : 2022)

3. Difficulty Level of Questions

The difficulty level of multiple choice questions can be calculated using the following formula :

P=B/JS

Information:
P= Difficulty index
B = The number of students who answered the question correctly
JS = The total number of students taking the test.

Interpretation of the results of calculating the difficulty index numbers for questions uses the following criteria (Arikunto : 2018)

Indeks Kesukaran (P)	Karakteristik	Keputusan
0,00 – 0,30	Sukar	Ditolak/direvisi
0,31 – 0,70	Sedang	Diterima
0,71 – 1,00	Mudah	Ditolak/direvisi

4. Distinguishing Chetns

To find the differentiating power, you can use the formula as follows

$$D = B_A / J_A - B_B / J_B = P_A - P_B$$

Information:
D = distinguishing power
Ja = number of upper group participants
JB = number of lower group participants
Ba = number of upper group participants who answered correctly
BB = number of lower group participants who answered correctly
PA=Proportion of upper group participants who answered correctly
PB=Proportion of lower group participants who answered correctly

Interpretation of the results of the calculation of differentiating power can use the following criteria

Table 2. Interpretation of the Different Power Index

Indeks Diskriminasi (D)	Karakteristik	Keputusan
0.00 – 0.20	Jelek (*Poor*)	Ditolak
0.21 – 0,40	Cukup (*Satisticfactory*)	Direvisi
0,41 - 0,70	Baik (*Good*)	Diterima
0,71 – 1,00	Baik Sekali (*Excellent*)	Diterima
Negatif	Semuanya tidak baik	Ditolak

5. Effectiveness of Distracting Questions

The formula for the distractor index is calculated using a formula

$$IP = \frac{P}{(N - B)/(n - 1)} \times 100\%$$

Information:

IP: Deception index

P: Number of students who chose the distractor

N: Number of students who took the test

B: Number of students who answered correctly on each question

n: Number of alternative answers (options)

1: Fixed number

The criteria for assessing the effectiveness of using distractors are as follows :

Table 3. Interpretation of the Question Distractor Effectiveness Index

Indeks Pengecoh (IP)	Karakteristik	Keputusan
76%-125%	Sangat Baik	Diterima
51%-75% atau 126%-150%	Baik	Diterima
26%-50% atau 151%-175%	Kurang Baik	Direvisi
0%-25% atau 176%-200%	Jelek	Direvisi
Lebih dari 200%	Sangat Jelek	Ditolak

Meanwhile, to conclude the effectiveness of the distractor's functioning on each question item, you can use criteria adapted from the Likert Scale as follows:

1) If there are three distractor answers that work then the question is said to have very good distractor effectiveness.

2) If there are two distractor answers that work then the question is said to have good distractor effectiveness.

3) If there is only one distractor answer that works then the question is said to have poor distractor effectiveness.

4) If all the distractor answers do not work then the question is said to have poor distractor effectiveness.

After analyzing the items based on validity, reliability, level of difficulty, distinguishing power and effectiveness of distractors. Next, it is necessary to follow up whether the questions need to be revised or discarded by first knowing the quality of the questions as a whole. So, with follow-up, the quality of the questions can be corrected as the next step in achieving improvement after the analysis of the questions is complete. The follow-up can be analyzed using a Likert scale which is grouped into 6 categories as follows:

Jumlah Kriteria yang Terpenuhi (Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh)	Kualitas Butir Soal	Revisi	Masuk Bank Soal
5	Sangat Baik	Terima	Ya
4	Baik	Terima	Ya
3	Sedang	Revisi	Belum
2	Kurang Baik	Revisi	Belum
1	Tidak Baik	Dibuang	Tidak
0	Sangat Tidak Baik	Dibuang	Tidak

RESULT AND DISCUSSION

Qualitative Analysis of Content Validity

This qualitative analysis was used to determine the content validity of the End of Year Assessment (PAT) questions in the PAI & BP class VIII academic year 2022/2023 at SMPN 2 Kuantan Mudik. This content validity analysis was carried out using source triangulation techniques. Where the results of this analysis will then be compared between one source and another. The three sources are panelists as reviewers who have been appointed by the researcher. Where the criteria used to determine this reviewer is that the reviewer has at least graduated from a Bachelor's degree majoring in Islamic Religious Education and taught Islamic Religious Education at the junior high school level. The three reviewers are R (SMPN 1 Kuantan Mudik), IK (SMP IP-ICBS Riau), and Y (SMPN 5 Kuantan Homecoming).

In the initial stage, the reviewer will be given a sheet containing the questions to be reviewed, the study format and study guidelines. After that the reviewers will work individually in different places. The reviewer simply puts a tick in the question review format column if the question item meets the existing criteria in the review aspect. After the study results were collected, the researcher asked for opinions and consulted the results with experts (expert judgment), namely Mrs. Z

Based on the results of the analysis in terms of material, construction and language aspects, from 50 items there are 41 items that have met all indicators in these three aspects. The items are items, namely item questions number 1, 2, 3, 5, 6, 7, 8, 9,

10, 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 33, 34, 36, 37, 38, 40, 41, 42, 43, 44, 45, 47, 48 and 49. high and acceptable and entered into the question bank.

Furthermore, for analysis of the level of thinking in terms of the proportion of cognitive levels. It is known that:

a) C1 (remembering) namely questions number 1, 2, 3, 4, 11, 13, 14, 19, 22, 26, 27, 28, 29, 32, 33, 35, 37 and 43.
b) C2 (understanding), namely questions number 5, 6, 7, 9, 10, 12, 15, 16, 18, 21, 30, 31, 36, 38, 39, 41, 44, 45, 46, 47, 48, 49 and 50.
c) C3 (apply), namely questions number 34, 40 and 42.
d) C4 (analyze), namely questions number 8, 17, 20, 23, 24 and 25.

Quantitative Analysis

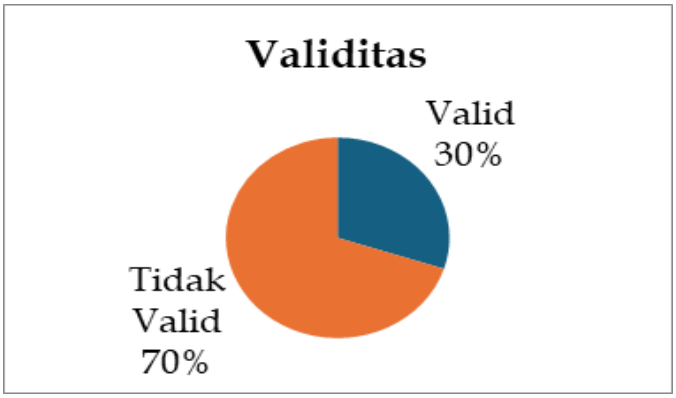
This quantitative analysis was carried out using measurement theory, namely classical measurement theory. The results of the analysis and discussion are as follows:

1. Question validity analysis

   Testing the validity of multiple choice questions was carried out using the biserial correlation coefficient formula which was carried out with the help of the SPSS application program. The results of calculating the validity of the question items will then be consulted with the rtable at a significance level of 5% for the sample (N=44), namely 0.2973. If rtable < rcount then the question item is declared valid. Conversely, if rtable > rcount then the question item is declared invalid

   Based on the results of the validity analysis of multiple choice questions using the SPSS program, it is known that 15 questions (30%) were declared valid, namely question items number 2, 4, 5, 8, 11, 18, 20, 31, 33, 35, 36, 38, 47, 48 and 49. Meanwhile, for invalid questions there are 35 questions (70%), namely questions number 1, 3, 6, 7, 9, 10, 12, 13, 14, 15, 16, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 34, 37, 39, 40, 41, 42, 43, 44, 45, 46, and 50. This indicates that of the total multiple choice questions, there are more invalid questions than valid questions.

Figure 1: PAT Question Item Validity Diagram

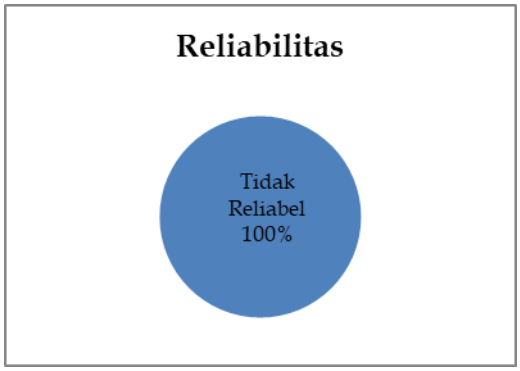


2. Analysis Of Reliability Problems

   The reliability testing of this multiple choice item is carried out using the Alpha Cronbach formula which is done with the help of the SPSS application program. The results of the calculation of the reliability of the items will be interpreted in a way if the results of R11 ≥ 0.60 then the problem can be declared reliable. Conversely, when the calculation results R11 ≤ 0.60 then the problem is declared not reliable.

   Based on the results of the reliability analysis of multiple choice questions using the SPSS program, it is known that none of the items (0%) are declared valid. As for the invalid questions, there are 50 items (100%) or the total item of items, namely item number 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and 50. This indicates that the overall grains of multiple choice are declared not reliable.

   Figure 2: Reliability Diagram of PAT Question Items

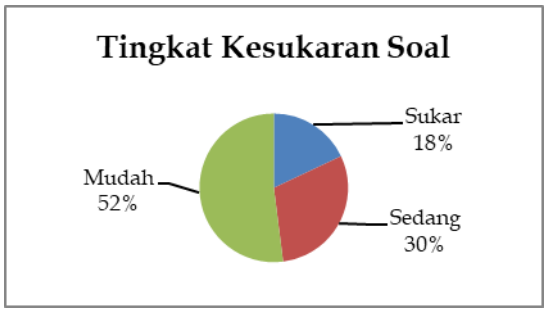


3. Analysis of the Difficulty Level of the Questions

   This difficulty level test was carried out using the SPSS program. The results are interpreted into three criteria, namely questions with a difficulty index (P) of 0.00 to 0.30 are questions that are classified as difficult, questions with a difficulty index (P) of 0.31 to 0.70 are questions that are classified as medium and questions with an index Difficulty (P) 0.71 to 1.00 is a question that is relatively easy.

   Based on the results of the analysis of the level of difficulty of multiple choice questions using the SPSS program, it is known that of the 50 questions there are 9 questions (18%) classified as difficult, namely questions number 8, 9, 12, 15, 16, 20, 31, 42 and 44 Then there are 15 questions (30%) classified as medium, namely questions number 2, 3, 4, 5, 7, 10, 11, 14, 17, 39, 41, 45, 46, 47 and 48. And there are 26 questions (52%) which are classified as easy, namely questions number 1, 6, 13, 18, 19, 21, 22, 23 , 24, 25, 26, 27, 28, 29, 30, 32, 33, 34, 35, 36, 37, 38, 40, 43, 49 and 50. So it can be concluded that the End of Year Assessment (PAT) questions in class VIII PAI & BP subjects at SMPN 2 Kuantan Mudik have easy difficulty level.

Figure 3: Diagram of Difficulty Levels of PAT Question Items

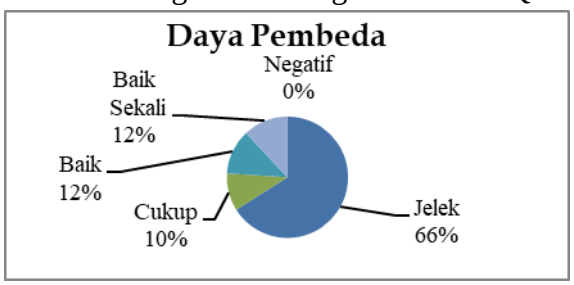


4. Analysis of the Discriminating Power of the Questions

   This differentiating power test was carried out to see the ability of the test items to differentiate students with high abilities from students with low abilities. This differentiating power analysis was carried out using the SPSS program. The results are interpreted into five criteria, namely questions with a discrimination index (D) of 0.00 to 0.20 have poor criteria, questions with a discrimination index (D) of 0.21 to 0.40 have sufficient criteria, questions with a discrimination index (D) 0.41 to 0.70 has good criteria, questions with a discrimination index (D) of 0.71 to 1.00 have very good criteria, and questions with a discrimination index (D) are negative having all the criteria is not good.

   Based on the results of the analysis of the differentiating power of multiple choice questions using the SPSS program, it is known that there are 33 questions (66%) classified as bad, namely question items number 1, 2, 4, 5, 8, 9, 10, 11, 16, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 35, 36, 38, 39, 40, 44, 45, 47, 48 and 49. Then there are 5 questions (10%) which are classified as sufficient, namely question items number 6, 12, 13, 19 and 32. Then there are 6 question items (12%) are classified as good. Then there were 6 questions (12%) that were classified as very good, and it was known that there were no question items that had a negative value, meaning that there were no question items that were all bad. So it can be concluded that the End of Year Assessment (PAT) questions in the PAI & BP class VIII subjects at SMPN 2 Kuantan Mudik have poor differentiating power.

   Figure 4: Differentiating Power Diagram of PAT Question Items



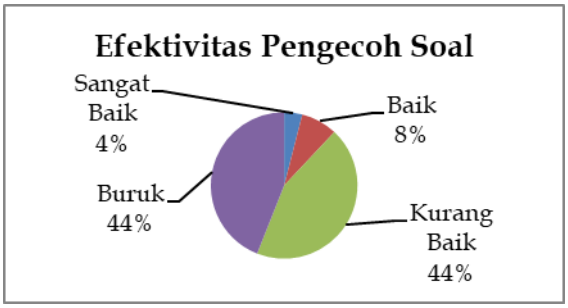
5. Analysis of the Effectiveness of Distracting Questions

   Testing the effectiveness of this question distractor was carried out using the help of the SPSS program. The results are interpreted into five criteria, namely options

with a distractor index (IP) of 76%-125% have very good criteria, options with a distractor index (IP) of 51%-75% or 126%-150% have good criteria, options with a distractor index (IP) 26%-50% or 151%-175% has poor criteria, options with a distractor index (IP) 0%-25% or 176%-200% has bad criteria and options with a distraction index (IP) of more than 200% have very bad criteria.

Based on the results of the analysis, it is known that there are 2 questions (4%) with very good question effectiveness, namely question items number 9 and 50. Then there are 4 question items (8%) with good question effectiveness, including question items number 16, 34, 40 and 45. Furthermore, there are 22 questions (44%) with poor distractor effectiveness, including question items number 1, 2, 3, 4, 5, 8, 10, 11, 13, 14, 17, 20, 32, 35, 38, 39, 41, 42, 43, 44, 48 and 49. And there are 22 questions (44%) with poor distractor effectiveness, namely questions numbers 6, 7, 12, 15, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 36, 37, 46 and 47. So the effectiveness of the distractor questions has the effectiveness of distracting questions that function less well or poorly.

Figure 5: Diagram of Effectiveness of Pat Item Chatting



The form of follow -up from the items in the year -end assessment (PAT) of PAI & BP Class VIII Subjects at SMPN 2 Kuantan Mudik. It is known that there are 34 items that if you want to be put back into the question bank, then it must go through improvement or revision first. The items referred to are item questions number 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 17, 18, 19, 20, 32, 33, 34, 35, 36, 36, 36, 36, 36 , 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49 and 50. Number 9, 12, 15, 16, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 and 44.

CONCLUSION

The quality of the questions in the End of Year Assessment (PAT) for Class VIII PAI and BP subjects at SMPN 2 Kuantan Mudik shows that based on content validity it is known that 94.6% of the questions meet the requirements for the material aspect, 97.6% for the construction aspect and 100 % for language aspects. Then, analyzing the level of thinking based on the proportion of cognitive level, it was found that there were 18 questions including C1 (Remembering), 23 C2 questions (Understanding), 3 C3 questions (Applying), and 6 C4 questions (Analyzing). Furthermore, based on classical measurement theory, it is known that the questions have a difficulty level that is relatively easy, the discriminatory power is classified as poor, and the effectiveness of distracting questions is classified as poor. So the follow-up carried out was 34 questions

(revised) and 16 questions should be (discarded and replaced). The questions in question are questions number 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 17, 18, 19, 20, 32, 33, 34, 35, 36 , 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49 and 50. Then the results were also obtained that there were 16 questions that should be discarded and would not be included in the question bank, namely question items number 9, 12, 15, 16, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 and 44

REFERENCES

Arifin, Zainal, *Evaluasi Pembelajaran*, Bandung: PT. Remaja Rosdakarya, 2013.

Arikunto, Suharsimi, *Dasar Dasar Evaluasi Pendidikan Edisi Ketiga*, Jakarta: Bumi Aksara, 2018.

Guspan dan Linovita, Guru Pendidikan Agama Islam di SMPN 2 Kuantan Mudik, *wawancara* (Bukit Kauman, 25 November 2023. Pukul 09.30 WIB).

Mania Sitti, Fitriani Nur, Ahmad Farham Majid, Nidya Nina Ichiana, dan Andi Ika Prasasti Abrar, *Analisis Butir Soal Ujian Akhir Sekolah,* [Al Asma : Journal of Islamic Education Vol. 2 No. 2 Tahun 2020] hlm. 274-284.

Muchlizani, Nurul, dkk. "Analisis Kualitas Butir Soal Ujian Akhir Semester Mata Pelajaran Akidah Akhlak Kelas V MI Radhiatul Adawiyah Makassar," [Jurnal Inspiratif Pendidikan Vol. XII, No. 1 Tahun 2023], hlm. 234.

Regeta, Neko Rossa, dkk, "Analisis Butir Soal Pilihan Ganda pada Materi PLSV dan PTLSV Siswa SMP Negeri 2 Wiradesa", [Prosandika Vol. 4, No. 1 Tahun 2022], Hlm. 477

Saat, Sulaiman dan Sitti Mania, *Pengantar Metodologi Penelitian Panduan Bagi Peneliti Pemula.* Gowa: Pusaka Almaida, 2019.

Sudijono, Anas, *Pengantar Evaluasi Pendidikan.* Jakarta: PT Raja Grafindo Persada, 2015.

Sugiyono, *Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, Kombinasi, R&D dan Penelitian Pendidikan)*. Bandung: Alfabeta, 2019.

Prawiki, Suci Mitra dan Helendra, "Analisis Kualitas Soal Ujian Akhir Semester Ganjil Tahun Pelajaran 2020/2021 Mata Pelajaran Biologi Kelas X SMA Negeri 1 Teluk Sebong", [Jurnal Biologi dan Pembelajarannya Vol. 17, No. 2 Tahun 2022], Hlm. 19.

Warju, dkk, "Analisis Kualitas Butir Soal Tipe HOTS pada Kompetensi Sistem Rem Siswa di Sekolah Menengah Kejuruan", [Jurnal Pendidikan Teknologi dan Kejuruan Vol. 17, No. 1 Tahun 2020], Hlm. 98