

Item Analysis of Arabic Thematic Questions to Determine Thinking Level Ability

Suleman D. Kadir¹✉ IAIN Sultan Amai Gorontalo, Indonesia
sulemand.kadir@gmail.com¹

Suharia Sarif² IAIN Sultan Amai Gorontalo, Indonesia
suhariasarif@iaingorontalo.ac.id²

Abdul Hafizh Nur Fuadi³ UIN Maulana Malik Ibrahim Malang, Indonesia
hafizhabdul47@gmail.com³

 <https://doi.org/10.58194/eloquence.v3i1.1498>

Corresponding Author: ✉ Suleman D. Kadir

Article History

Received:
06-03-2024
Accepted:
01-04-2024
Published:
30-04-2024

ABSTRACT

Background: Thematic Arabic language test items are crucial to be analyzed, but the interesting aspect is to determine whether the Arabic language questions used meet the criteria of HOTS (High Order Thinking Skill) questions.

Purpose: To analyze HOTS (High Order Thinking Skills)-based thematic Arabic language questions, which involves the analysis of test items that meet the criteria and possess high-level thinking skills.

Method: This research employs a mixed-method approach, combining qualitative and quantitative methods, as the analysis in this research requires examination from these two perspectives to generate accurate verbal and numerical results. The research sample consisted of 101 individuals from 3 classes. Data collection was conducted by distributing the test items to the students and analyzing them using the Anates application..

Results and Discussion: The analysis results show that the existing test items still contain a low level of LOTS (Low Order Thinking Skill) thinking, therefore improvements and revisions are necessary so that they can be re-tested in the next stage. Specifically, based on the analysis, there are several questions that still fall into the C3 criteria (items 6 and 15), C2 (item 14), and C1 (items 7-12). Meanwhile, in terms of quality, as measured through the Anates application, the XY correlation has a value of 0.63, the test score reliability is 0.78, and the difficulty level is found in items 2, 14, 11, and 15, which still need to be corrected and revised.

Conclusions and Implications: The thematic questions that have been tested meet the HOTS criteria, but in terms of quality, they are still relatively moderate, with some questions leaning towards LOTS, therefore requiring revisions. Research Implications: The findings of this study can be used as a reference for educators to design and analyze HOTS-based questions in Arabic language learning, particularly in relation to the enhancement of students' higher-order thinking skills. These results can also encourage the development of more comprehensive assessment instruments that align with the current curricular demands.

Keywords:

Thematic Question Items; Thinking Level Ability; HOTS; Arabic Language

ABSTRAK

Latar Belakang: Butir soal tematik bahasa Arab sangat penting dianalisis, namun hal sangat menarik adalah menjawab apakah soal bahasa Arab yang digunakan memenuhi kriteria soal-soal HOTS (High Order Thinking Skill)

Tujuan: Untuk menganalisis soal tematik berbahasa Arab berbasis HOTS (High Order Thinking Skills) yaitu analisis butir soal yang masuk ke kriteria dan memiliki daya berfikir tingkat tinggi.

Metode: Penelitian ini menggunakan metode gabungan antara kualitatif dan kuantitatif sebab analisis di dalam penelitian ini memerlukan analisis dari dua sudut pandang tersebut agar menghasilkan hasil yang akurat secara verbal dan numerik. Sampel penelitian berjumlah 101 orang yang terdiri dari 3 kelas. Cara pengumpulan data dilakukan dengan mendistribusikan soal kepada peserta didik dan menganalisisnya menggunakan aplikasi Anates.

Hasil dan Pembahasan: Hasil analisis bahwa butir soal yang ada masih mengandung tingkat berfikir rendah LOTS (low order thinking skill), oleh sebab itu perlu diadakan perbaikan dan revisi agar bisa diujikan kembali pada tahap berikut. Khususnya berdasarkan hasil analisis bahwa ada beberapa soal yang masih masuk kriteria C3 yakni soal nomor 6 dan 15, C2 terdapat pada soal nomor 14, C1 terdapat pada soal nomor 7-12. Sedangkan berdasarkan kualitasnya yang diukur melalui aplikasi anates bahwa korelasi XY memiliki nilai sebesar 0.63 reliabilitas nilai tesnya angka 0.78 dan tingkat kesukaran ditemukan pada soal nomor 2, 14, 11, dan 15 yang masih perlu dikoreksi dan direvisi.

Kesimpulan dan Implikasi: Soal tematik yang sudah diujikan masuk kriteria HOTS berkualitas cukup dari sisi kualitasnya, namun beberapa soal mengarah kepada LOTS sehingga perlu direvisi. Implikasi penelitian ini adalah hasil penelitian ini dapat digunakan sebagai referensi bagi pendidik untuk merancang dan menganalisis soal-soal berbasis HOTS dalam pembelajaran bahasa Arab, khususnya terkait dengan peningkatan kemampuan berpikir tingkat tinggi peserta didik.

Butir Soal Tematik; Kemampuan Tingkat Berfikir; HOTS; Bahasa Arab

Kata Kunci



Copyright: © 2024 by the author(s).

This is open access article under the

[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

INTRODUCTION

Learning assessment is part of the learning process that must be tried. Learning assessment is an activity to control, guarantee, and determine the quality of learning on various learning components in each learning process, level, and type of learning as a form of accountability for the implementation of learning. In learning assessment, several stages must be passed, namely measurement, assessment, and evaluation. These three stages are always related to one another and cannot be separated.[1]

So that improving the quality of education and learning is the main concern after the UN (National Examination) was changed to AN (National Assessment) in 2020 as a form of evaluation of inputs, processes, and learning outcomes in the domain of improving the quality of learning and learning outcomes of students.[2] One of the assessment items in the AN (National Assessment) that is used as an evaluation of national education is AKM (Minimum Competency Assessment) as a measurement of the literacy and numeracy ability of students. To be able to have this ability, students need skills in analyzing and trying to solve problems from information obtained at learning time as an improvement to the quality of students. To arrive at that stage, students must be presented with skills in thinking to analyze and find a solution to the problems faced. So what can be done is by testing HOTS (Higher Order Thinking Skills) based questions for the ability to think at a high level which is presented with several question items. Since K13 is applied, the questions used for students are HOTS.[3]

Analysis of Arabic thematic questions using one level of high thinking skills, namely HOTS, is a study in this paper. Because the HOTS question points have not been a serious concern in several studies related to the assessment of Arabic language learning. When viewed so far, thematic

tests are still rare and less desirable and considered as a form of assessment in educational evaluation. Therefore, when reviewed in depth, there will be problems with this. So, more research is needed to obtain authentic and valid information. This research needs to be carried out because this study focuses on the procedures for analysis to determine the quality or absence of the questions tested to students as an exercise for them in analytical thinking and solutions to the problems they face.[4]

Several studies have been carried out before, such as Fauziyah (2020) and colleagues who analyzed related question items using HOTS on UAS. Laila (2020) analyzed the question items and concentrated on the construction of question items from a qualitative perspective. The same thing was done by Maulana and Sanusi (2020) analyzing the question points at UAS in Madrasah Ibtidiah 2018.[5]–[7]

So based on this, the author examines thematic questions in Arabic because previous research was more oriented towards analyzing the question items on the UAS exam which were adjusted to the study of interest and the direction of purpose that became the concentration of the research they studied. However carefully examined some writings have similarities with the writings raised in this study, namely the writings studied by Fina and Zidni which are both objects of study direction are Arabic thematic questions but with different material themes.[8]

This writing aims to analyze the thematic question points in Arabic language learning with the level of high-level thinking skills HOTS in class XI MAN 1 Gorontalo Regency with the theme of material *الحج والعمرة* and *تكنولوجيا الإعلام والاتصال*. This research requires mapping objective learning outcomes, planning further teaching materials, by measuring the quality of the learning process, and also as a prediction of further learning. These question points are a form of assessment in Arabic language learning that is important for graduation in Arabic language learning. This thematic test makes it very possible to consider and predict failures in teaching and learning with the closest Measurement time lag and makes educators more aware of the object of mapping the cognitive abilities of their students.[9]

The side of accuracy and efficiency used to test the competence of students in understanding the theme of the material *الحج والعمرة* and *تكنولوجيا الإعلام والاتصال* which will be decisive in the wearability of questions in further learning. Questions that are less representative to be tested are validity, reliability, discriminating power, and low deception will be revised and improved to make them worthy of testing. As for the questions that have high accuracy, they will be stored as exam questions in the question bank. The most important thing for HOTS questions in thematic question tests is needed to be tested to measure the competence of students with the cognitive level of HOTS. For problems that have a high level of thinking skills, HOTS can provide critical-analytical thinking skills training to find solutions to existing problems.[10]

This research was carried out at MAN 1 Gorontalo Regency with the subject of class XI research for the 2021/2022 school year by distributing thematic test questions to students in the class. This place is the object of research because no one has carried out quality measurement of thematic questions based on HOTS before in Arabic language learning. The reason for this research is because there is still a lack of this research carried out. The purpose of conducting research is to measure the quality of questions so that the achievement of the cognitive domain of students can be known through the HOTS question points.

LITERATURE REVIEW

Item analysis is a systematic process to evaluate the effectiveness of each item or question in a test.[11] This analysis is crucial to determine the quality of questions used in an assessment

instrument. Through item analysis, teachers or researchers can identify the strengths and weaknesses of each test item, so that revisions or improvements can be made to enhance the validity and reliability of the instrument.[12] Thus, item analysis becomes a crucial step in ensuring that the assessment instrument used is truly capable of accurately and comprehensively measuring the students' abilities.

1. Quantitative Item Analysis Techniques (Using Anates Application)

Item analysis can be conducted through a quantitative approach, one of which is by utilizing the Anates application. Anates enables the measurement of various aspects of item quality, such as difficulty level, discriminating power, validity, and reliability.[13] Through Anates, teachers or researchers can obtain relevant statistical data to evaluate the quality of each test item in detail. The analysis results can then be used as a basis for decision-making regarding the revision, improvement, or replacement of items that do not meet the criteria of good quality.

2. Implications of Item Analysis Results on Revision and Improvement of Item Quality

The findings from item analysis have very important implications for teachers or researchers in improving the quality of the assessment instruments used. Items identified as having low difficulty level, discriminating power, validity, or reliability need to be revised or even replaced.[14] The revision process must consider the characteristics of good test items, such as alignment with learning indicators, clarity of wording, and accuracy of answer choices. By systematically making improvements based on the analysis results, the quality of assessment instruments, particularly thematic Arabic language questions, can be continuously enhanced.[15] This will ultimately support the valid and reliable measurement of students' higher-order thinking skills.

3. Higher Order Thinking Skills (HOTS)

Higher Order Thinking Skills (HOTS) refer to the ability to not only remember and understand, but also analyze, evaluate, and create.[16] In the context of Arabic language learning, HOTS becomes crucial as it can encourage students to not only memorize vocabulary or grammar rules, but also develop critical and creative thinking skills in using the Arabic language.[17] The theoretical foundation of HOTS in Arabic language learning is based on cognitive theories that emphasize more complex mental processes, such as problem-solving, decision-making, and analytical thinking.

a. Strategies for Developing HOTS-based Questions in Arabic Language Learning

The development of HOTS-based questions in Arabic language learning can be done through several strategies, including: (1) formulating learning indicators that are oriented towards higher-order thinking skills; (2) developing contextual and engaging question stimuli; (3) constructing questions that demand analysis, evaluation, and creativity; (4) using varied question formats, such as essays, complex multiple-choice, and case studies; and (5) conducting regular item analysis to ensure the quality of HOTS-based questions. These strategies can be applied in an integrated manner when designing HOTS-based assessment instruments for Arabic language learning.[18]

b. The Role of HOTS in Improving Students' Arabic Language Proficiency

The implementation of HOTS in Arabic language learning plays an important role in enhancing students' language skills. HOTS-based questions can encourage students to think critically in understanding Arabic texts, analyzing sentence structures, evaluating information, and using Arabic creatively and innovatively.[19] Additionally, HOTS can also help students develop communication skills, problem-solving, and reflective thinking, which are essential competencies in Arabic language learning. Therefore, the development of HOTS-based questions in Arabic language assessment can significantly contribute to the comprehensive improvement of students' Arabic language proficiency.

METHOD

The method used in this study is combined (qualitative and quantitative) because in this study researchers see the quality side by analyzing it using HOTS high-level thinking skills. The quantitative side is to find out the calculation results of each respondent who answers correctly and precisely.

The method of data collection from this study is through class sampling tested in three classes, namely Religion, Language, and Social Studies classes in the MAN 1 education unit of Gorontalo district. Each class is selected one class for research purposes. The number of students in religious classes is 30 people, language 36 people, and social studies 35, so the total sample of research subjects is 101-person students. The theme of the material tested is thematic questions related to *الحج والعمرة* and *تكنولوجيا الإعلام والاتصال* which are in Arabic textbooks published by the Ministry of Religion of the Republic of Indonesia which were released in 2021 with a total of 15 questions. Students do these questions with the help of e-learning media which are facilities available in madrasah.

This research was carried out in several steps. The first step is to prepare the question items to be tested. The second step is to test the question items on students who the subject of research are. Test results that have been worked on by students through e-learning will be compiled and converted to Excel form so that they are easily transferred to the analysis application of multiple-choice question analyses. The third step is to calculate the quality of the questions that have been tested on students. After obtaining the results of the application calculation, the quality of the questions is presented in the form of a table so that the fourth step can be carried out, namely the analysis process and the next stage is conclusions. The scheme of this study is as follows:

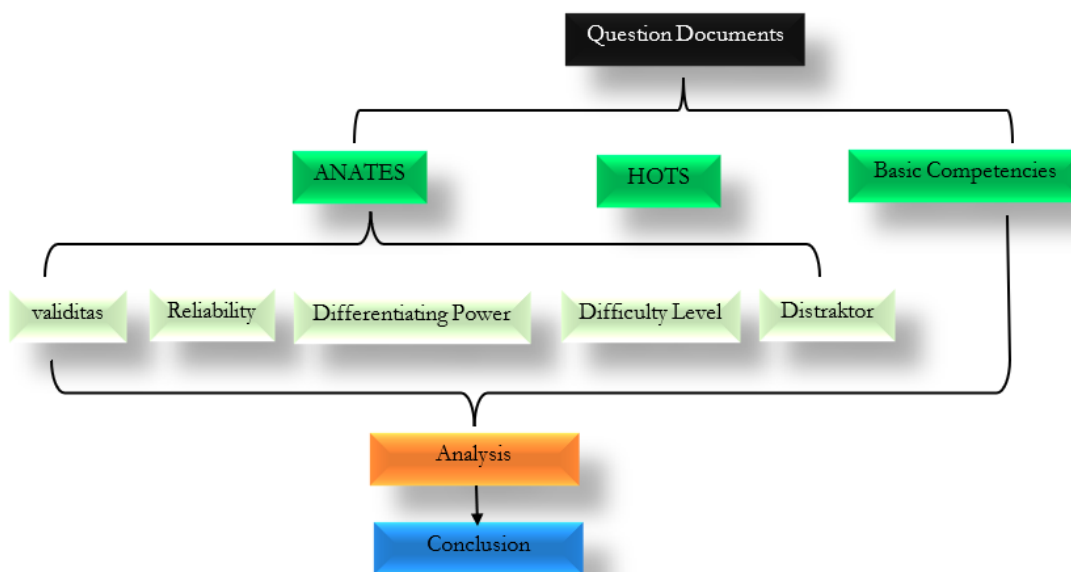


Figure 1. HOTS Question Point Research Scheme

RESULT AND DISCUSSION

1. Basic Competency Analysis

The theme of the material that will be given to students is four competencies plus language rules in the material *الحج والعمرة* and *تكنولوجيا الإعلام والاتصال* KD (Basic competencies) on the theme of this material as follows:

Table 1. Table 1. Congestive level analysis on the material theme
تكنولوجيا الإعلام والاتصال والحج والعمرة

Theme	No. KD	KD (Basic competence)	level
معرفة الجسم	3.7	Understand the social function, structure of the text, and linguistic elements (sounds, words, and meanings) of the text related to the theme: مكة والمدينة والحج والعمرة which involves speech acts stating and asking actions/events in the past by paying attention to the form, meaning, and function of تصريف الفعل الماضي اللغوي	C5
	3.8	Analyze ideas from Arabic texts related to the theme: مكة والمدينة والحج والعمرة by paying attention to the form, meaning, and function of the grammatical arrangement تصريف الفعل الماضي اللغوي	C4
	4.7	Demonstrate speech acts stating and questioning actions/events in the past by paying attention to the form, meaning, and function of تصريف الفعل الماضي اللغوي both orally and in writing.	C4
	4.8	Assess the results of the analysis of ideas from Arabic texts related to the theme مكة والمدينة والحج والعمرة by taking into account the form, meaning, and function of the grammatical arrangement of تصريف الفعل الماضي اللغوي both orally and in writing	C6
الاستيعاب والتفكير الناقد	3.9	Understand the social function, structure of the text, and linguistic elements (sounds, words, and meanings) of the text related to the theme: (الأونلاين، الجوال، الحاسب، الاتصال) تكنولوجيا الإعلام والاتصال which involves the act of speech stating and asking actions/events in the present and future by taking into account the form, meaning, and function of تصريف الفعل المضارع اللغوي	C4
	3.10	Evaluate the ideas of the Arabic text related to the theme: (الأونلاين، الجوال، الحاسب، الاتصال) تكنولوجيا الإعلام والاتصال by taking into account the form, meaning, and function of the grammatical arrangement تصريف الفعل المضارع اللغوي	C5
	4.9	Demonstrate speech acts stating and asking current and future actions/events by taking into account the form, meaning, and function of تصريف الفعل المضارع اللغوي both orally and in writing.	C4
	4.10	Presenting the results of the evaluation of ideas from Arabic texts related to the theme تكنولوجيا الإعلام والاتصال by taking into account the form, meaning, and function of the grammatical arrangement of تصريف الفعل المضارع اللغوي both orally and in writing.	C5

Source.[20]

Mapping the ability of the cognitive level of students to achieve KD, the distribution of question items in the four competencies plus language rules are weighed based on adequacy for KD achievement that adjusts to the question indicators. The distribution of question items can be seen below:

Table 2. Distribution of material question items on the theme
تكنولوجيا الإعلام والاتصال والحج والعمرة

Quantities of the Items in Question	Question categories
4	المفردات
3	الاستماع والكلام
4	القرءاة
2	الكتابة
2	القواعد
15	Number of questions

2. HOTS Operational Analysis

HOTS questions are not questions made to make it difficult for an answerer or respondent but are intended to use thinking activities that produce high reasoning to be able to solve and find solutions to problems faced when presented with problems that demand fast, correct, and precise solutions.[21]

Question items that contain high-level thinking skills HOTS is a demand for answerers to use deep analytical thinking power to always be precise and correct in answering the questions presented and train and familiarize their reasoning power to run well. If viewed in general, questions that contain high-level thinking HOTS must be different from other questions, and it will be difficult to find the correct answer. But this is where students are trained to have deep reasoning power about every problem faced and demand it to find solutions to these problems.[22]

The characteristic of HOTS question items is always to ask and direct the answerers to the content of several verbs, namely, creating (C6), evaluative (C5), or analytical (C4). As for the characteristics that still contain verbs; understand (C2), remember (C1), or apply (C3) it can be ascertained that the problem is at the level of LOTS (Lower Order Thinking Skills or in the group of lower order thinking abilities. If in the preparation of the problem, it can be through various operational verbs (KKO) objectives, namely C4, C5, and C6 in the revised version of Bloom's taxonomy. Which refers to the grid of questions contained in Core Competencies (IC) and Basic Competencies (KD) as in KMA 183 2019 related to the Islamic religious education curriculum and Arabic language learning in madrasah.[23]

Table 3. Description of HOTS Material Questions

تكنولوجيا الإعلام والاتصال والحج والعمرة

Question No.	Material	No. KD	Question Indicator	Level cognitive
1	الحج والعمرة	3.7	Several kinds of worship are presented. Students analyze umrah and hajj	C4
2		3.7	Pictures of worship activities are presented. Learners identify and associate it with Umrah and Hajj	C4
3		4.7	Served <i>hiwar</i> . Learners are asked to associate communication with inferring the place where the conversation takes place	C5
4		3.8	Presented reading texts about Umrah and Hajj. Learners conclude the difference between Umrah and Hajj	C5
5		3.8	Presented reading text. Learners analyze underlined text.	C4
6		4.8	Presented intermittent words. Learners are asked to sort the word into a correct sentence	C3
7		4.8	Presented a choice of rule texts about تصريف الفعل الماضي اللغوي students identify and choose the right rule choices in the text.	C1
8		4.8	Presented the <i>jumlab</i> (sentences) short. Learners identify the rules تصريف الفعل الماضي اللغوي	C1
9	تكنولوجيا الإعلام والاتصال	3.9	Presented types of images of information technology and transportation. Learners categorize by type	C1
10		3.9	Presented types of images of information technology and transportation. Learners categorize by type.	C1
11		4.9	Short communication is presented using المضارع اللغوي. Learners identify different sentence structures.	C1
12		4.9	Served hiwar. Learners are to complete readings according to language situations.	C1
13		3.10	Presented reading text. Students are asked to conclude the content of the reading.	C5
14		4.10	Presented readings. Students interpret the meaning of underlined phrases.	C2
15		4.10	Presented intermittent words. Learners are asked to sort into a correct sentence.	C3

Based on the table of the results of the analysis of the quality of the question items, a total of 15 questions tested were found to have 9 questions that entered the LOTS quality level, namely questions number 6 and 15 (C3), 14 (C2) and 7-12 (C1). If you look at such a question number, it does not fall into the category of questions for HOTS or higher-order thinking skills. So that in the future when testing again, it needs to be revised and corrected qualitatively. This form of HOTS-level assessment is needed to measure the achievement of student competence in learning and can be used as reflection material in compiling progress reports on student learning outcomes and improving the quality of learning in a better direction. Using questions that have a HOTS level will make students familiar with challenges that lead to the development of students' potential to produce quality output. So based on this, in learning Arabic as a foreign language, more adaptation is needed for native speakers.[24]–[26]

3. Question Quality

a. Validitas

The validity of the question item is a measurement instrument used in measuring what must be measured. In this case, the researcher intends to measure the suitability of the question items with the syllabus and the Learning Implementation Plan (RPP).[27] The validity of the Anates application is by correlating the scores of the question items with the overall score. The results of the validity test of the question items in the following table:

Table 4. Significance of Question Items

No. Question Items	Correlation	Significance
1	0.583	Significance
2	0.525	Significance
3	0.484	Significance
4	0.478	-
5	0.253	-
6	0.583	Significance
7	0.256	-
8	0.548	Significance
9	0.571	Significance
10	0.658	Very significance
11	0.587	Significance
12	0.561	Significance
13	0.591	Significance
14	0.514	Significance
15	0.515	Significance

Table 5. Limits of Significance of Correlation-Coefficients

Df (N-2)	P=0.05	P=0.01	df (N-2)	P= 0.05	P=0.01
10	0.576	0.708	60	0.250	0.325
15	0.482	0.606	70	0.233	0.302
20	0.423	0.549	80	0.217	0.283
25	0.381	0.496	90	0.205	0.267

30	0.349	0.449	100	0.195	0.254
40	0.304	0.393	125	0.174	0.228
50	0.273	0.354	150	0.159	0.208

Based on Table 4 above, it can be seen that invalid SOA items are found in question numbers 4, 5, and 7. Therefore, the invalid questions must be excluded or revised to the questions as a further improvement. With the measurement of the validity of question items, it can be known with certainty which question items need to be maintained and which must be corrected by revision or discarded.[28]

Besides the questions must be valid, the questions must also meet the standards of reliability. The standard reliability value in the size of the value range of 0.8 – 1 which leads to a high-value measure of 0.6 – 0.8 is in the sufficient category, 0.4 – 0.6 is in the rather low category, 0.2 – 0.4 is the low category, and 0 – 0.2 is included in the very low category, meaning it shows unreliable. Reliability is the reliability of the test. A test that can be said to be reliable is that there are consistent results when measuring something that is measured.[29]

For measurements in the state's application related to reliability, this is measured by halving by comparing an even score (X) and an odd score (Y) using the Spearman-Brown formula. The state program used to measure question reliability data resulted in an average value of 9.50, a standard deviation of 3.56, a correlation of XY with a value of 0.63, and a test reliability of 0.78. So based on the results of the data, it is included in the category of reliability of the questions tested, including reliability on a sufficient scale.[30]

b. Differentiating Power

To measure the differentiating power section in the state's application, it is done by dividing the upper group and the lower group in answering the question correctly.[31] That is, the results of obtaining scores are graded from the highest answered correctly to the most correct. The function of differentiating power is to know and distinguish students who have a level of intelligence from the very and less. The way to calculate the distinguishing power in the state's application is by subtracting the upper group that answers correctly from the lower group that produces the correct answer in each question that has been done.[32] The differentiating power of Arabic questions by testing using the Anates application gets the following value results:[33],[34]

Table 6. The Distinguishing Power of the Question Item

No. Items	Top Group	Bottom Group	different	Index DP (%)
1	28	9	19	70.37
2	29	15	14	51.85
3	26	9	17	62.96
4	20	4	16	59.25
5	19	12	7	25.92
6	23	7	20	74.07
7	18	8	10	37.03
8	26	7	19	70.37
9	24	7	17	62.96
10	27	8	19	70.37
11	26	12	14	51.85
12	24	5	18	66.66
13	22	3	19	70.37

14	26	15	11	40.74
15	20	10	10	37.03

c. Difficulty Level

The level of difficulty in each question item is a number that indicates the difficulty or ease in the question item. The criteria for a good question are questions that fall into the category of not and also not too difficult, as what is meant is the level of difficulty of the question should be included in the medium criteria. Because, if a question contains risks too easily, it will cause student motivation to be lower. Vice versa, if a question contains criteria that are too difficult it will lead to despair for students because the problem they face is something beyond the reach of their reasoning power.[35]

Table 7. Difficulty Level Interpretation of Data

No. Item	Total Correct	Difficulty Level (%)	Category
1	66	65,34	Medium
2	89	88.12	It's easy
3	63	62,38	Medium
4	44	43.56	Medium
5	67	66.34	Medium
6	59	58.42	Medium
7	45	44.55	Medium
8	52	51.49	Medium
9	56	55.46	Medium
10	66	65.35	Medium
11	81	80.19	Easy
12	63	62.38	Medium
13	46	45.54	Medium
14	87	86.14	It's Easy
15	72	71.29	Easy

Based on the results of the analysis in Table 7, two questions can be obtained that are included in the very easy question criteria 2 and 4 (13.33%), two questions are included in the easy criteria, namely questions number 11 and 15 (13.33%), and 11 other questions are included in the medium criteria (73.33%). Therefore, based on the results of the analysis, question items number 2, 14, 11, and 15 need to be revised again so that the level of difficulty in the question items can be stabilized and then the remaining questions can still be maintained to be tested at the next stage. Therefore, it can be known based on the results that have been obtained that none of the questions tested are included in the criteria for difficult and very difficult questions.

d. Effectiveness of Deceivers

The distractor in the question item can function if the student chooses at least 5% of the total. The question is for students who do not understand the material well in the question items.[36] Meanwhile, based on other opinions question deceivers can function well if a question deceiver has great appeal to students or test takers who have not and do not understand the concept of a material or in this case lack of mastery of the material faced by them. In addition, the purpose of this is an effort to find out and weigh the unequal grasp of students with each other.[37]

Table 8. The Effectiveness of the Problem Deceiver

No. items	Multiple Choice				
	a	b	c	d	e
1	**6	+6	-15	+6	++9
2	++4	**87	+2	++3	++5
3	-3	-4	**64	-15	-15
4	**45	+9	+9	+10	-28
5	**68	-13	+5	-3	+12
6	-17	++9	**60	-5	++10
7	++15	**46	++11	-22	+7
8	++10	++10	**53	+18	++10
9	+7	++12	**57	++12	++13
10	-4	-1	**67	-23	+6
11	**86	-11	++5	-2	-1
12	+6	**62	+13	++10	++10
13	+21	+12	-6	++17	**45
14	-11	**86	+2	-1	-1
15	+4	++6	-12	++6	+4

Information: **: Answer key, ++ : Excellent, + : Good,
 -: not good enough _: bad —: very bad

From the results of the data in the table, it was found that some questions in numbers 4, 10, 11, and 14 had minus 2-3 which stated that the deceivers in the question items were very bad. So that from here it can be used as an evaluation for the next and revision material for the questions earlier so that it can be used again at the next stage of the exam.[38]

CONCLUSION AND IMPLICATIONS

Based on the discussion of the results of data analysis the question items in the thematic questions that have been applied and tested in MAN 2 Gorontalo Regency produce analysis results that the existing question items still contain a low level of thinking LOTS (low order thinking skill), therefore it is necessary to make improvements and revisions so that they can be tested again at the next stage. Especially based on the results of the analysis several questions are still included in the C3 criteria, namely questions number 6 and 15, C2 is in question number 14, and C1 is in questions number 7-12. Meanwhile, based on the quality measured through the state application, the XY correlation has a value of 0.63, the reliability of the test score is 0.78, and the level of difficulty is found in questions number 2, 14, 11, and 15 which still need to be corrected and revised. Therefore, based on the results obtained after going through several stages of analysis, shows that the thematic questions that have been tested are included in sufficient quality criteria in terms of quality.

Thematic questions based on HOTS are instruments that must be held to increase the reasoning power of students toward higher order thinking in educational institutions. Therefore, the question compiler must have sufficient basic competence and master the elements of HOTS in making HOTS-based questions as an instrument so that students can use a high level of reasoning power. Therefore, in the preparation of the problem, it is necessary to pay attention to all the elements needed to be analyzed objectively in terms of quality and quantity.

BIBLIOGRAPHY

- [1] A. N. Phafiandita, A. Permadani, A. S. Pradani, and M. I. Wahyudi, "Urgensi Evaluasi Pembelajaran di Kelas," *JIRA J. Inov. dan Ris. Akad.*, vol. 3, no. 2, pp. 111–121, Feb. 2022, <https://doi.org/10.47387/jira.v3i2.262>
- [2] S. d. Kadir and M. Arif, "Dinamika Perkembangan Kurikulum Bahasa Arab Di Indonesia Ditinjau Secara Historis," *Al-kilmah Pendidik. Bhs. Arab dan Hum.*, vol. 1, no. 2, 2022. <https://doi.org/10.52266/al-afidah.v2i2.586>
- [3] D. P. D. Santi, E. S. Praja, and S. Setiyani, "Peningkatan Kompetensi Guru : Membuat Soal HOTS dan Pengecoh Tipe Multiple Choice," *ABDIMASKU J. Pengabd. Masy.*, vol. 4, no. 2, p. 139, May 2021, <https://doi.org/10.33633/ja.v4i2.212>
- [4] S. D. Kadir and H. Hairuddin, "Analisis MaFuL Muthlaq Berdasarkan Makna Dalam Al-Qur'an Juz 29," *Assuthur J. Pendidik. Bhs. Arab*, vol. 1, no. 2, pp. 120–130, 2022, <https://doi.org/10.58194/as.v1i2.477>
- [5] M. Laili, "Ketepatan Kontruksi Butir Pilihan Ganda Bahasa Arab," *ALSUNYAT J. Penelit. Babasa, Sastra, dan Budaya Arab*, vol. 3, no. 2, pp. 111–124, 2020, <https://doi.org/10.17509/alsuniyat.v3i2.25272>
- [6] I. Rahmi Nur Fauziah, S. Syihabudin, and A. Sopian, "Analisis Kualitas Tes Bahasa Arab Berbasis Higher Order Thinking Skill (HOTS)," *لساننا J. Ilmu Bhs. Arab dan Pembelajarannya*, vol. 10, no. 1, p. 45, Aug. 2020, <https://doi.org/10.22373/ls.v10i1.7805>
- [7] D. Maulana and A. Sanusi, "Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Bersama Daerah (UAMBD) Madrasah Ibtidaiyah Tahun 2017-2018," *Ta'lim al-'Arabiyah J. Pendidik. Bhs. Arab Kebahasaaraban*, vol. 4, no. 1, pp. 12–24, Jun. 2020, <https://doi.org/10.15575/jpba.v4i1.8054>
- [8] F. A. Kafi and Z. Ilma, "Analisis Soal Tematik Bahasa Arab berbasis HOTS pada Kelas XII MAN 3 Jember," *Al-Fusha Arab. Lang. Educ. J.*, vol. 4, no. 1, pp. 1–9, Jan. 2022, doi: 10.36835/ALFUSHA.V4I1.729. <https://doi.org/10.62097/alfusha.v4i1.729>
- [9] S. D. Kadir, M. Jundi, S. A. Laubaha, I. Ali, M. Z. Adam, and Y. Kasan, "Korelasi Bahasa dan Pikiran dalam Perspektif Al-Qur'an," *Fikroh J. Pemikir. dan Pendidik. Islam*, vol. 15, no. 1, pp. 1–17, Jan. 2022, <https://doi.org/10.37812/fikroh.v15i1.294>
- [10] S. A. D. Febrianti, I. W. Widiani, and K. Y. Ana, "Higher-Order Thinking Skill (HOTS) Instrument-Based Cognitive Evaluation in Grade V Elementary School Students," *Think. Ski. Creat. J.*, vol. 4, no. 2, p. 48, Sep. 2021, <https://doi.org/10.23887/tscj.v4i2.38570>
- [11] I. L., "Evaluasi dalam Proses Pembelajaran," *Eval. Dalam Proses Pembelajaran*, vol. 9, no. 2, p. 344, 2019. <https://doi.org/10.35673/ajmpi.v9i2.427>
- [12] N. Azma and S. Mania, "Development of A Diagnostic Test in The Form of A Three Tier Essay Test to Identify Understanding of Mathematical Concepts on The Material of The System of Linear Equations of Three Variables Pengembangan Tes Diagnostik Berbentuk Three Tier Essay Test untuk Mengidentifikasi Pemahaman Konsep Matematika pada Materi Sistem Persamaan Linear Tiga Variabel," vol. 6, no. 1, pp. 1–23, 2024.
- [13] Y. F. Basri and E. M. Karima, "Kualitas Butir Soal Penilaian Tengah Semester Ganjil Mata Pelajaran Sejarah Terhadap Hasil Belajar Kognitif Siswa di Fase E SMK Negeri 1 Painan," *J. Pendidik. Tambusai*, vol. 7, no. 3, pp. 21160–21171, 2023.
- [14] S. Ansyoriah, E. Boeriswati, and S. S. Asep Supriyana, *Evaluasi Pembelajaran Bahasa*. Penerbit Adab, 2024.

- [15] M. A. Musa and R. Mutiah, "Analisis butir soal bahasa arab di mtsn kota parepare," pp. 11–23.
- [16] S. Y. Erlangga, E. A. Poort, P. H. Winingsih, O. Manasikana, and A. Dimas, "Meta-Analisis: Effect size Model Pembelajaran Berbasis Masalah pada Kemampuan Berpikir Tingkat Tinggi (HOTS) dan Pemahaman Konseptual Siswa dalam Fisika," *Compt. J. Ilm. Pendidik. Fis.*, vol. 9, no. 2, pp. 185–198, 2023, <https://doi.org/10.30738/cjipf.v9i2.15685>
- [17] S. Ammade, *Implementasi Model Pembelajaran Problem Based Learning (PBL) pada Keterampilan English Listening dengan Menggunakan Analisis STAR (Situation, Task Action, Result)*, no. January. 2024.
- [18] A. M. Aziz, F. Atmajaya, A. W. Yusuf, and A. Hermawan, "Pembelajaran Bahasa Arab Berbasis Higher Order Thinking Skills (Hots) Dalam Membentuk Pembelajar Yang Kompeten," *Al-Fakkaar*, vol. 5, no. 1, pp. 60–78, 2024, [Online]. Available: <https://ejournal.unisda.ac.id/index.php/ALF/article/view/5770>
- [19] A. S. P. Meliala, "Muatan HOTS Pada Latihan Soal Buku Teks Bahasa Arab Kelas VI Madrasah Ibtidaiyah Kemenag," *Ihya al-Arabiyah J. Pendidik. Bhs. dan Sastra Arab*, vol. 9, no. 2, p. 115, 2023, <https://doi.org/10.30821/ihya.v9i2.17901>
- [20] M. 1 K. Gorontalo, *Soal Tematik Bahasa Arab MAN 1 Kabupaten Gorontalo*. Kab. Gorontalo: Perpustakaan MAN 1 Kab. Gorontalo, 2021.
- [21] R. A. Nugroho, *HOTS (Kemampuan Berpikir Tingkat Tinggi: Konsep, Pembelajaran, Penilaian, dan Soal-soal)*. Jakarta: PT Gramedia Widiasarana Indonesia, 2018.
- [22] Ali Mufti, "Project-Based Learning untuk Meningkatkan Kemampuan Berpikir Tingkat Tinggi pada Mata Pelajaran Bahasa Arab," *Al-Ma'rifah*, vol. 19, no. 1, pp. 13–22, Apr. 2022, <https://doi.org/10.21009/almakrifah.19.01.02>
- [23] A. Muradi, F. Mubarak, R. Darmawaty, and A. R. Hakim, "Higher Order Thinking Skills Dalam Kompetensi Dasar Bahasa Arab," *Arab. J. Arab. Stud.*, vol. 5, no. 2, p. 177, 2020, <https://doi.org/10.24865/ajas.v5i2.293>
- [24] A. Fanani and D. Kusmaharti, "Pengembangan Pembelajaran Berbasis Hots (Higher Order Thinking Skill) di Sekolah Dasar Kelas V," *J. Pendidik. Dasar*, vol. 9, no. 1, pp. 1–11, May 2018, <https://doi.org/10.21009/10.21009/JPD.081>
- [25] N. H. Bin Nafsah and A. M. R. A. J. Manan, "Musykilāt Tadrīs Mahārah al-Kalām fī Ma'had Dār at-Tālibīn al-Qādiri 2 Gumukmas Jember," *Al-Fusha Arab. Lang. Educ. J.*, vol. 3, no. 1, pp. 21–32, Jan. 2021, <https://doi.org/10.36835/alfusha.v3i1.431>
- [26] A. Saefullah Kamali, "Pengaruh Soal Tipe Hots terhadap Kemampuan Berpikir Kreatif Siswa," *Cakrawala Pedagog.*, vol. 3, no. 2, pp. 128–131, Oct. 2019, <https://doi.org/10.51499/cp.v3i2.110>
- [27] A. M. Farida, "Validitas dan Reliabilitas dalam Analisis Butir Soal," *Al-Mu'arrib J. Pendidik. Bhs. Arab*, vol. I, no. 1, pp. 34–44, 2021. <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- [28] I. R. N. Hula, "Preferensi Abu Dawud Sulaiman Bin Najjah dalam Kaidah Rasm Al-Mushaf/Preference of Abu Dawud Sulaiman Bin Najjah In Rules of Rasm Al-Mushaf," *Diwan J. Bhs. dan Sastra Arab*, vol. 6, no. 2, 2020, [doi: https://doi.org/10.24252/diwan.v6i2.13969](https://doi.org/10.24252/diwan.v6i2.13969)
- [29] S. Hajaroh, "Perbandingan Reliabilitas Butir Soal Bahasa Arab Antara Pilihan Ganda dengan Menjodohkan pada Tes Buatan Guru," *eL-HIKMAH J. Kaji. dan Penelit. Pendidik. Islam*, vol. 10, no. 2, pp. 157–178, Dec. 2016, <https://doi.org/10.20414/elhikmah.v10i2.215>
- [30] I. R. N. Hula and Z. P. Mariana, Ana, "Pengembangan Hybrid Learning berbasis Aplikasi

- Computer Assisted Test (CAT) pada Program Arabic Proficiency Test,” *Tadbir J. Manaj. Pendidik. Islam*, vol. 9, no. 1, pp. 103–125, 2021, <https://doi.org/10.30603/tjmpi.v9i1.2063>
- [31] N. Hasaniyah *et al.*, “Development of Arabic Speaking Teaching Materials based on Tourism Anthropology: A Validator’s Assessment,” *Int. J. Membr. Sci. Technol.*, vol. 10, no. 2, pp. 1130–1143, Jun. 2023, <https://doi.org/10.15379/ijmst.v10i2.1161>
- [32] N. Fikriyah, “Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas VII Semester Ganjil SMP Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019/2020,” *Maharaat J. Pendidik. Bhs. Arab*, vol. 3, no. 2, pp. 128–140, Sep. 2021, <https://doi.org/10.18196/mht.v3i2.10501>
- [33] Muhammad Jundi, M. Z. Adam, I. Ali, S. D. Kadir, and S. A. Laubaha, “Himpunan Mahasiswa Jurusan dan Matrikulasi Bahasa Arab bagi Mahasiswa Baru,” *J. Naskhi J. Kaji. Pendidik. dan Bhs. Arab*, vol. 4, no. 1, pp. 1–9, Apr. 2022, <https://doi.org/10.47435/naskhi.v4i1.798>
- [34] S. A. Laubaha, L. G. Otaya, Z. Yasin, I. R. N. Hula, H. Hairuddin, and M. Z. Adam, “Pengembangan Bahan Ajar Mahāratul Kalām Berbasis Kearifan Lokal Provinsi Gorontalo,” *Al Mi’yar J. Ilm. Pembelajaran Bhs. Arab dan Kebahasaaraban*, vol. 7, no. 1, p. 459, 2024, <https://doi.org/10.35931/am.v7i1.3340>
- [35] D. Maulana and A. Sanusi, “Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Bersama Daerah (Uambd) Madrasah Ibtidaiyah Tahun 2017-2018,” *Ta’lim al-’Arabiyah J. Pendidik. Bhs. Arab Kebahasaaraban*, vol. 4, no. 1, pp. 12–24, Jun. 2020, <https://doi.org/10.15575/jpba.v4i1.8054>
- [36] I. R. N. Hula and A. Kasim, “Al-Qawaid al-Sittah dalam Rasm al-Mushaf (Six Rules of Rasm and the Uniqueness in Qur’an),” *’A Jamiy J. Bhs. dan Sastra Arab*, vol. 10, no. 2, p. 385, Sep. 2021, <https://doi.org/10.31314/ajamiy.10.2.385-418.2021>
- [37] A. Rama, R. R. Putra, Y. Huda, and R. Lapisa, “Pengembangan E-Modul Menggunakan Aplikasi Flip Pdf Professional Pada Mata Kuliah Analisis Kurikulum Pendidikan Dasar,” *JRTI (Jurnal Ris. Tindakan Indones.)*, vol. 7, no. 1, p. 42, Feb. 2022, <https://doi.org/10.29210/30031473000>
- [38] M. Jundi, Y. Kasan, and S. D. Kadir, “Perbandingan Ma’ful Mutlaq Bahasa Arab dan Cognate Object Bahasa Inggris serta Analisis Sintaksis Ma’ful Mutlaq pada Juz ’Amma,” *Lisanan Arab. J. Pendidik. Bhs. Arab*, vol. 5, no. 2, pp. 153–178, Dec. 2021, <https://doi.org/10.32699/liar.v5i2.2190>