

Non-Test Assessment Innovation Performance For Maharah Kalam Through Youtube In The Pandemic Era

R. Umi Baroroh¹, Nafisatun Nisa²

^{1,2}Sunan Kalijaga Islamic State University, Indonesia

umi.baroroh@uin-suka.ac.id¹, 20204021013@student.uin-suka.ac.id²

Abstract

This study aims to innovate the development of the mahârah kalâm assessment instrument in the pandemic era through youtube media and to analyze the instrument developed using the Rasch model analysis. This research uses the R&D (Research and Development) method. The instrument items made were 21 items developed from four characteristics: pronunciation, smoothness, vocabulary, grammar, and understanding, with a total of 31 students. The assessment developed is in the form of a rating scale with 1-4 criteria. Data analysis was carried out by Winstep software. Based on the results of this research from the output of the Winstep program, it can be seen that the instrument made has a good reliability value and can measure students' abilities. This is indicated by the item reliability score of 0.72, the person reliability value of 0.82, the Cronbach alpha value of 0.87, the MNSQ Infit and MNSQ Outfit values , respectively 1.01 and 1.00, the ZSRD infit 0.00 and Outfit ZSTD of -0.01. At the same time, the number of items that fit is 11 items and ten items that are misfits.

Keywords: Rasch Analysis Model; Fit; Reliabilitas; Non-Test Assessment; Pandemic

INTRODUCTION

Arabic is an important learning material taught at various levels of education, from elementary school to university. Learning Arabic continues to develop along with the times. In the current era, Arabic is still mandatory for schools or madrasas throughout Indonesia and is also an option for local content in various schools. In the pandemic era, learning experienced changes and shifts both in terms of fulfilling the learning method, the material, and the media used in the learning process, this was due to the demands for distance learning.

Distance learning has a significant impact on the education process in Indonesia. Among them have an impact on shifting the learning process to online through various digital applications.(Hasnawati, 2021) Arabic learning is also done online through various application media such as WhatsApp, Facebook, Instagram, Twitter, YouTube, and other application media. This is the impact and implications of technological developments in today's era.(Septantiningtyas, 2018)

The learning process does not only include the delivery of material but starts from designing the objectives, curriculum, content, or material to the assessment or evaluation of learning. In evaluating learning, a teacher is also required to be able to use learning media well in learning kalam during a pandemic (Abdulatif, 2021)

In this case, teachers can take advantage of various learning applications to train students' language skills. As research conducted by Sa'addah Fiddaraini, that online maharah kalam learning is carried out through zoom meetings, and it is quite effective but there are obstacles such as many students do not take part in learning activities because of traffic jams with weak signals. Then the learning evaluation is carried out by conducting conversations via voice notes on WhatsApp then students record the screen as evidence. After that, the teacher assesses the students' speaking ability in Arabic, starting from the intonation and the correctness of the pronunciation.(Fiddaraini & , Farid Permana, 2020) However, in this case, the author assumes that the methods and media used are less effective and less objective because the teacher does not know the expressions and faces of students in practicing smaharoh kalam, and this can be done by students by reading the text. Therefore, the researchers tried to innovate the development of Maharoh Kalam assessment through YouTube media, so that teachers can observe all aspects that must be mastered by students in Mahârah Kalâm.

However, in preparing the assessment criteria, various procedures must be considered to obtain a valid and objective test instrument. A good assessment process depends on a good test instrument as well. Assessment serves to evaluate the learning process that has been carried out. In addition, to determine student learning outcomes, level of ability and improve the quality of student learning. (Azizah & Wahyuningsih, 2020)

Dr. Abdul Munip in his book "Arabic learning assessment" states that learning assessment must be based on several principles, namely valid, educating, competency-oriented, fair and objective, open, sustainable, comprehensive, and meaningful.(Abdul Munip, 2017) Broadly speaking, learning kalam aims so that students can pronounce the sounds of letters and words in Arabic well, and according to grammatical rules, able to express their ideas orally. The type of assessment that is suitable for use in learning kalam is the type of non-test assessment. That is the type of assessment that does not use tests in its instrument, because non-test assessments are used to assess the process and results. Non-test assessment can be used to measure students' abilities in the cognitive, affective, and psychomotor domains.(Abdul Munip, 2017)

The author assumes that the non-test type of assessment is often based on an assessment that is less objective (tends to be subjective). Therefore, it is very necessary to take the right steps in the preparation of a valid and objective instrument that can measure well. From the results of the assessment, a score will be obtained where the score indicates the student's ability. However, this requires further analysis to get good quality questions, the quality of students, and the comparisons that are measured.(Boone et al., 2014) The approach that is widely used in the analysis of test results is classical test theory. Where the classic test only emphasizes the visible score of one exam, and the results obtained are concluded as a person's ability (ability) for the exam he is taking. While the level of student ability is directly proportional to the level of difficulty of the questions. The level

of success of students in working on the questions depends on the level of ability and the level of difficulty of the questions.(Sumintono, 2016)

In this case, the suitable analytical model used to identify instrument items is the Rasch model. The Rasch model was developed by George Rasch which is an analytical model of item response theory (Item Response Theory). Rasch theory develops a measurement model that determines the relationship between the level of student ability (personability) and the level of difficulty of the item (Item different). (Sumintono, 2016) Rasch is an analytical model that is used to reduce and minimize the existence of invalid and non-objective questions. The analysis using the Rasch model results in a statistical analysis of fit (fit statistics) which provides information to researchers whether the data obtained ideally illustrates that people who have high abilities provide patterns of answers to items according to their level of difficulty..(Azizah & Wahyuningsih, 2020)

The Rasch model uses logarithmic functions to produce measurements with equal intervals. Which will produce logit (log odds units), showing the relationship between student abilities depending on the level of difficulty of the questions. So the equation can be written as follows:

Probability of success	=	Respondent's ability	-	The item difficulty level
------------------------	---	----------------------	---	---------------------------

The advantage of the Rasch model compared to other analytical models is that the Rasch model is an analytical model that meets the criteria for an objective measurement concept in the social sciences and education. So that the quality of measuring social fields with the Rasch model analysis will have the same quality as measurements on the physical dimensions in the field of physics which always produce definite and objective measurement results.(Sumintono, 2016)

The Rasch model meets the criteria for the concept of objective measurement in the social sciences and education as stated by Mok and Wright, namely:

1. Gives a linear measure with equal intervals
2. Carry out the right estimation process
3. Finding items that are inappropriate (misfits) or uncommon (outliers)
4. Overcome lost data
5. Produces replicable measurements (independent of the parameters studied).(Sumintono, 2016)

From the results of the analysis of the Rasch model, it can be seen that first, the scores obtained are in the form of a scale with equal and linear intervals derived from ratio data and not in the form of raw score data. Second, the Rasch model produces a process of estimating a person's ability or the level of difficulty of the questions that will have a more precise estimation value and can be compared with each other because they have the same logit unit. Third, being able to find out the order of questions from easy to difficult, the inaccuracy/consistency of answers from respondents, fourth, by ranking students' abilities and the level of difficulty of the questions will be known if there is missing data, fifth, from the logit scale,

will bring up a value that depends on the pattern of responses given so that the measurement is independent. (Sumintono, 2016)

METHOD

The research method used is the R&D (Research and Development) method. The model developed is the Liu model. The steps of Liu's model in developing a test instrument using the Rasch model are: a. define constructs that can be characterized by linear properties, namely by identifying basic competencies, b. identify abilities according to different levels of a predetermined construct, c. determine the resulting space of the students' abilities, tested with a representative sample of the population, e. applying the Rasch model, f. review items according to statistics and revise if necessary, g. develop documentation for the instrument. (Ibnu et al., 2019)

Data analysis was carried out quantitatively using the Winsteps program to obtain item parameters that fit the Rasch model. The size of the item reliability determines the reliability, and the overall reliability value of the item is indicated by the magnitude of Cronbach's alpha value. Then there is Outfit MNSQ, Outfit ZSRD, and the correlation value of the item with the question as a whole which shows the limit of items that are declared fit with the Rasch model. Items that are declared fit with the Rasch model if the Outfit MNSQ value is between 0.5 to 1.5, Outfit ZSRD has a value between -2.0 to 2.0, and the correlation value of the item with a total score is between 0.4 to 0.85. (Azizah & Wahyuningsih, 2020)

RESULTS AND DISCUSSION

The test assessment technique used in speaking skills (*mahârah kalâm*) is a non-test type of performance technique. The test instrument developed in this non-test technique is in the form of a rating scale. The rating scale is used to assess respondents based on references to certain characteristics and to determine the abilities and characteristics of respondents who are varied and multilevel. (Suparman, 2016) With this assessment, the assessor gives value to the mastery of the students' speaking competence/skills based on the specified scale. The rating scale ranges from 1-4 criteria, with a scale of 1 = very bad, 2 = not very good, 3 = quite good, 4 = very good. (Sugiono, 2015)

The number of instruments developed is as many as 225 questionns. The respondents were PBA S1 Semester 5 students with a total of 31 students. The item of the instrument was developed referring to the indicator of mastery of maharah kalam. Following are the indicators: (Abdul Munip, 2017)

Table 1.1 Indicators of the maharah kalam instrument

Aspects of Pronunciation (<i>an Nuthq</i>)	Students are able to pronounce the sounds of Arabic letters fluently
	Students are able to pronounce the sounds of letters that are similar and close together correctly
	Students are able to distinguish the pronunciation of long and short harokat
	Students are able to pronounce clearly (<i>jahr</i>)
	Students are able to pronounce Arabic words with the right intonation

Smoothness (<i>as sur'ah</i>)	Students are able to express their thoughts and words fluently Students are able to think and speak spontaneously using Arabic Students convey Arabic sentences without memorizing
Vocabulary (<i>al-mufradah</i>)	Students have a wealth of vocabulary in Arabic Students can use various vocabulary according to the current context Students use some Arabic cultural expressions obtained from watching videos
Grammar (<i>al-qowaid</i>)	Students are able to express their thoughts by using the right nahwiyah structure Students are able to apply mudzkar muannats Arabic grammar correctly Students are able to apply Arabic grammar mufrod, tasniyha, plural Students are able to apply Arabic grammar adverbs (<i>hâl, tamyîz, 'adad, harf jar</i>) Students are able to apply Arabic grammar I'rob sentences Students are able to apply Arabic grammar in the form of words (<i>tashrif</i>)
Understanding	Students are able to convey the core of the discussion from the video they have watched Students can re-communicate sentences The sentences conveyed can be understood by the listener Students can express the words that are listened to in other sentences correctly

Analysis of the Rasch model with Winstep Software on the Maharah Kalam Assessment Instrument

The results of the analysis of 21 items with the number of respondents as many as 31 students obtained the following values:

1. Person Measure (0.86) means that it is greater than the item measure logit value (0.0) so that it shows the tendency of respondents to have good criteria on statements in various items.
2. The Cronbach alpha value obtained is (0.87), meaning that the interaction between the person and the item as a whole is reliable with a very good reliable predicate.
< 0.5: Poor 0.5–0.6: Poor 0.6–0.7: Fair 0.7–0.8: Good > 0.8: Very good
Then the reliable predicate 0.87 is quite good.
3. Value of Person Reliability and Item Reliability
< 0.67: Weak, 0.67 – 0.80: Fair, 0.81 – 0.90: Good 0.91 – 0.94: Very good > 0.94: Excellent. (Sumintono & Widhiarso, 2013)

From the results of the analysis, the value of person reliability is 0.87 and item reliability is 0.72, it can be concluded that the consistency of the answers/reliability level of the respondents is quite good and the quality of the items/reliability level of the items is quite good.

4. The value of InfitMean Square Statistic (Infit MNSQ) and OutfitMean Square Statistic (Outfit MNSQ) on the table person average value is 1.01 and 1.00 respectively. While the ideal value is 1.00 where the closer to 1.00, the better.(Sumintono & Widhiarso, 2013) While the Infit ZSTD and Outfit ZSTD values have an average value in the table person respectively, namely 0.00 and -0.01, and the table items respectively -0.07 and 0.01. Where ideally is 0.0

where if it is close to 0.0 then the quality is getting better. So that the instrument made is by the Rasch model and can be used as a measuring tool for maharah kalam.

- The Separation Index is used to calculate the strata of the item difficulty level which is useful for identifying groups of item difficulty levels. The greater the value of separation indicates the quality of the instrument, the more equations used are the formula:

$$H = \frac{[(4 \times \text{separation}) + 1]}{3}$$

The separation value obtained is 2.54, so $H = [(4 \times 2.54) + 1] / 3 = 3.72$ is rounded up to 4.

Table. 1.2 Summary Statistic

Criteria	Result Value	Predicate
Measure person	0,86	Good
Measure Item	0,0	Good
Alpha Cronbach	0,87	Very good
Infit MNSQ	1,01	Good
Outfit MNSQ	1,00	Good
Infit ZSTD	0,00	Good
Outfit ZSTD	-0,01	Good
Person Reliability	0,87	Good
Item Reliability	0,72	Quite Good
Separation	3,72	Good

SUMMARY OF 31 MEASURED PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	59.8	21.0	.86	.41	1.01	.00	1.00	-.01
SEM	1.3	.0	.23	.01	.07	.22	.07	.22
P. SD	6.9	.0	1.23	.04	.39	1.20	.40	1.20
S. SD	7.0	.0	1.25	.04	.40	1.22	.41	1.22
MAX.	75.0	21.0	3.74	.45	1.86	2.11	1.87	2.24
MIN.	47.0	21.0	-1.11	.35	.29	-2.45	.28	-2.45
REAL RMSE	.45	TRUE SD	1.15	SEPARATION	2.54	PERSON RELIABILITY	.87	
MODEL RMSE	.42	TRUE SD	1.16	SEPARATION	2.79	PERSON RELIABILITY	.89	
S. E. OF PERSON MEAN = .23								

Figure 1.1 Person Measure

SUMMARY OF 21 MEASURED ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	88.3	31.0	.00	.34	.98	-.07	1.00	.01
SEM	1.3	.0	.15	.00	.06	.24	.06	.24
P.SD	5.9	.0	.67	.02	.28	1.07	.28	1.07
S.SD	6.1	.0	.69	.02	.28	1.10	.29	1.09
MAX.	100.0	31.0	1.48	.37	1.50	1.72	1.54	1.90
MIN.	74.0	31.0	-1.45	.30	.46	-2.37	.46	-2.30
REAL RMSE	.35	TRUE SD	.57	SEPARATION	1.61	ITEM	RELIABILITY	.72
MODEL RMSE	.34	TRUE SD	.58	SEPARATION	1.72	ITEM	RELIABILITY	.75
S.E. OF ITEM MEAN = .15								

Figure 1.2 Item Measure

Analysis Of The Characteristics Of The Items And The Ability Of Respondents

Based on the results of the analysis through winstep, providing information from both the items and the respondents have differences in each item and student. As stated by Arikunto quoted by Ibnu et al, that a question that can be answered by smart students and students who are not smart is a form of the question that is not good because there is no differentiator. (Ibnu et al., 2019) From the winstep data, it is known that the distribution of student abilities and the level of difficulty of the questions are known. Students who can answer difficult questions have a high level of ability. The results of the analysis show that several students have high abilities, or have good criteria for money kalam proficiency, namely students with codes 005P, 0010P, 0012L, 0025L with a logit value of +2. In addition, the data shows students who have intermediate abilities with a logit value of +1, they are students with codes 007L, 0011P, 0014P, 0013P, 0017P, 004P, 0026P, 002P, 003P, 006P, 009P. The students who are classified as having low abilities are if they reach the logit-2 value. From the data, there are no students who are classified as low. However, 2 students achieved a logit value of -1, namely students with codes 0024P and 0027L. The data is shown in Figure 1.3.

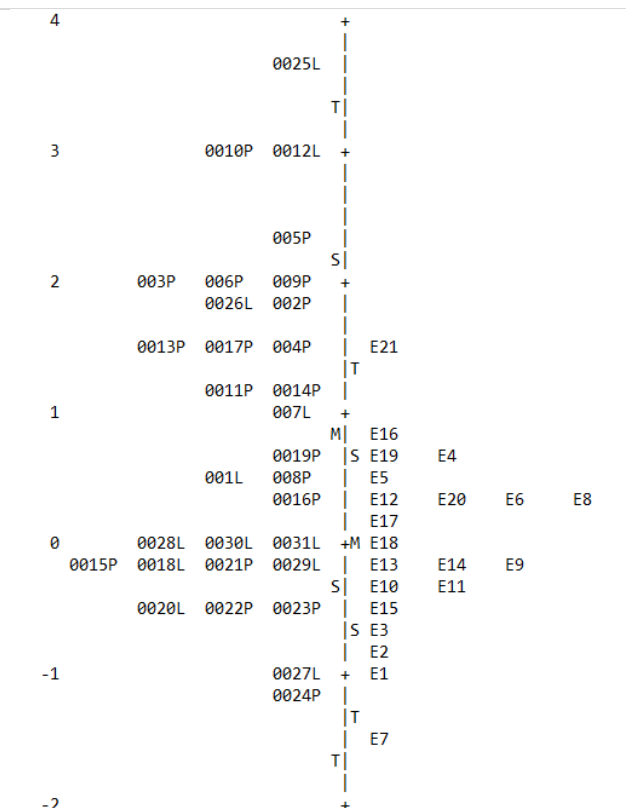


Figure 1.3 Person Map Item

From Figure 1.3, it can also be seen that the items have a high level of difficulty, which is shown in question code E21. And the item with the lowest level of difficulty is with a logit value below -1 which is indicated by the question code E7. This means that the instrument items in logit -1 need to be revised and improved again to fit the Rasch model. While the items that are between logit -0 to +0 are items in the medium difficulty level.

Instrument Reliability

Table 1.2 Instrument Reliability

Variable	Reliability	Separation
Item	0,72	2,48
Person	0,87	3,72

From the results of the analysis obtained data that the reliable value of the items is good, namely > 0.8 . In addition to this, it is said that the greater the value of separation, the better the quality of the questions. From the results obtained, it is known that the real value obtained is good and can identify groups of questions and groups of respondents.

Item Validity

Through winstep analysis, it can be seen whether all the items of the instrument are fit or not. The items in the research are said to be fit if the MNSQ

output is in the range of values from 0.5 to 1.5. Of the 21 items analyzed, 18 items are fit and 3 items that are not. The items that do not fit are items with codes E21, E6, and E10 with Outfit MNSQ values, respectively, namely 1, 54, 1.53, and .46 (out of the ideal limit). While the criteria for the suitability of the items that are fit that can be accepted are those that have an Outfit ZSTD value between -2.0 to +2.0. The data that shows the items that do not fit according to the Outfit ZSTD value are items E10 with a value of -2.30.

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
21	74	31	1.48	.30	1.43	1.66	1.54	1.90	A .69	.58	48.4	61.2	E21
6	86	31	.29	.33	1.50	1.72	1.53	1.70	B .71	.52	61.3	63.6	E6
7	100	31	-1.45	.37	1.40	1.39	1.31	1.09	C .74	.52	64.5	70.0	E7
13	90	31	-.16	.34	1.34	1.21	1.28	1.00	D .57	.51	64.5	67.8	E13
5	84	31	-.50	.32	1.11	.51	1.20	.77	E .17	.53	51.6	63.1	E5
17	87	31	.18	.33	1.10	.44	1.16	.64	F .36	.52	54.8	65.4	E17
18	89	31	-.05	.34	1.11	.48	1.12	.50	G .54	.51	67.7	67.2	E18
8	86	31	.29	.33	1.09	.42	1.08	.38	H .73	.52	58.1	63.6	E8
20	85	31	.40	.33	1.02	.16	1.08	.38	I .73	.53	58.1	63.7	E20
19	82	31	.71	.32	.95	-.13	1.06	.32	J .45	.54	64.5	63.5	E19
12	86	31	.29	.33	.91	-.25	1.01	.12	K .63	.52	58.1	63.6	E12
4	83	31	.61	.32	1.00	.08	.99	.06	J .15	.54	54.8	63.4	E4
16	80	31	.91	.31	.98	.00	1.00	.09	I .72	.55	61.3	62.7	E16
2	96	31	-.91	.36	.94	-.11	.95	-.07	H .26	.51	74.2	70.5	E2
3	94	31	-.65	.36	.81	-.63	.85	-.48	G .39	.51	74.2	70.0	E3
9	90	31	-.16	.34	.81	-.65	.83	-.54	F .13	.51	74.2	67.8	E9
14	90	31	-.16	.34	.77	-.81	.80	-.65	E .73	.51	67.7	67.8	E14
15	93	31	-.53	.35	.67	-1.25	.64	-1.39	D .70	.51	74.2	69.2	E15
1	97	31	-1.04	.36	.62	-1.51	.62	-1.46	C .46	.51	80.6	70.6	E1
11	91	31	-.28	.35	.56	-1.81	.55	-1.81	B .32	.51	83.9	68.2	E11
10	91	31	-.28	.35	.46	-2.37	.46	-2.30	A .19	.51	90.3	68.2	E10
MEAN	88.3	31.0	.00	.34	.98	-.07	1.00	.01			66.1	66.2	
P. SD	5.9	.0	.67	.02	.28	1.07	.28	1.07			10.7	2.9	

Figure 1.3 Item Statistic Misfit

The item is said to be fit if it has a correlation point measure in the Rasch model or referred to as (biserial points) in classical theory which shows the difference between items. The items received are items that have scores ranging from 0.4 to 0.85. The items that meet the point measure correlation value criteria are 13 items and 8 items that do not meet (> 0.4), namely items E9, E4, E5, E10, E2, E11, E17, E3. The following shows the correlation point measure data in Figure 1.4 below:

INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
.81	-.65	.83	-.54	.13	.51	74.2	67.8	E9
1.00	.08	.99	.06	.15	.54	54.8	63.4	E4
1.11	.51	1.20	.77	.17	.53	51.6	63.1	E5
.46	-2.37	.46	-2.30	.19	.51	90.3	68.2	E10
.94	-.11	.95	-.07	.26	.51	74.2	70.5	E2
.56	-1.81	.55	-1.81	.32	.51	83.9	68.2	E11
1.10	.44	1.16	.64	.36	.52	54.8	65.4	E17
.81	-.63	.85	-.48	.39	.51	74.2	70.0	E3
.95	-.13	1.06	.32	.45	.54	64.5	63.5	E19
.62	-1.51	.62	-1.46	.46	.51	80.6	70.6	E1
1.11	.48	1.12	.50	.54	.51	67.7	67.2	E18
1.34	1.21	1.28	1.00	.57	.51	64.5	67.8	E13
.91	-.25	1.01	.12	.63	.52	58.1	63.6	E12
1.43	1.66	1.54	1.90	.69	.58	48.4	61.2	E21
.67	-1.25	.64	-1.39	.70	.51	74.2	69.2	E15
1.50	1.72	1.53	1.70	.71	.52	61.3	63.6	E6
.98	.00	1.00	.09	.72	.55	61.3	62.7	E16
1.09	.42	1.08	.38	.73	.52	58.1	63.6	E8
.77	-.81	.80	-.65	.73	.51	67.7	67.8	E14
1.02	.16	1.08	.38	.73	.53	58.1	63.7	E20
1.40	1.39	1.31	1.09	.74	.52	64.5	70.0	E7

Figure 1.4 Poin Measure Correlation

The following is a summary table that shows the items that fit according to the Rasch model:

Table 1.3 Summary of questions

No.	Question Code	Description
1	E1	<i>Fit</i>
2	E2	<i>Misfit</i>
3	E3	<i>Misfit</i>
4	E4	<i>Misfit</i>
5	E5	<i>Misfit</i>
6	E6	<i>Misfit</i>
7	E7	<i>Fit</i>
8	E8	<i>Fit</i>
9	E9	<i>Misfit</i>
10	E10	<i>Misfit</i>
11	E11	<i>Misfit</i>
12	E12	<i>Fit</i>
13	E13	<i>Fit</i>
14	E14	<i>Fit</i>
15	E15	<i>Fit</i>
16	E16	<i>Fit</i>
17	E17	<i>Misfit</i>
18	E18	<i>Fit</i>
19	E19	<i>Fit</i>
20	E20	<i>Fit</i>
21	E21	<i>Misfit</i>

From that table can be known that the youtube online learning-based kalam assessment instrument that fit according Rasch model is the following indicators,

those are students are able to pronounce the sounds of Arabic letters fluently, students are able to think and speak spontaneously using Arabic, students speak Arabic without memorizing, students are able to express their thoughts by using the right nahwiyah structure, students are able to apply mudzkar muannats Arabic grammar correctly, students are able to apply Arabic grammar mufrod, tasniyha, plural, students are able to apply Arabic grammar adverbs (*hâl, tamyiz, 'adad, harf jar*), students are able to apply Arabic grammar i'rab sentences, students are able to convey the core of the discussion from the video they have watched, students can re-communicate sentences, the sentences conveyed can be understood by the listener.

The items that misfit is following the indicators, those are students are able to pronounce the sounds of letters that are similar and close together correctly, students are able to distinguish the pronunciation of long and short harokat, students are able to pronounce clearly (*jahr*), students are able to pronounce Arabic words with the right intonation, students are able to express their thoughts and words fluently, students have a wealth of vocabulary in Arabic, students can use various vocabulary according to the current context, students use some Arabic cultural expressions obtained from watching videos students have a wealth of vocabulary in Arabic, students can use various vocabulary according to the current context, students use some Arabic cultural expressions obtained from watching videos, students are able to apply Arabic grammar in the form of words (*tashrîf*). students can express the words that are listened to in other sentences correctly

CONCLUSION

Based on the results of research with the analysis of the Rasch model on the Maharah Kalam assessment instrument, it can be seen that the instrument made has a good reliability value and can measure students' abilities. This is indicated by the item reliability score of 0.72, the person reliability value of 0.82, the Cronbach alpha value of 0.87, the MNSQ Infit and MNSQ Outfit values respectively 1.01 and 1.00, and the ZSRD infit 0. ,00 and Outfit ZSTD of. -0.01. While the number of items that fit according to the Rasch model, 11 questions are fit and 10 questions that are misfit. This means that 11 questions are by the Rasch model, and 10 items need to be revised and improved so that they can be used as measuring tools that are by the Rasch model.

REFERENCES

- Abdul Munip. (2017). *Penilaian Pembelajaran Bahasa Arab*. Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri (UIN) Sunan Kalijaga Yogyakarta.
- Abdulatif, S. (2021). Dampak Pandemi Terhadap Eksistensi Pendidikan di Era Digital. *Jurnal Pendidikan Tambusai*, 5 (1), 1567–1570.
- Azizah, A., & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *JUPITEK: Jurnal Pendidikan Matematika*, 3(1), 45–50.

<https://doi.org/10.30598/jupitekvol3iss1pp45-50>

- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rating Scale Surveys. In *Rasch Analysis in the Human Sciences*. https://doi.org/10.1007/978-94-007-6857-4_2
- Fiddaraini, S., & Farid Permana, D. A. R. (n.d.). (2020). Pembelajaran Maharah Kalam Kelas Daring Komunitas Cangkal Bahasa Arab Sa'adah. *Proceeding NATHLA: Al-Nadwah Al-'Alamiyyah Fi Ta'Lim Al-Lughah Al-'Arabiyyah (International Conference on Arabic Language Teaching)*, 1(1), 109–116.
- Hasnawati. (2021). Aktivitas Digital: Dampak Penerapan Pembelajaran Daring. *Jurnal Literasi Digital*, 1(3), 148–154.
- Ibnu, M., Indriyani, B., Inayatullah, H., & Guntara, Y. (2019). Aplikasi Rasch Model: Pengembangan Instrumen Tes Untuk Mengukur Miskonsepsi Mahasiswa Pada Materi Mekanika. *Prosiding Seminar Nasional Pendidikan FKIP*, 2(1), 205–210.
- Septantingtyas, N. (2018). Pengaruh Pembelajaran Jarak Jauh Dengan Aplikasi Google Class Terhadap Hasil Belajar Mahasiswa. *Edureligia; Jurnal Pendidikan Agama Islam*, 2(2), 131–135. <https://doi.org/10.33650/edureligia.v2i2.714>
- Sugiono. (2015). *Metode Penelitian dan Pengembangan (Research and Development)*. ALFABETA.
- Sumintono, B. (2016). Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (Assessment For Learning). *Makalah Dipresentasikan Dalam Kuliah Umum Pada Jurusan Statistika, Institut Teknologi Sepuluh November, Surabaya, 17 Maret 2016., March*, 1–19. http://eprints.um.edu.my/15876/1/ITS_rasch_model_asesment_for_learning.pdf
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. Trim Komunikata Publishing House.
- Suparman, U. (2016). *Penilaian dalam pembelajaran bahasa*. Suluh Media.