JRERE 11 (2) 2022 166-177

Journal of Research and Educational Research Evaluation

http://journal.unnes.ac.id/sju/index.php/jere

The Analysis of Question Item on Tarikh Subject with Rasch Model Theory

Joko Subando[1]✉, Minhayati Saleh[2], Ngatimin Ngatimin[3]

[1]Islamic Institute of Mamba'ul 'Ulum Surakarta, Indonesia
[2]State Islamic University of Walisongo Semarang, Indonesia
[3]MIPK Muhammadiyah Kenteng Boyolali, Indonesia

Article Info

History Articles
Received:
16 May 2022
Accepted:
19 June 2022
Published:
30 August 2022

Keywords:
Analisis, Tarikh, Rasch Model

Abstract

This research has a purpose to find out the level of students' ability to work on the questions, the suitability of the individuals with the items, the level of difficulty of the items, the suitability of the items, and the interaction of individuals with the items by using the Rasch model theory. The data collection technique is with documentation, the data collected is in the form of answers from VII grader students of Madrasah Quraniyah Karanganyar taken from the end of the odd semester exam for the 2021/2022 school year. The number of respondents is 23 students and the number of questions is 25 items. The data are analyzed by using the Rasch model theory with the help of the Winstep program. The research result shows that (1) there are 2 students with high abilities (ex11, ex19) and 3 students with low abilities (ex3, ex9, ex12); (2) students who do not fit the model or have misconceptions about the questions are ex16, ex12, ex22, ex09, ex15, and ex21; (3) the most difficult questions are considered to be 2 items (B12, B29), the others are categorized as a medium; (4) the items that do not fit the model are B12, B20 and B24. Research finding generally shows that the test takers' abilities are above the item difficulty level, so the questions must be improved further.

✉Correspondence Address :
Post Graduate, Islamic Institute of Mamba'ul 'Ulum Surakarta,
Indonesia, Jalan Sadewa No 14 Serengan Surakarta
E-mail : jokosubando@iimsurakarta.ac.id

p-ISSN 2252-6420
e-ISSN 2503-1732

INTRODUCTION

According to the National Education System Constitution number 20 the year 2003, teachers are professional educators with the main task of educating, teaching, guiding, directing, training, assessing, and evaluating students in early childhood education through formal education, basic education, and secondary education. In carrying out their professional duties, teachers are obliged to plan the lessons, implement a quality learning process, and assess and evaluate the learning outcomes of the Constitution (2017).

According to Riadi (2018), in carrying out the evaluation process, teachers must be able to measure the competencies that have been achieved by students from each learning process or after several lesson units, so that teachers can make a decision on the students, whether there is a need for improvement or remedial and determine the next lesson plan both in terms of material and strategic plans.

Therefore, teachers are at least able to compile test and non-test instruments, able to make decisions for the position of their students, whether the expectation of optimal mastery has been achieved or not. The skills that must be possessed by the teachers then become a routine activity, namely making tests, taking measurements, and evaluating the competencies of their students so that they can determine further learning policies.

Meanwhile, teachers at Madrasah Quraniyah Al Husnayain Surakarta have participated in an evaluation study of the Islamic Religious Education (PAI) learning system. Every time there are mid-semester examinations and end-of-semester assessments, teachers are required to prepare exam questions. However, their understanding of the test preparation instruments has not yet been measured, especially in item analysis (interview with Sri Mulyani, 12/3/2022). An understanding of item analysis is an important thing in assembling questions so that the package of questions is well prepared and able to measure what should be measured.

Two major theories are often used for question item analysis, namely classical test theory and modern test theory or the Rasch model (Aziz, 2015; Mardapi, 1998; (Aziz, 2015; Mardapi, 1998; Triono, Sarno dan Sungkono, 2020). The modern test theory is developed because of the weakness in the classical test theory which states that the level of question item difficulty is highly dependent on the respondent when calibrating the items. When the calibration is done and it turns out that students have high ability, the level of question item difficulty will be low and when students have low ability calibration, the question item difficulty level will be high (Sumintono, 2018). The Rasch model theory can cover these shortcomings. It means that the level of difficulty and the discriminating power of the questions do not depend on the respondents during calibration, so the Rasch model theory is recommended in question item analysis and the development of question banks (Suyata, Mardapi, Kartowagiran, & Retnawati, 2011).

Based on the literature search, there have been many studies related to question item analysis using classical test theory and the Rasch model. (Setianingsih, 2018) conducted a test analysis to get an idea of the quality of the test, both the overall quality of the test and the quality of each question item. The characteristics of the test quality are measured based on good discriminatory power, distractor function, and a reasonable level of difficulty. This research is considered to be quantitative research. The locations of this research are at some MTs Ma'arif NU Kemranjen, namely MTs Ma'arif NU 2 Kemranjen, MTs Ma'arif NU 3 Kemranjen, and MTs Ma'arif NU 4 Kemranjen of Banyumas Regency. The object of research is the final exam sheet for the second semester of eight graders in the subject of Islamic Cultural History and student answer sheets. Data collection methods that are used in this research consist of observation, interview, and documentation. The result shows that the 50 question items from the second-semester final exam of the Islamic Cultural History subject which are followed by 191 students at three MTs Ma'arif NU Kemranjen Banyumas Regency is concluded to be not good. Based on the analysis carried out from the level of difficulty, discriminating power, and distractor function, it can be seen that the level of difficulty includes categories of very difficult in 2 questions, difficult in 5 questions, moderate in 33 questions, easy in 5 questions, and very easy in 5 questions. Distinguishing power includes very good in 4 questions, good in 22 questions, enough in 12 questions, bad in 10 questions, and very bad in 2 questions. Judging from the distractor function, it is found that the distractor function with the criteria of functioning is 33. There are 13 question items of which only two of the distractors function well. There are 2 question items in which only one distractor works. There are 2 question items in which the distractor function does not work.

Another interesting thing is the result of research (Fernanda & Hidayah, 2020) which conducts question item analysis by using classical test theory and the Rasch model. The result of the research shows that the analysis which uses the Rasch model is better than the classical test theory. The advantages are the ability of the Rasch model theory in detecting questions that are not answered by the students, its ability to detect the presence of DIF, and so on. This research means to strengthen the statement of Suyata, Mardapi, Kartowagiran, & Retnawati, (2011) as previously stated.

Ramadhanti, Rahmatullah, Wilujeng, and Chusna (2021) conduct a study to determine the ability of students to solve ICT literacy problems by defining indicators, accessing, managing, combining, evaluating, creating, and communicating. The test is conducted on 24 students spread across various classes at SMP Negeri 2 Fakfak. The instrument that is used in this research consists of 20 multiple choice questions. This research belongs to a descriptive study using the Rasch model. The result shows that the average score of students in solving decisive problems is 49, 27 in accessing, 18 in managing, 35 in combining, 20 in evaluating,11 in creating, and 15 in communicating with an overall average of 21.71 ± 14.62. Rasch model shows that students 01L, 02L, and 18P have high intelligence while students with the lowest abilities are coded as 07P, 12P, 09P, 08P, 05P, 13P, 14P, 16P, 19P, 21P, and 24P. The result shows that ICT literacy skills are still low, especially in the indicator of creation.

However, in general, the question item analysis that uses the Rasch model in some research above only discusses the aspect of the student's ability level. It does not comprehensively discuss both the compatibility of the person with the compatibility of the question items and the map of the person with the items which become the advantages of the Rasch model so further research is needed (Aziz, 2015; Wahyudi, Setyowati, & Partini, 2020). There are several reasons why this research is important to do (1) the availability of information on the suitability of the person with the question is useful to find out whether there are students' misconceptions about the questions. If there is a misconception, it can be improved by using corrective learning. Reason number (2) is about the existence of information on the suitability of question items that is useful for detecting whether students answer correctly, guess, or work on questions by using their knowledge. Reason number (3) is based on literature searches and field preliminary research that show there has been no question item analysis which is conducted at the Al Husnayain Islamic School. Thus, this research is strongly needed. Reason number (4) shows that Al Husnayain’s teachers are more focused on the good learning process but are not followed a good evaluation plan, especially the preparation of a quality test question package. Whereas the forms of quality questions are useful in measuring and obtaining real learning outcomes. The learning outcome will be a reflection of the learning process and evaluation materials for learning. Reason number (5) shows that this research is important to provide a complete perspective on the application of the Rasch model theory in item analysis. This research aims to determine the ability level of the students at Madrasah Quraniyah Al Husnayain in working on Tarikh questions, the suitability of students with the question items, the level of difficulty of the Tarikh question items, the suitability of the Tarikh question items, and the interaction map of students' abilities with the question items.

METHODS

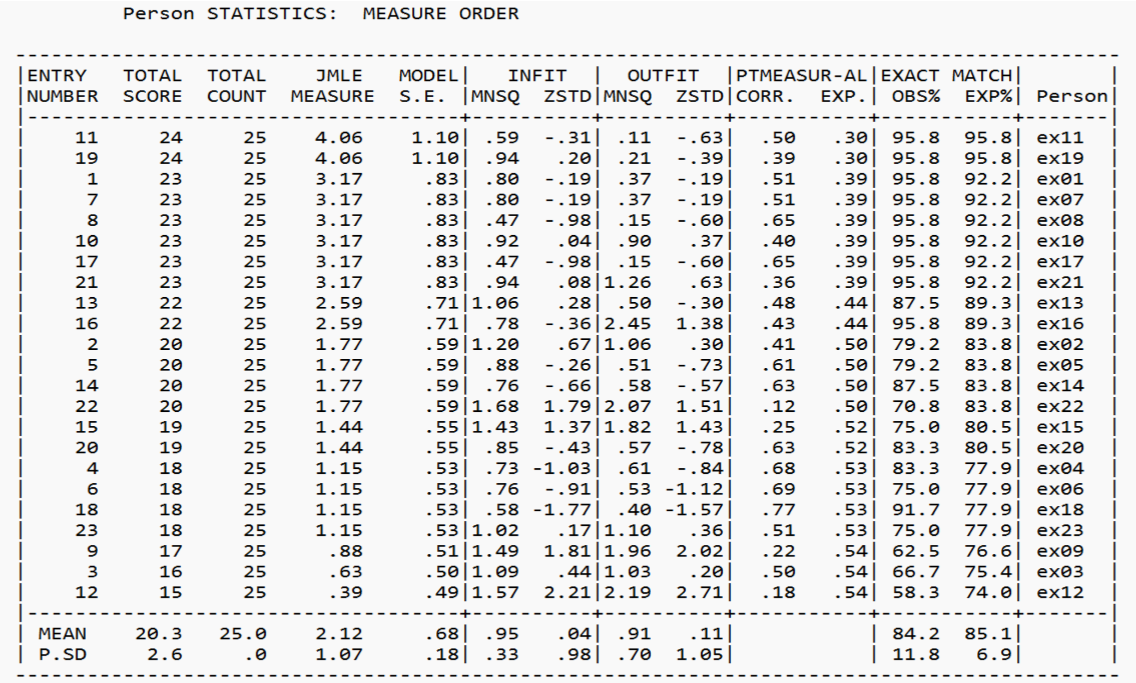
This research is categorized as descriptive research with a quantitative approach. The purpose of this research is to determine the characteristics of the question items, the characteristics of the examinees, and the interaction between the two. Research respondents are 23 students of seventh grader class at Madrasah Quraniyah Al Husnayain Karanganyar. The research data is in the form of student answers on the subject of Tarikh at the end of the odd semester final exam for the 2021/2022 school year which is taken by using documentation techniques. The number of multiple-choice questions which is analyzed consists of 25 items. The data that have been collected are analyzed by using the Rasch model theory with the help of the Winstep program.

RESULTS AND DISCUSSION

Individual Ability Level

The individual ability level in table 1 is shown in the JMLE measure column in the logit unit. Logit describes the ability of examinees which has a relationship that the greater the logit value, the higher the ability of the examinees, and this will be positively correlated with the total score, namely the correct answers from the examinees (H Untary, Risdianto, & Kusen, 2020). Meanwhile, the total count in the table column shows the number of questions that must be done by the examinees.

Table 1. Individual Ability



```
         Person STATISTICS:  MEASURE ORDER

--------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL    JMLE   MODEL|   INFIT  |  OUTFIT  |PTMEASUR-AL|EXACT MATCH|       |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Person|
|------------------------------------+----------+----------+-----------+-----------+-------|
|    11     24     25     4.06   1.10| .59  -.31| .11  -.63|  .50   .30| 95.8  95.8| ex11  |
|    19     24     25     4.06   1.10| .94   .20| .21  -.39|  .39   .30| 95.8  95.8| ex19  |
|     1     23     25     3.17    .83| .80  -.19| .37  -.19|  .51   .39| 95.8  92.2| ex01  |
|     7     23     25     3.17    .83| .80  -.19| .37  -.19|  .51   .39| 95.8  92.2| ex07  |
|     8     23     25     3.17    .83| .47  -.98| .15  -.60|  .65   .39| 95.8  92.2| ex08  |
|    10     23     25     3.17    .83| .92   .04| .90   .37|  .40   .39| 95.8  92.2| ex10  |
|    17     23     25     3.17    .83| .47  -.98| .15  -.60|  .65   .39| 95.8  92.2| ex17  |
|    21     23     25     3.17    .83| .94   .08|1.26   .63|  .36   .39| 95.8  92.2| ex21  |
|    13     22     25     2.59    .71|1.06   .28| .50  -.30|  .48   .44| 87.5  89.3| ex13  |
|    16     22     25     2.59    .71| .78  -.36|2.45  1.38|  .43   .44| 95.8  89.3| ex16  |
|     2     20     25     1.77    .59|1.20   .67|1.06   .30|  .41   .50| 79.2  83.8| ex02  |
|     5     20     25     1.77    .59| .88  -.26| .51  -.73|  .61   .50| 79.2  83.8| ex05  |
|    14     20     25     1.77    .59| .76  -.66| .58  -.57|  .63   .50| 87.5  83.8| ex14  |
|    22     20     25     1.77    .59|1.68  1.79|2.07  1.51|  .12   .50| 70.8  83.8| ex22  |
|    15     19     25     1.44    .55|1.43  1.37|1.82  1.43|  .25   .52| 75.0  80.5| ex15  |
|    20     19     25     1.44    .55| .85  -.43| .57  -.78|  .63   .52| 83.3  80.5| ex20  |
|     4     18     25     1.15    .53| .73 -1.03| .61  -.84|  .68   .53| 83.3  77.9| ex04  |
|     6     18     25     1.15    .53| .76  -.91| .53 -1.12|  .69   .53| 75.0  77.9| ex06  |
|    18     18     25     1.15    .53| .58 -1.77| .40 -1.57|  .77   .53| 91.7  77.9| ex18  |
|    23     18     25     1.15    .53|1.02   .17|1.10   .36|  .51   .53| 75.0  77.9| ex23  |
|     9     17     25      .88    .51|1.49  1.81|1.96  2.02|  .22   .54| 62.5  76.6| ex09  |
|     3     16     25      .63    .50|1.09   .44|1.03   .20|  .50   .54| 66.7  75.4| ex03  |
|    12     15     25      .39    .49|1.57  2.21|2.19  2.71|  .18   .54| 58.3  74.0| ex12  |
|------------------------------------+----------+----------+-----------+-----------+-------|
| MEAN     20.3   25.0     2.12    .68| .95   .04| .91   .11|           | 84.2  85.1|       |
| P.SD      2.6     .0     1.07    .18| .33   .98| .70  1.05|           | 11.8   6.9|       |
--------------------------------------------------------------------------------------------
```

Based on table 1 above, information is obtained that the examinees who have the highest ability, namely ex11 and ex19, can do 24 questions (total score) correctly from the 25 questions (total count) given, followed by ex01, ex07, ex08, ex10, ex17, and ex21 because they can work correctly as many as 23 items out of 25 items which are tested. The examinee who has the lowest ability is Ex12 because this student is only able to work correctly as many as 15 items (total score) of the 25 items (total count) tested.

According to Sumintono and Widhiarso (2014), the ability of examinees to answer questions correctly is measured on a logit scale. Students who have the ability in the same logit can answer the questions correctly as well. In the Winstep application, students' abilities are shown in the measure column. Ex19 and ex19 have the same logit, which is 4.06, both can answer 24 questions correctly. Ex1, ex7, ex8, ex10, ex17, and ex21 have a logit of 3.17 and can answer 23 questions correctly. According to Helverasary Untary, Risdianto, and Kusen (2002), the ability of students measured in logit has the same scale so information is obtained that Ex19 students (4.06 logit) have 2 times more abilities than Ex2 (1.77 logits).

According to Subando (2022), the criteria for the examinee's ability can be categorized into three, namely if the value of measure>M+S is categorized as high, M-1S to M+1S is categorized as moderate, and measure<M-1S is categorized as low, see the table below:

Table 2. Classification of the Examinee’s Ability

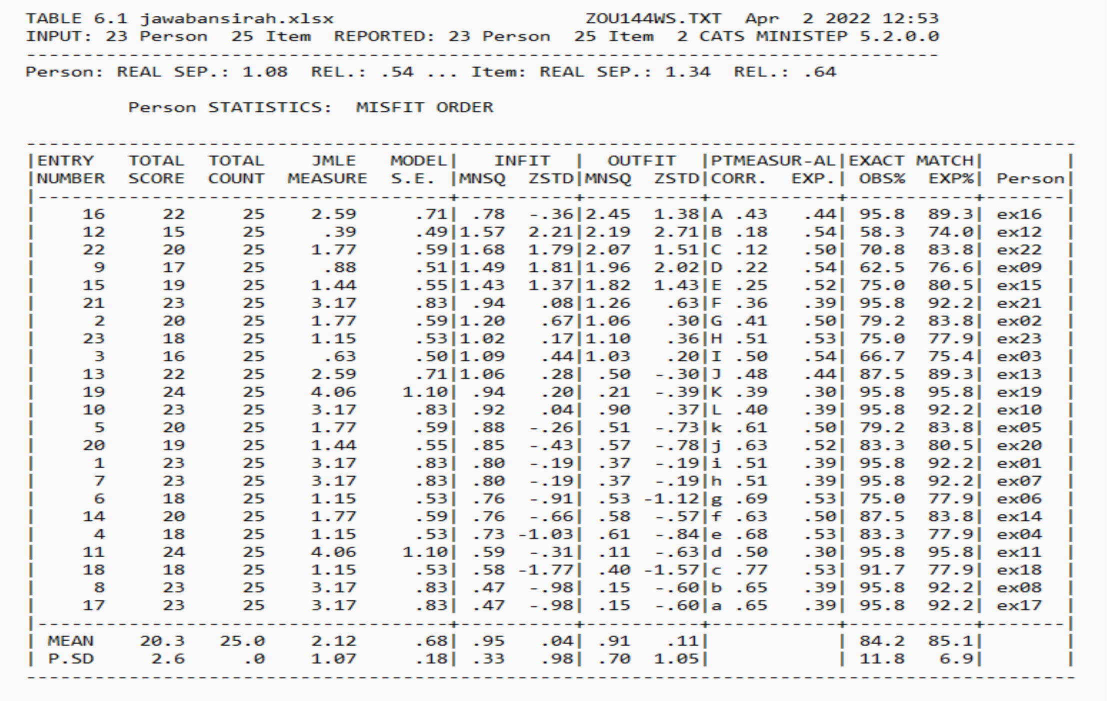
Num	Criteria	Cut score	Criteria
1	>M+1S	Logit>3.19	high
2	M-1S -M+1S	1,05≤logit≤3.19	moderate
3	< M-1S	Logit<1.05	low

Based on the above criteria, there are 2 examinees with high ability, 20 examinees with moderate ability, and 3 examinees with low ability.

Individual Fit with the Model

The level of individual fit with the model is presented by Winstep in the output table of person statistics: misfit order. The table will present the person who does not fit at the top.

Table 3. Individual Fit to Model

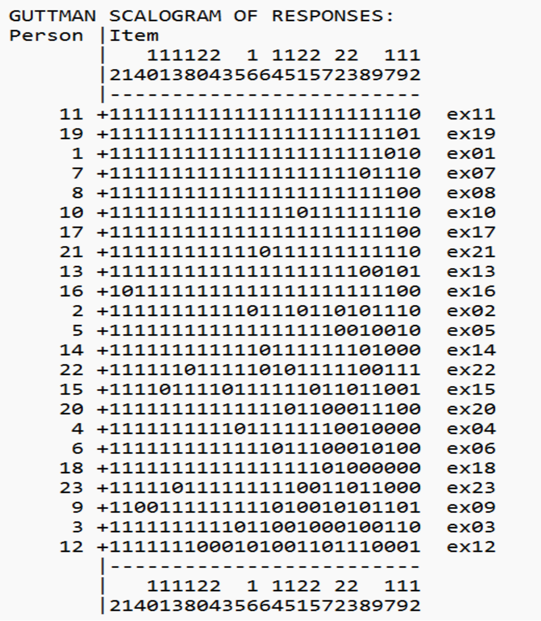


```
TABLE 6.1 jawabansirah.xlsx                      ZOU144WS.TXT  Apr  2 2022 12:53
INPUT: 23 Person  25 Item  REPORTED: 23 Person  25 Item  2 CATS MINISTEP 5.2.0.0
--------------------------------------------------------------------------------
Person: REAL SEP.: 1.08  REL.: .54 ... Item: REAL SEP.: 1.34  REL.: .64

        Person STATISTICS:  MISFIT ORDER

--------------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL     JMLE    MODEL|   INFIT   |  OUTFIT   |PTMEASUR-AL |EXACT MATCH|        |
|NUMBER  SCORE  COUNT  MEASURE    S.E. |MNSQ   ZSTD|MNSQ   ZSTD|CORR.  EXP. | OBS%  EXP%| Person |
|--------------------------------------+-----------+-----------+------------+-----------+--------|
|    16     22     25     2.59      .71| .78   -.36|2.45   1.38|A .43    .44| 95.8  89.3| ex16   |
|    12     15     25      .39      .49|1.57   2.21|2.19   2.71|B .18    .54| 58.3  74.0| ex12   |
|    22     20     25     1.77      .59|1.68   1.79|2.07   1.51|C .12    .50| 70.8  83.8| ex22   |
|     9     17     25      .88      .51|1.49   1.81|1.96   2.02|D .22    .54| 62.5  76.6| ex09   |
|    15     19     25     1.44      .55|1.43   1.37|1.82   1.43|E .25    .52| 75.0  80.5| ex15   |
|    21     23     25     3.17      .83| .94    .08|1.26    .63|F .36    .39| 95.8  92.2| ex21   |
|     2     20     25     1.77      .59|1.20    .67|1.06    .30|G .41    .50| 79.2  83.8| ex02   |
|    23     18     25     1.15      .53|1.02    .17|1.10    .36|H .51    .53| 75.0  77.9| ex23   |
|     3     16     25      .63      .50|1.09    .44|1.03    .20|I .50    .54| 66.7  75.4| ex03   |
|    13     22     25     2.59      .71|1.06    .28| .50   -.30|J .48    .44| 87.5  89.3| ex13   |
|    19     24     25     4.06     1.10| .94    .20| .21   -.39|K .39    .30| 95.8  95.8| ex19   |
|    10     23     25     3.17      .83| .92    .04| .90    .37|L .40    .39| 95.8  92.2| ex10   |
|     5     20     25     1.77      .59| .88   -.26| .51   -.73|k .61    .50| 79.2  83.8| ex05   |
|    20     19     25     1.44      .55| .85   -.43| .57   -.78|j .63    .52| 83.3  80.5| ex20   |
|     1     23     25     3.17      .83| .80   -.19| .37   -.19|i .51    .39| 95.8  92.2| ex01   |
|     7     23     25     3.17      .83| .80   -.19| .37   -.19|h .51    .39| 95.8  92.2| ex07   |
|     6     18     25     1.15      .53| .76   -.91| .53  -1.12|g .69    .53| 75.0  77.9| ex06   |
|    14     20     25     1.77      .59| .76   -.66| .58   -.57|f .63    .50| 87.5  83.8| ex14   |
|     4     18     25     1.15      .53| .73  -1.03| .61   -.84|e .68    .53| 83.3  77.9| ex04   |
|    11     24     25     4.06     1.10| .59   -.31| .11   -.63|d .50    .30| 95.8  95.8| ex11   |
|    18     18     25     1.15      .53| .58  -1.77| .40  -1.57|c .77    .53| 91.7  77.9| ex18   |
|     8     23     25     3.17      .83| .47   -.98| .15   -.60|b .65    .39| 95.8  92.2| ex08   |
|    17     23     25     3.17      .83| .47   -.98| .15   -.60|a .65    .39| 95.8  92.2| ex17   |
|--------------------------------------+-----------+-----------+------------+-----------+--------|
| MEAN    20.3   25.0     2.12      .68| .95    .04| .91    .11|            | 84.2  85.1|        |
| P.SD     2.6     .0     1.07      .18| .33    .98| .70   1.05|            | 11.8   6.9|        |
--------------------------------------------------------------------------------------------------
```

According to Susongko (2016), Wahyudi, Setyowati, and Partini (2020), Tabatabaee-Yazdi, Motallebzadeh, Ashraf, and Baghaei (2018), Sumintono (2018), Ee and Yeo (2018), Suraji, Totok Sumaryanto, and Khumaedi (2019), Upegui-Arango et al. (2020), Müller (2020), Hamdu, Fuadi, Yulianto, and Akhirani (2020), individual abilities match the model if (1) outfit Z-standard (ZSTD) scores: -2.0 < ZSTD < +2.0; (2) Outfit Mean Square (MNSQ) value: 0.5 < MNSQ < 1.5; (3) Point Measure Correlation (Pt Mean Corr) value: 0.4 <Pt Measure Corr < 0.85. Based on these criteria, the individual abilities that do not fit the model are ex16, ex12, ex22, ex09, ex15, and ex21.

Individuals that do not fit the model can be analyzed with a scalogram (H Untary et al., 2020). Winstep gives the following scalogram result:

Table 4. Scalogram



```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |   111122  1 1122 22  111
       |2140138043566451572389792
       |-------------------------
    11 +1111111111111111111111110  ex11
    19 +1111111111111111111111101  ex19
     1 +1111111111111111111111010  ex01
     7 +1111111111111111111101110  ex07
     8 +1111111111111111111111100  ex08
    10 +1111111111111110111111110  ex10
    17 +1111111111111111111111100  ex17
    21 +1111111111110111111111110  ex21
    13 +1111111111111111111100101  ex13
    16 +1011111111111111111111100  ex16
     2 +1111111111101110110101110  ex02
     5 +1111111111111111110010010  ex05
    14 +1111111111110111111101000  ex14
    22 +1111110111110101111100111  ex22
    15 +1111011110111111011011001  ex15
    20 +1111111111111101100011100  ex20
     4 +1111111111011111110010000  ex04
     6 +1111111111111011100010100  ex06
    18 +1111111111111111101000000  ex18
    23 +1111101111111110011011000  ex23
     9 +1100111111111010010101101  ex09
     3 +1111111111011001000100110  ex03
    12 +1111111000101001101110001  ex12
       |-------------------------
       |   111122  1 1122 22  111
       |2140138043566451572389792
```

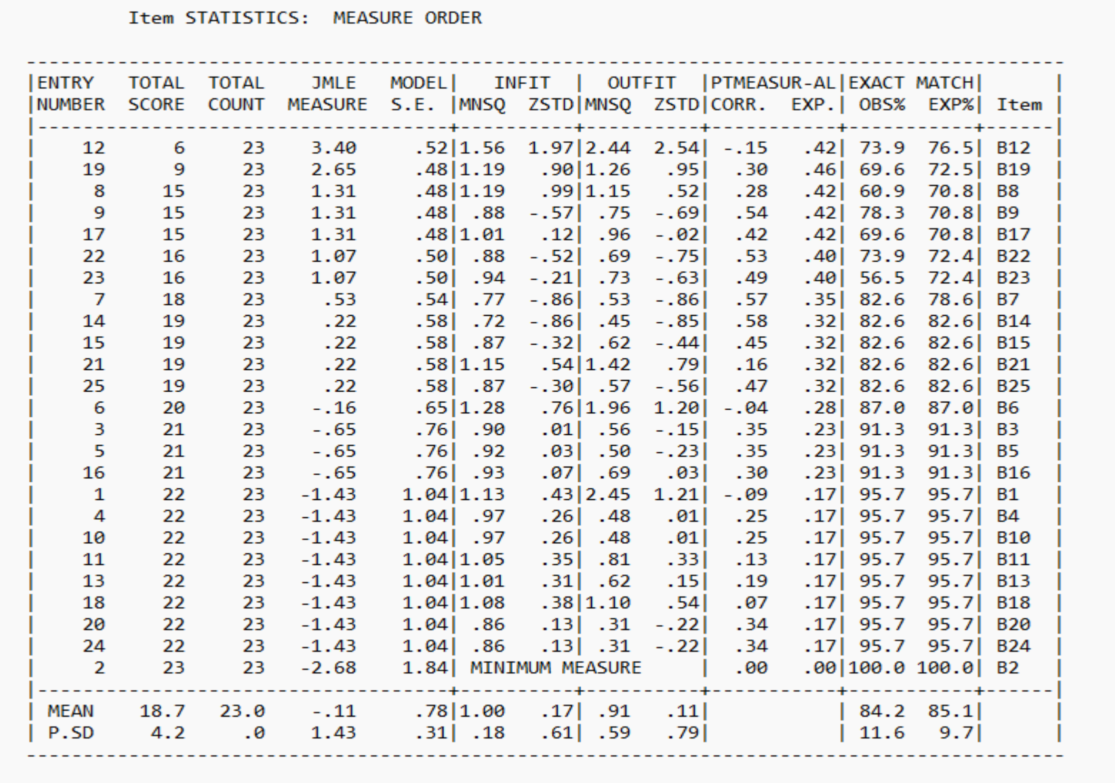
The number 1 indicates the correct answer and the number 0 indicates the wrong answer. The further to the right indicates the more difficult items and the further to the left indicates the easier items.

From the scalogram above, information is obtained that ex12 has an inconsistent pattern of answers because easy items cannot be answered correctly but difficult items can be answered correctly, this raises the assumption that respondent makes guesses in answering exam questions, as well as respondents ex09 and ex03. Ex15 and Ex22 also show a pattern of answers to several question items with a low level of difficulty that cannot be answered correctly meanwhile question items with a high level of difficulty can be answered correctly (H Untary et al., 2020). However, there is quite valuable information that the pattern of answers is not the same, thus it can be assumed that there is no mutual cheating among the examinees.

Item Difficulty Level

The item difficulty level is presented by Winstep in the output table item STATISTICS: MEASURE ORDER. The entry number shows the items that are sorted based on the items' difficulty level (item measure). In the last column, there is an item that shows the name of the item. The total score shows the number of questions that can be answered correctly from all the questions (total count) presented. The JMLE measure shows a measure of the difficulty level of an item in logit units. The questions are sorted from the hardest put on the top to the easiest put on the bottom row (Subando, 2022).

Table 5. Item Difficulty Level



```
          Item STATISTICS:  MEASURE ORDER

--------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL     JMLE   MODEL|   INFIT  |  OUTFIT  |PTMEASUR-AL|EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE   S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|-------------------------------------+----------+----------+-----------+-----------+------|
|    12      6     23     3.40     .52|1.56  1.97|2.44  2.54| -.15   .42| 73.9  76.5| B12  |
|    19      9     23     2.65     .48|1.19   .90|1.26   .95|  .30   .46| 69.6  72.5| B19  |
|     8     15     23     1.31     .48|1.19   .99|1.15   .52|  .28   .42| 60.9  70.8| B8   |
|     9     15     23     1.31     .48| .88  -.57| .75  -.69|  .54   .42| 78.3  70.8| B9   |
|    17     15     23     1.31     .48|1.01   .12| .96  -.02|  .42   .42| 69.6  70.8| B17  |
|    22     16     23     1.07     .50| .88  -.52| .69  -.75|  .53   .40| 73.9  72.4| B22  |
|    23     16     23     1.07     .50| .94  -.21| .73  -.63|  .49   .40| 56.5  72.4| B23  |
|     7     18     23      .53     .54| .77  -.86| .53  -.86|  .57   .35| 82.6  78.6| B7   |
|    14     19     23      .22     .58| .72  -.86| .45  -.85|  .58   .32| 82.6  82.6| B14  |
|    15     19     23      .22     .58| .87  -.32| .62  -.44|  .45   .32| 82.6  82.6| B15  |
|    21     19     23      .22     .58|1.15   .54|1.42   .79|  .16   .32| 82.6  82.6| B21  |
|    25     19     23      .22     .58| .87  -.30| .57  -.56|  .47   .32| 82.6  82.6| B25  |
|     6     20     23     -.16     .65|1.28   .76|1.96  1.20| -.04   .28| 87.0  87.0| B6   |
|     3     21     23     -.65     .76| .90   .01| .56  -.15|  .35   .23| 91.3  91.3| B3   |
|     5     21     23     -.65     .76| .92   .03| .50  -.23|  .35   .23| 91.3  91.3| B5   |
|    16     21     23     -.65     .76| .93   .07| .69   .03|  .30   .23| 91.3  91.3| B16  |
|     1     22     23    -1.43    1.04|1.13   .43|2.45  1.21| -.09   .17| 95.7  95.7| B1   |
|     4     22     23    -1.43    1.04| .97   .26| .48   .01|  .25   .17| 95.7  95.7| B4   |
|    10     22     23    -1.43    1.04| .97   .26| .48   .01|  .25   .17| 95.7  95.7| B10  |
|    11     22     23    -1.43    1.04|1.05   .35| .81   .33|  .13   .17| 95.7  95.7| B11  |
|    13     22     23    -1.43    1.04|1.01   .31| .62   .15|  .19   .17| 95.7  95.7| B13  |
|    18     22     23    -1.43    1.04|1.08   .38|1.10   .54|  .07   .17| 95.7  95.7| B18  |
|    20     22     23    -1.43    1.04| .86   .13| .31  -.22|  .34   .17| 95.7  95.7| B20  |
|    24     22     23    -1.43    1.04| .86   .13| .31  -.22|  .34   .17| 95.7  95.7| B24  |
|     2     23     23    -2.68    1.84| MINIMUM MEASURE     |  .00   .00|100.0 100.0| B2   |
|-------------------------------------+----------+----------+-----------+-----------+------|
| MEAN     18.7   23.0     -.11     .78|1.00   .17| .91   .11|           | 84.2  85.1|      |
| P.SD      4.2     .0     1.43     .31| .18   .61| .59   .79|           | 11.6   9.7|      |
--------------------------------------------------------------------------------------------
```

From table 5 above, information is obtained that the item measure is correlated with the total score, a high item measure value has a small total score, while a small item measure has a large total score. Sumintono (2018) states that items that have the same logit will have the same number of participants who can answer correctly. Items B8, B9, and B17 with a value of 1.31 can be answered correctly by 15 examinees. This shows that the logit has the same scale because the logit scale is the same, item B12 with a logit of 3.40 means that it has an estimated difficulty level of 3 times higher compared to item B8 which has a logit value of 1.31. Items B22 and B23 with a logit of 1.07 have an estimated level of difficulty 5 times higher compared to items B14, B15, B21, and B25 with a logit of 0.22 (H Untary et al., 2020).

From the table above, information has also been obtained that item B12 is the most difficult item with a logit value of 3.40 and 6 examinees can answer B12 correctly out of 23 examinees. Item B19 is ranked number 2 with a logit value of 2.65 and this item can be answered correctly by 9 out of 23 examinees.

The easiest item is item B2 with a logit value of -2.68 and 23 examinees answered correctly. It means that item B2 can be answered correctly by the examinees. Item B24 is more difficult than item B2 because 22 of the 23 examinees answer correctly and the logit score is -1.43.

According to Mardapi (1998), and Pratama (2020) the level of question item difficulty can be classified into three. If the logit value is >2, question items are categorized as very difficult. If the logit value is <-2, question items are categorized as very easy. And if the logit value is between -2 to 2, question items are categorized as moderate. Based on these criteria, the very difficult question items are considered to be 2 items, 23 question items for moderate difficulty, and there is no question item belongs to very easy, see table 6.

According to Susongko (2016), Purnamasari, Hadi, and Istiyono (2018), the ideal item has a logit value between -2 to +2. Thus, the Tarikh question items which are developed by Husnaiayin teachers can be concluded that, from the aspect of the difficulty level of the items, 92% of the items are ideal and only 8% are not ideal.

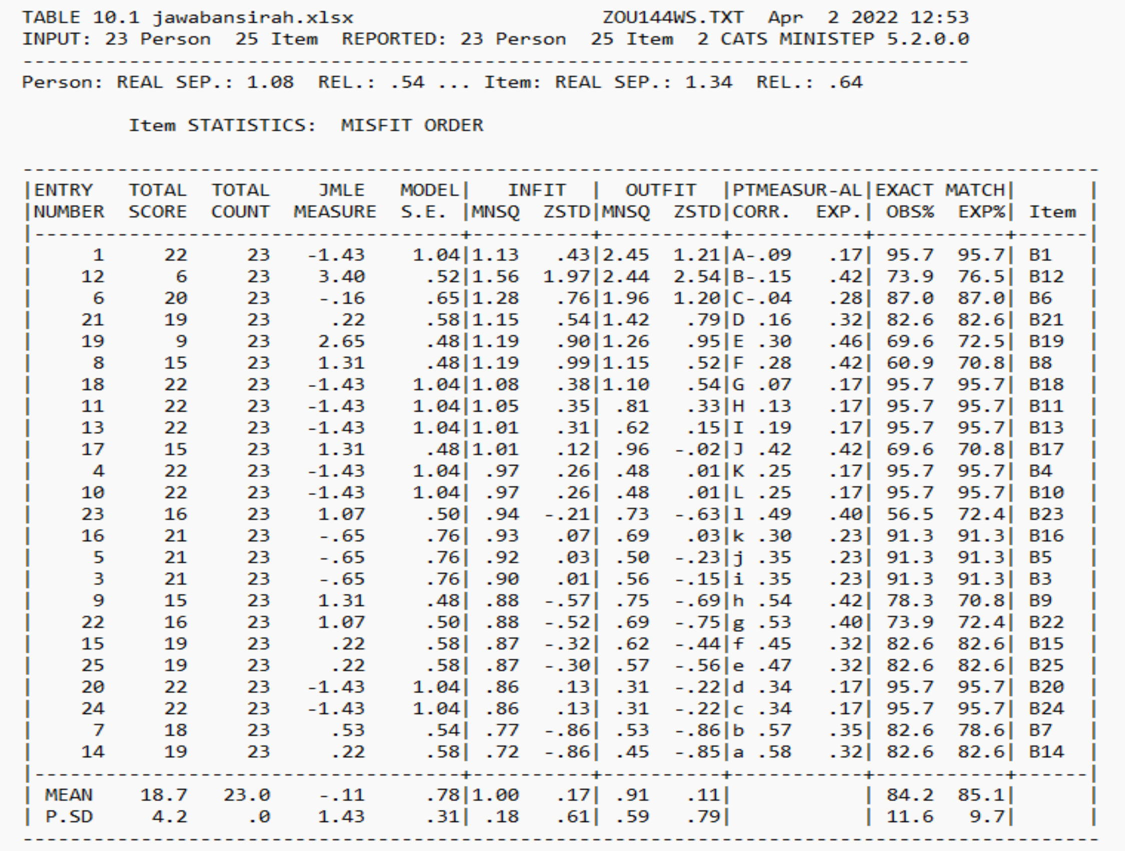
Table 6. Classification of Question Item Difficulty Level

Num	Criteria Score	Criteria	Total	Percentage
1	Measure >+2	Very difficult	2	8
2	-2 s/d 2	Moderate	23	92
3	Measure<-2	Very easy		
	Total		25	100

Analysis of the Fit Level of Question Items with the Model

The fit level of items is displayed by Winstep in the output of table 10, Item STATISTICS: MISFIT ORDER. Misfit items are placed in the top row, see table 7.

Table 7. Item Fit with Model

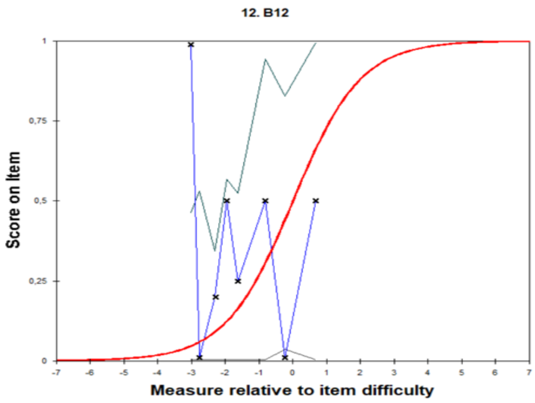


```
TABLE 10.1 jawabansirah.xlsx                    ZOU144WS.TXT  Apr  2 2022 12:53
INPUT: 23 Person  25 Item  REPORTED: 23 Person  25 Item  2 CATS MINISTEP 5.2.0.0
--------------------------------------------------------------------------------
Person: REAL SEP.: 1.08  REL.: .54 ... Item: REAL SEP.: 1.34  REL.: .64

         Item STATISTICS:  MISFIT ORDER

------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL     JMLE   MODEL|   INFIT  |  OUTFIT  |PTMEASUR-AL|EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE   S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|------------------------------------+----------+----------+-----------+-----------+------|
|     1     22     23    -1.43   1.04|1.13   .43|2.45  1.21|A-.09   .17| 95.7  95.7| B1   |
|    12      6     23     3.40    .52|1.56  1.97|2.44  2.54|B-.15   .42| 73.9  76.5| B12  |
|     6     20     23     -.16    .65|1.28   .76|1.96  1.20|C-.04   .28| 87.0  87.0| B6   |
|    21     19     23      .22    .58|1.15   .54|1.42   .79|D .16   .32| 82.6  82.6| B21  |
|    19      9     23     2.65    .48|1.19   .90|1.26   .95|E .30   .46| 69.6  72.5| B19  |
|     8     15     23     1.31    .48|1.19   .99|1.15   .52|F .28   .42| 60.9  70.8| B8   |
|    18     22     23    -1.43   1.04|1.08   .38|1.10   .54|G .07   .17| 95.7  95.7| B18  |
|    11     22     23    -1.43   1.04|1.05   .35| .81   .33|H .13   .17| 95.7  95.7| B11  |
|    13     22     23    -1.43   1.04|1.01   .31| .62   .15|I .19   .17| 95.7  95.7| B13  |
|    17     15     23     1.31    .48|1.01   .12| .96  -.02|J .42   .42| 69.6  70.8| B17  |
|     4     22     23    -1.43   1.04| .97   .26| .48   .01|K .25   .17| 95.7  95.7| B4   |
|    10     22     23    -1.43   1.04| .97   .26| .48   .01|L .25   .17| 95.7  95.7| B10  |
|    23     16     23     1.07    .50| .94  -.21| .73  -.63|l .49   .40| 56.5  72.4| B23  |
|    16     21     23     -.65    .76| .93   .07| .69   .03|k .30   .23| 91.3  91.3| B16  |
|     5     21     23     -.65    .76| .92   .03| .50  -.23|j .35   .23| 91.3  91.3| B5   |
|     3     21     23     -.65    .76| .90   .01| .56  -.15|i .35   .23| 91.3  91.3| B3   |
|     9     15     23     1.31    .48| .88  -.57| .75  -.69|h .54   .42| 78.3  70.8| B9   |
|    22     16     23     1.07    .50| .88  -.52| .69  -.75|g .53   .40| 73.9  72.4| B22  |
|    15     19     23      .22    .58| .87  -.32| .62  -.44|f .45   .32| 82.6  82.6| B15  |
|    25     19     23      .22    .58| .87  -.30| .57  -.56|e .47   .32| 82.6  82.6| B25  |
|    20     22     23    -1.43   1.04| .86   .13| .31  -.22|d .34   .17| 95.7  95.7| B20  |
|    24     22     23    -1.43   1.04| .86   .13| .31  -.22|c .34   .17| 95.7  95.7| B24  |
|     7     18     23      .53    .54| .77  -.86| .53  -.86|b .57   .35| 82.6  78.6| B7   |
|    14     19     23      .22    .58| .72  -.86| .45  -.85|a .58   .32| 82.6  82.6| B14  |
|------------------------------------+----------+----------+-----------+-----------+------|
| MEAN    18.7   23.0     -.11    .78|1.00   .17| .91   .11|           | 84.2  85.1|      |
| P.SD     4.2     .0     1.43    .31| .18   .61| .59   .79|           | 11.6   9.7|      |
------------------------------------------------------------------------------------------
```

In addition to detecting the level of difficulty of the question item, Winstep also detects the level of question item fit (item fit). Items that meet the fit criteria will be able to carry out their measurement functions well, but items that do not fit will not be able to carry out their measurement functions properly. Questions that do not fit raise students' misconceptions about the questions. This information is important for teachers to improve the quality of teaching so as not to create misconceptions. The level of conformity of items is displayed by Winstep in the output table 10 Item STATISTICS: MISFIT ORDER.

Based on the criteria for the fit of question items and the misfit table, items B12, B20, and B24 the MNSQ, ZSTD, and

Ptmeasure corr outfit values do not meet the criteria so the items are stated as a misfit and raise students' misconceptions about the questions (Putri, Kartono, & Supriyadi, 2020). Meanwhile, the other question items have criteria that are accepted and some are rejected so that the misfit can be tolerated. The misfit in item B12 is also seen from the ICC graph in Picture 1.

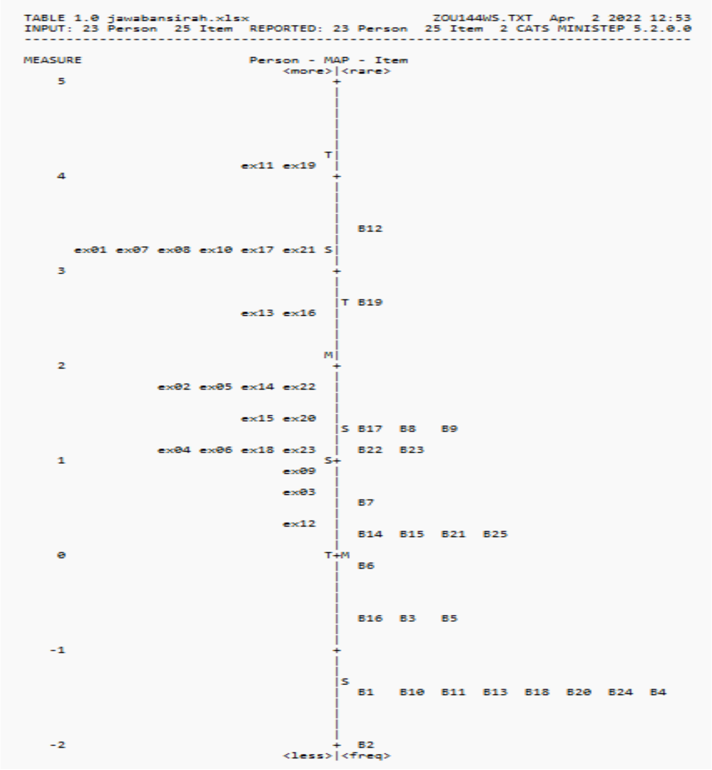


Picture 1. ICC Graphic

The black cross (x) indicates the response is outside the fit curve, so item 10 is a misfit item.

Abilities map and item difficulty level

The interaction pattern between individuals and question items is presented in the form of a diagram in Picture 2. The left-wing describes the abilities of students ranging from those with the highest abilities (located at the top) to students with the lowest abilities (bottom row). The right-hand wing is the item's difficulty level, starting from the most difficult item (top) to the easiest item (bottom). Both (ability and item difficulty level) is on the same scale so they can be compared (Rahim & Haryanto, 2021).



```
TABLE 1.0 jawabansirah.xlsx                        ZOU144WS.TXT  Apr  2 2022 12:53
INPUT: 23 Person  25 Item  REPORTED: 23 Person  25 Item  2 CATS MINISTEP 5.2.0.0
--------------------------------------------------------------------------------

MEASURE                    Person - MAP - Item
                                <more>|<rare>
    5                                 +
                                      |
                                      |
                                      |
                                      |
                                      |
                                     T|
                           ex11 ex19  |
    4                                 +
                                      |
                                      |
                                      |
                                      |
                                      |  B12
                                      |
       ex01 ex07 ex08 ex10 ex17 ex21 S|
                                      |
    3                                 +
                                      |
                                      |
                                      |T B19
                           ex13 ex16  |
                                      |
                                      |
                                      |
                                     M|
    2                                 +
                                      |
                 ex02 ex05 ex14 ex22  |
                                      |
                                      |
                           ex15 ex20  |
                                      |S B17  B8   B9
                                      |
                 ex04 ex06 ex18 ex23  |  B22  B23
    1                                S+
                                ex09  |
                                      |
                                ex03  |
                                      |  B7
                                      |
                                ex12  |
                                      |  B14  B15  B21  B25
                                      |
    0                                T+M
                                      |  B6
                                      |
                                      |
                                      |
                                      |
                                      |  B16  B3   B5
                                      |
                                      |
   -1                                 +
                                      |
                                      |
                                      |S
                                      |  B1   B10  B11  B13  B18  B20  B24  B4
                                      |
                                      |
                                      |
                                      |
   -2                                 +  B2
                                <less>|<freq>
```

Picture 2. Ability map and difficulty level of items

One of the advantages of the Rasch model and Winstep is that it provides a map of the ability and difficulty level of items on the same scale. The person map and also item map can be seen in Picture 2. The left side describes the ability of the examinees and the right side describes the level of item difficulty.

From the map, information is obtained that, in general, the ability of the examinees is above the item difficulty level or the item difficulty level is below the examinees' ability. Item B12 is the most difficult item, and ex 11, and ex19 are the examinees who have the highest ability. Although item B12 is the most difficult, the item is still below the ability of ex11 and ex19, so both have a fairly high chance of answering correctly (Aprilia, Lidinillah, & Giyartini, 2021).

However, item B12 is far above the ability of ex03 and ex09 so both of them have a small chance of being able to answer item B12 correctly.

Examinee ex12 is the one with the lowest abilities but based on the map above, ex12 abilities are above the level of difficulty points B14, B15, B21, B25, B6, B16, B3, B5, B1, B10, B11, B18, B20, B24, B4, and B2. It means that examinee ex12 has a great chance of being able to answer these questions correctly. Thus, in general, the items are very easy and are below the ability of the examinees. However, there is one item that is outside T (outlier item), namely item B2, the item is categorized as a very difficult item.

Based on the analysis of the individual ability level, the suitability of the individual ability with the model, the difficulty level of the item, the suitability of the item difficulty level with the model, and the ability map with the item suitability level, it is found that the student's ability level is above the question difficulty level and the proportion of the item difficulty level is less than ideal so it needs to be improved or revised dealing with the questions. From the explanation above, it can also be seen that the contribution of the Rasch Model Theory in producing quality questions and important information about the ability of the examinees is indeed beneficial.

CONCLUSION

Based on the results of the research and discussion above, it can be concluded that (1) 2 students have high abilities (ex11, ex19), 20 students are in moderate abilities, and 3 students are in low abilities (ex3, ex9, ex12). Conclusion number (2) tells that there are students who experience misconceptions or do not fit with the model, they are ex16, ex12, ex22, ex09, ex15, and ex21. Conclusion number (3) tells about the most difficult items which are found in 2 items (B12 and B29). Those items are categorized as moderate difficulty levels as many as 23 items, and there are no items that are categorized as very easy. Conclusion number (4) shows that the items that do not fit are B12, B20 and B24. Conclusion number (5) is about the interaction map of ability and item difficulty level that states the examinees' ability is above the item difficulty level or the item difficulty level is below the examinees' ability. Based on the conclusions above, the questions are then revised so that the level of difficulty from the questions is equal to the ability of the examinees and the proportion of the difficulty level of the questions is ideal.

REFERENCES

Aprilia, M., Lidinillah, D. A. M., & Giyartini, R. (2021). Pengembangan Instrumen Penilaian Kreativitas Siswa melalui Analisis Rasch Model di Sekolah Dasar. Jurnal Basicedu, 5(4), 2302-2310.

Aziz, R. (2015). Aplikasi model Rasch dalam pengujian alat ukur kesehatan mental di tempat kerja. Psikoislamika: Jurnal Psikologi dan Psikologi Islam, 12(2), 29-39.

Ee, N. S., & Yeo, K. J. (2018). Item Analysis for the Adapted Motivation Scale Using Rasch Model. International Journal of Evaluation and Research in Education, 7(4), 264-269.

Fernanda, J. W., & Hidayah, N. (2020). Analisis kualitas soal ujian statistika menggunakan classical test theory dan rasch model. Square: Journal of Mathematics and Mathematics Education, 2(1), 49-60.

Hamdu, G., Fuadi, F., Yulianto, A., & Akhirani, Y. (2020). Items quality analysis using rasch model to measure elementary school students' critical thinking skill on stem learning. JPI (Jurnal Pendidikan Indonesia), 9(1), 61-74.

Mardapi, D. (1998). Analisis butir dengan teori tes klasik dan teori respons butir. Jurnal Kependidikan, 28(2).

Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? Journal of Statistical Distributions and Applications, 7(1), 1-12.

Pratama, D. (2020). Analisis kualitas tes buatan guru melalui pendekatan item response theory (IRT) model rasch. Tarbawy: Jurnal Pendidikan Islam, 7(1), 61-70.

Purnamasari, U. D., Hadi, S., & Istiyono, E. (2018). The Characteristic Of Islamic Religion And Character Education Test Using Rasch Model. Jurnal Mudarrisuna: Media Kajian Pendidikan Agama Islam, 8(2), 372-382.

Putri, B. S. F., Kartono, K., & Supriyadi, S. (2020). Analysis Of Essay Test Instruments Using Higher Order Thinking Skill (HOTS) at High School Mathematics Students Using The Rasch Model. Journal of Research and Educational Research Evaluation, 9(2), 58-69.

Ramadhanti, D., Rahmatullah, R., Wilujeng, I., & Chusna, D. S. A. (2021). Profile of ICT Literacy Capability Using Rasch Model at SMP Negeri 2 Fakfak. Indonesian Journal of Applied Science and Technology, 2(3), 89-95.

Rahim, A., & Haryanto, H. (2021). Implementation of Item Response Theory (IRT) Rasch Model in Quality Analysis of Final Exam Tests in Mathematics. Journal of Research and Educational Research Evaluation, 57-65.

Riadi, A. (2018). Kompetensi Guru dalam pelaksanaan evaluasi pembelajaran. ITTIHAD, 15(28), 52-67.

Setianingsih, R. (2018). Analisis Kualitas Butir Soal Ujian Akhir Semester 2 Mata Pelajaran Sejarah Kebudayaan Islam Kelas VIII di MTs Ma'arif NU Kemranjen Kabupaten Banyumas Tahun Ajaran 2016/2017. IAIN Purwokerto.

Subando, J. (2022). Evaluasi Hasil Belajar Pendidikan Agama Islam. Klaten: Lakeisha.

Suraji, S., Totok Sumaryanto, F., & Khumaedi, M. (2019). The analysis of instrument of the ability to acting and thinking creatively based Rasch model. Journal of Research and Educational Research Evaluation, 8(1), 48-56.

Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. Paper presented at the 1st International Conference on Education Innovation (ICEI 2017). Atlantis Press.

Sumintono, B., & Widhiarso, W. (2014). Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi): Trim Komunikata Publishing House.

Susongko, P. (2016). Validation of science achievement test with the rasch model. Jurnal Pendidikan IPA Indonesia, 5(2), 268-277.

Suyata, P., Mardapi, D., Kartowagiran, B., & Retnawati, H. (2011). Model Pengembangan Bank Soal Berbasis Guru Dan Mutu Pendidikan. Jurnal Kependidikan: Penelitian Inovasi Pembelajaran, 41(2).

Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and Validation of a Teacher Success Questionnaire Using the Rasch Model. International Journal of Instruction, 11(2), 129-144.

Triono, D., Sarno, R., & Sungkono, K. R. (2020). Item Analysis for Examination Test in the Postgraduate Student's Selection with Classical Test Theory and Rasch Measurement Model. Paper presented at the 2020 International Seminar on Application for Technology of Information and Communication (iSemantic).

Undang–Undang, R. (2017). Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. Direktorat Jendral Kelembagaan IPTEK & DIKTI.

Untary, H., Risdianto, E., & Kusen. (2002). Analisis Data Penelitian dengan Model Rash dan Winstep. Bogor: Halaman Moeka Publishing.

Untary, H., Risdianto, E., & Kusen. (2020). Analisis Data Penelitian dengan Model rach dan Winstep. Jakarta: Halaman Moeka Publishing.

Upegui-Arango, L. D., Forkmann, T., Nielsen, T., Hallensleben, N., Glaesmer, H., Spangenberg, L., . . . Boecker, M. (2020). Psychometric evaluation of the Interpersonal Needs Questionnaire (INQ) using item analysis according to the Rasch model. PloS one, 15(8), e0232030.

Wahyudi, A., Setyowati, A., & Partini, S. (2020). Analisis Model Rasch Pada Pengembangan Skala Resiliensi. Jurnal Fokus Konseling, 6(2), 68-74.