CONCIENCIA: Journal of Islamic Education
Vol. 23 No. 2, December 2023, 147-156
p-ISSN: 1412-2545 e-ISSN: 2622-7304 http://jurnal.radenfatah.ac.id/index.php/conciencia

ANALYSIS OF ITEM QUALITY IN ISLAMIC RELIGIOUS EDUCATION SUBJECTS WITH RASCH MODELING

Evita[1], Eva Dwi Kumala Sari[2], Erni Susieni[3], Badrud Tamam[4]

[1,2,3]Sekolah Tinggi Ilmu Tarbiyah Al-Marhalah Al-'Ulya Bekasi, West Java, Indonesia
[4]Universitas Wiralodra Indramayu, West Java, Indonesia
Email: eva@almarhalah.ac.id

Abstract: The role of teachers in schools is expected, especially in evaluating student learning; teachers must pay attention to the quality of the items made. This study aims to determine the instrument's validity or the quality of the item's suitability with the model known as Item FIT (item suitability) on the End of Semester Assessment of Islamic Religious Education. The sample in this study was 216 respondents, using cluster random sampling. The results obtained; 1) the end-of-semester test given is categorized as easy when viewed from the level of student ability above the level of difficulty of the question; 2) the level of student ability is moderate with logit (0) producing very high information. Then, it will produce low information at the low or high ability level. It means that the PAI test produces maximum information when given to individuals with moderate ability. 3) Item fit criteria from 216 students, there are 93% that fit the model (fit model), and these students are said to fit the model because they have met at least two existing criteria; 4) The results of the analysis of the summary table above, 40 items analyzed there are 33 items (83%) that fit the model (fit model). 5) Reliability for Pearson reliability of 0.83 and item reliability of 0.97, it can be concluded that the consistency of student answers is good and the quality of the items in the PAI test is excellent reliability; 6) Gender bias analysis can be seen that 2 of the 40 items are DIF or need to be corrected because they harm certain gender groups.

Keywords: Question Items, Islamic Education, Rasch Modeling.

INTRODUCTION

Education is a conscious effort to prepare students for their future roles through counseling, education, and training activities. (Wahyudin, 2016)

Education is a learning process that occurs throughout life in various places and situations, aiming to impact each individual's development positively. That education continues throughout life (Pristiwanti et al., 2022).

Evaluation in education is assessing a student's growth and progress toward the goals or values set out in the curriculum. Evaluating something is based on predetermined criteria or objectives and deciding what to evaluate. Assessment is an essential part of the teacher's professional duties in learning. Educators can use assessment to determine how much performance they can get from the learning process (Hamzah, 2014).

Evaluation activities have significant benefits not only in education but also in learning activities. This is because the results of learning activities carried out can be seen through evaluation, and the follow-up that will be carried out can be decided based on these results (Wahyudi, 2016).

Islamic teachings also apply the assessment applied by Allah SWT. Allah says in the Qur'an that student evaluation work is an essential task in a series of educational processes carried out by educators. (Ihsan & Ihsan, 2001).

أحَسِبَ ٱلنَّاسُ أَن يُتۡرَكُوٓاْ أَن يَقُولُوٓاْ ءَامَنَّا
وَهُمۡ لَا يُفۡتَنُونَ . وَلَقَدۡ فَتَنَّا ٱلَّذِينَ مِن قَبۡلِهِمۡۖ
فَلَيَعۡلَمَنَّ ٱللَّهُ ٱلَّذِينَ صَدَقُواْ وَلَيَعۡلَمَنَّ
ٱلۡكَٰذِبِينَ.

The meaning: " (2) Do people think that they are left to say: "We have believed," while they are not tested again? (3) And indeed we have tested those who were before them, so indeed Allah knows those who are true and indeed He knows those who lie". (Q.S. Al-Ankabut: 2-3); (Taufiq, 2018).

The purpose of educational evaluation is to determine the level of knowledge of students so that the level of intelligence of these students can be known. The evaluation also helps teachers and schools see if the learning process is successful and if the education process is running well. (Halik, Mania & Nur, 2019). Usually, to measure or evaluate the level of students' abilities and understanding, teachers can give tests. Therefore, evaluation skills and abilities are professional skills that every teacher or prospective teacher must have.

Arikunto says that a test is a tool or way to find or measure something in an atmosphere with certain rules. (Arikunto, 2016).

Tests are used as tools to measure various performances and collect data. A test must have validity, i.e., be able to measure what it is supposed to measure, and it must have reliability, i.e., be repeatable with high consistency. Measurement refers to the numerical value that results from a test. Then, the data that has been obtained is evaluated.(Gumantan, Mahfud, and Yuliandra 2020)

The standard of success of an educational process can be seen from the quality of learning outcomes assessment that is adjusted to the applicable curriculum standards. Therefore, assessment, in this case, occupies an important place to be able to measure the success of learning (Cahyati, 2018).

Item analysis is an activity that teachers must do to improve the quality of written questions. This activity is the process of collecting, summarizing, and using information from student responses to make decisions about each assessment (Mahendra, 2019). The item analysis process certainly requires accuracy so that the results obtained can be used and provide accurate information. There are many analysis tools that have been developed as options for users. However, of course, it is necessary to pay attention to the suitable and accurate analysis tool to use. The development of analytical tools is quite extensive. There is the modern valuation approach and the classic approach. These two approaches each have advantages and disadvantages. Of course, in choosing this approach, it is necessary to know the reliability of each of the two.

There are two types of theories in test instrument analysis, namely Classical Test Theory (CTT) and modern test theory (Item Response Theory / IRT). In analyzing test instruments, the theory that is often used is classical test theory using the Iteman application. However, the use of modern test theory in test instrument analysis is rarely done. The advantage of classical test theory is its affordable practicality and efficiency. However, the drawback of this theory lies in the grain characteristics that depend on the sample taking the test. Therefore, the results of these tests can be inconsistent or change if the sample being tested also changes. In the case of a difficult test question, the student's ability will appear low, while in the case of an easy test question, the student's ability will appear high.

Rasch modeling is one of the most famous modern analytical theories. Rasch modeling has the ability to predict missing data. Rasch modeling provides relevant information about test takers and the quality of the questions because it creates a consistent measurement ratio. However, there is a drawback to the Rasch model: The mathematical equations in IRT are more challenging to understand than the equations in CTT, so a computer is required. (Purniasari, Masykuri, and Ariani 2021)

Rasch Measurement Model (RMM) is the simplest IRT model with a 1-parameter model but has very strong assessment or analysis capabilities. (Fischer & Molenarr, 1995; Haiyang, 2010; Sumintono, 2016). Georg Rasch used a logarithmic function to overcome inequalities between intervals. (Bond & Fox, 2015; Fischer & Molenaar, 1995; Sumintono, 2016; Van Zile-Tamsen, 2017). Analysis with Rasch modeling produces a statistical analysis of fit (fit statistics) that provides information to researchers whether the data obtained ideally describes that people who have high ability provide a pattern of answers to items according to their level of difficulty. (Sumintono 2016)

The purpose of this study was to determine the quality of items from the end-of-semester test on the subject of Islamic Religious Education (PAI) with Rasch modeling.

METHODS

This study aims to analyze the quality of multiple-choice items on the End of Semester Assessment of Islamic Religious Education class X at Bina Karya Mandiri Vocational School. The analysis in this study used Rasch modeling analysis with the help of the Winstep program. Rasch modeling has the advantage that test items and test takers are independent of each other, whereas, in classical theory, tests depend on test takers. Rasch modeling can produce a measuring scale with the same interval precision and has units. Rasch modeling, in addition to being able to see the level of item difficulty, is used to see the validity of the instrument or the quality of the fit of the item to the model, known as item fit (item fit). Item FIT explains whether items function normally when making measurements or not. If an item is found that is not FIT (misfit), then the item is categorized as invalid, indicating that there has been a misconception (what the statement means but is interpreted differently by the respondent).Rasch modeling also detects if there are respondents whose response patterns do not match. What is meant by different response patterns is the discrepancy between the answers given based on their abilities compared to the ideal model. This can be used by teachers to find out the consistency of student thinking or to find out if cheating is being done. As in the examination of items in terms of fit in the model, the same criteria are used in the examination of persons. According to Boone et al., the criteria used to check the suitability of unsuitable persons (outliers or misfits). Rasch modeling knows the description of the distribution of students' or respondents' abilities and the distribution of the difficulty level of questions with the same scale, the level of students' ability to take the PAI test, the item fit criteria, Item fit explains whether the items function normally to make measurements or not. If there is an item that does not fit, it is an indication that there are misconceptions in students about the item, item reliability Pearson, and gender bias.

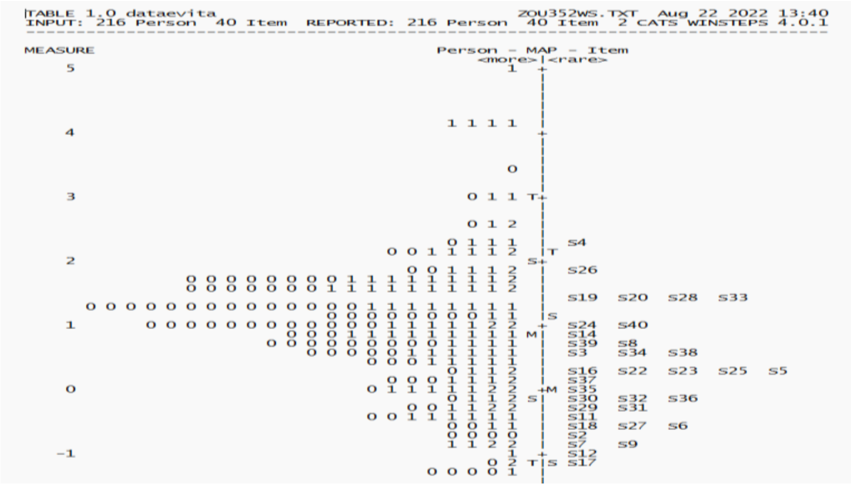
Setting, population, and sample

Rasch modeling alsodetects if there are respondents whose response patterns do not match. What is meant by different response patterns is the discrepancy between the answers given based on their abilities compared to the ideal model. This can be used by teachers to find out the consistency of student thinking or to find out if cheating is being done. As in the examination of items in terms of fit in the model, the same criteria are used in the examination of persons. According to Boone et al., the criteria used to check the suitability of unsuitable persons (outliers or misfits). Rasch modeling knows the description of the distribution of students' or respondents' abilities and the distribution of the difficulty level of questions with the same scale, the level of students' ability to take the PAI test, the item fit criteria, Item fit explains whether the items function normally to make measurements or not. If there is an item that does not fit, it is an indication that there are misconceptions in students about the item, item reliability Pearson, and gender bias.

The PAI questions were in the form of multiple-choice questions, as many as 40. PAS scores were collected according to the selected sample; then, the scores were analyzed with Rasch modeling assisted bytheWinstepapplication.

RESULT

Analysis with Rasch modeling described by the Wright map image below is to describe the distribution of student or respondent abilities and the distribution of question difficulty levels on the same scale.

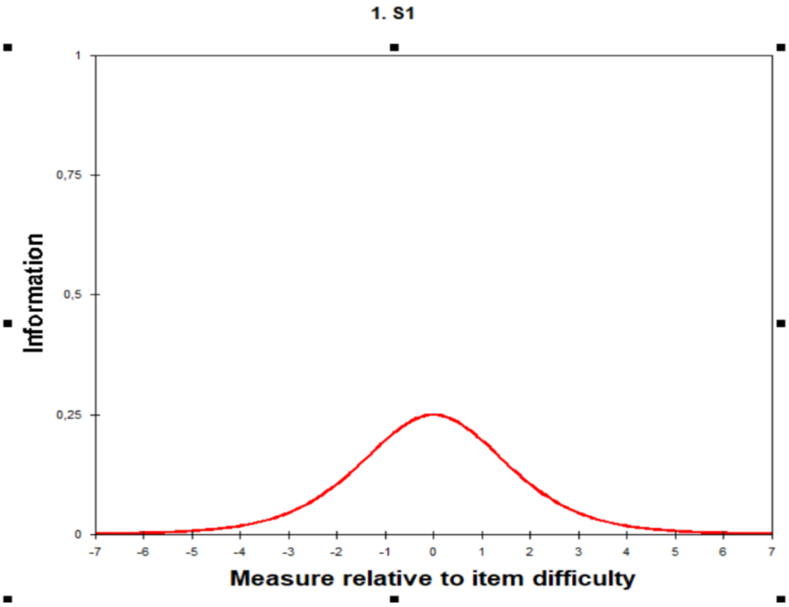


```
TABLE 1.0 dataevita                              ZOU352WS.TXT  Aug 22 2022 13:40
INPUT: 216 Person  40 Item  REPORTED: 216 Person  40 Item  2 CATS WINSTEPS 4.0.1
--------------------------------------------------------------------------------

MEASURE                                  Person - MAP - Item
                                             <more>|<rare>
     5                                          1  +
                                                   |
                                                   |
                                                   |
                                                   |
                                                   |
                                          1 1 1 1  |
     4                                             +
                                                   |
                                                   |
                                                   |
                                                0  |
                                                   |
                                                   |
     3                                      0 1 1 T+
                                                   |
                                                   |
                                            0 1 2  |
                                                   |
                                          0 1 1 1  |  S4
                                    0 0 1 1 1 1 2  |T
     2                                            S+
                                      0 0 1 1 1 2  |  S26
                0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2  |
                0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2  |
                                                   |  S19  S20  S28  S33
      0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1  |
                              0 0 0 0 0 0 0 0 1 1  |S
     1      0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 2 2  +  S24  S40
                          0 0 0 1 1 1 1 1 1 1 1 1 M|  S14
                        0 0 0 0 0 0 0 0 0 1 1 1 1  |  S39  S8
                            0 0 0 0 0 1 1 1 1 1 1  |  S3   S34  S38
                                  0 0 0 1 1 1 1 1  |
                                          0 1 1 2  |  S16  S22  S23  S25  S5
                                    0 0 0 1 1 1 2  |  S37
     0                            0 1 1 1 1 1 2 2  +M S35
                                          0 1 1 2 S|  S30  S32  S36
                                      0 0 1 1 2 2  |  S29  S31
                                  0 0 1 1 1 1 1 1  |  S11
                                          0 0 1 1  |  S18  S27  S6
                                          0 0 0 0  |  S2
                                          1 1 2 2  |  S7   S9
    -1                                          1  +  S12
                                              0 2 T|S S17
                                        0 0 0 0 1  |
```

Picture 1. ***Wright's map (Person-Item Map)***

The left wright map illustrates student ability, and the one on the right illustrates the distribution of logit values for item difficulty. From this picture, it can be concluded that the end-of-semester test given is categorized as easy if you look at the level of student ability above the level of difficulty of the question. Many students were able to answer the questions maximally, and all were correct because the students' abilities were higher than the level of difficulty of the questions.



Picture 2. Tes Information Function

Based on Picture 2, the X-axis describes the level of students' ability to take the PAI test, and the Y-axis is the magnitude of the information function of the PAI test. The results show that a moderate level of student ability with logit (0) produces very high information. Then, at a low or high ability level, it will produce low information. This means that the PAI test produces maximum information when given to individuals with moderate ability. The same scale shows the ability of each student to work on each question; the level of student ability is indicated by the logit score. A high logit score indicates a high level of ability in working on the PAI test. The level of individual ability can be seen from the statistical results (Pearson measure) of the Winstep analysis on Rasch. The individual ability level was 216 students, and the highest student results were found in students with numbers 143,110,113,128,158 and 19. The low ability level is in students with numbers 49,124, 82,55,52, and 30.

Table 1.
Student Ability Category Results
(*Pearson Ability*)

No	Kategori	Jumlah	Persentase
1	Very high	23	11%
2	High	154	71%
3	Low	39	18 %

Table 1. Explaining the level of student ability has a very high category of 11%, a high 71%, and a low category of 18%. Olsen explained in his research that a person is able to solve items with difficulty

levels gradually to solve a problem depending on the ratio between the person's ability and the difficulty level of the item (Olsen, 2003).

Pearson (respondent) is said to be suitable if it behaves consistently with what the model expects. This is in accordance with the item fit criteria: the outfit mean square (MNSQ) value, the outfit Z-standard (ZSTD) value, and the point measure correlation (Pt Mean Corr) value. Respondents are said to be suitable if the MNSQ value is 0.5<MNSQ<1.5, the ZSTD value is between -2<ZSTD<2, and the point measure correlation value is not negative and is at 0.4<Pt Measure Correlation < 0.85. The results of the model fit analysis can be seen in table 2 below:

Table 2.
Pearson Match Result

No	Criteria	Lots of People (*Pearson*)	
		Fit	**Tidak *Fit***
1	0,5<MNSQ<1,5	201 people	15 people
2	-2<ZSTD<2		
3	0,4<Pt*Measure Correlation*<0,85		
Percentage		93%	7%

Based on the results of the Winstep program analysis summarized in table 2. Of the 216 students, 93% fit the model (fit model), and these students are said to fit the model because they have met at least two existing criteria.

Furthermore, the level of item fit with the model is said to fit the model if it describes consistently with what is expected by the model. Item fit explains whether the items function normally to make measurements or not. If an item is found that does not fit, it is an indication that students have misconceptions about the item (Sumintono & Widhiarso, 2015). According to Boone, Staver, and Yale (2014), there are three criteria that must be met, namely the outfit mean square (MNSQ) value, the Z-standard outfit (ZSTD) value, and the point measure correlation (Pt Mean Corr) value. Respondents are said to be suitable if the MNSQ value is 0.5 < MNSQ < 1.5, the ZSTD value is between -2 < ZSTD < 2, and the point measure correlation value is not negative and is at 0.4 < Pt Measure Correlation < 0.85. Sumintono and Widhiarso (2015) say that the ZSTD value is strongly influenced by the sample size. So, when the sample size is large, it is inevitable that the ZSTD value has a range above 3. The summary of item suitability is presented in table 3. Below:

Table.3.
Item Match Results

No	Criteria	Number of Question Items (*Item*)	
		Fit	**Not Fit**
1	0,5<MNSQ<1,5	33 item	7 item
2	-2<ZSTD<2		
3	0,4<Pt *Measure Correlation*<0,85		
Percentage		83%	17%

The results of the analysis from the summary table above show that for the 40 items analyzed, there are 33 items (83%) that fit the model (fit model). It is said to fit the model because it has met two of the three criteria, namely the outfit mean square (MNSQ) value, the Z-standard outfit (ZSTD) value, and the point measure correlation (Pt Mean Corr) value (Boone et al., 2016).

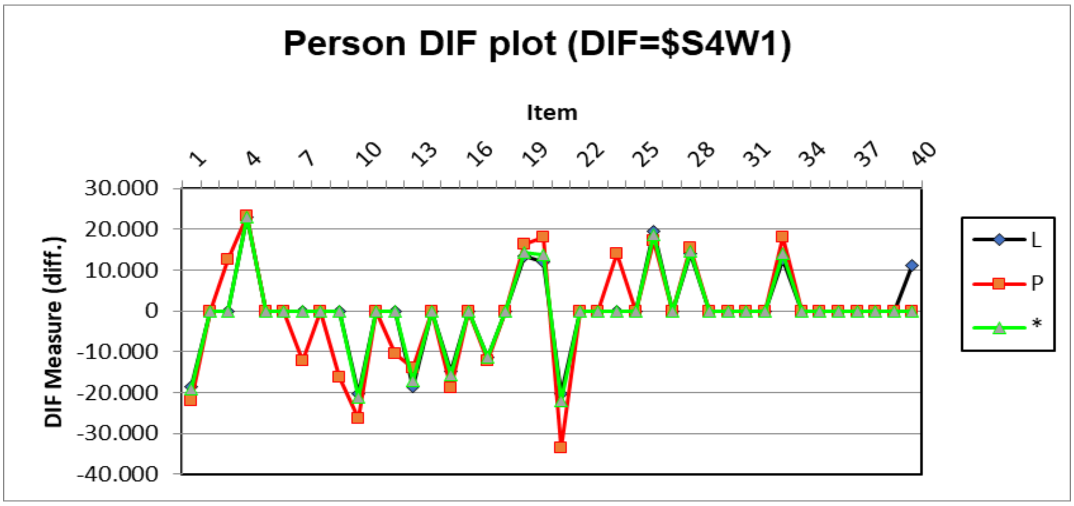
In items 20, 25, and 38, the item criteria are not met, so the questions are categorized as unreasonable and need to be corrected or replaced because they do not meet the three predetermined criteria. This is done with the aim that, in the future, it can produce more accurate test scores because they are tested with appropriate and quality items. For other items, they can be accepted and categorized as good questions because they have met the three predeterminedstandards.

Table.4.
Pearson and Item Reliability Results

	Separation	Reliability	Alpha Cronbach
Person	2.22	0.83	0.86
Item	5.80	0.97	

Table 4 shows that the Cronbach alpha value is 0.86, which states that there is an excellent interaction between Pearson and the items as a whole. While the reliability for Pearson's reliability is 0.83 and item reliability is 0.97, it can be concluded that the consistency of student answers is good, and the quality of the items in the PAI test has a special reliability aspect. At the same time, the separation value is to show the grouping of items and pearson. This shows that the item separation value is in a large category, which shows that the greater the separation value, the better the test quality in the overall respondent and item because it can identify the respondent group and the item group. The item separation value is 5.80; this indicates that the item consists of very difficult, difficult, medium, easy, and very easy items. At the same time, the Pearson category is categorized as 3, namely very high, high, and low ability students.

Rasch modeling analysis outputs can also see gender bias in an analysis to see the fairness of questions or tests that students do when viewed from gender. Item bias, which is more commonly known in item analysis, namely differential item functioning (DIF) in RASCH modeling, basically uses the Mantel Haenzel method. The presence or absence of DIF can be determined by looking at the resulting DIF analysis table. If there is a probability value <0.05 in the item, it can be said that the item contains DIF, which can harm certain gender groups.



Picture 3: Different Item Function

Upon examining the image above, it shows that item 1 harms women more than men, item 4 benefits men more than women, item 10 harms women more than men, item 19 benefits men more than women, item 22 harms women more than men, item 40 benefits men more. Gender bias analysis can be seen that 2 out of 40, namely items number 5 and 24, these items display probability results <0.05, so it can be said that the item contains DIF or needs to be corrected because it harms certain gender groups.

DISCUSSION

The purpose of this study was to determine the quality of teacher-made tests in PAI subjects. Based on some previous research results, many categorize teacher-made tests as weak and unable to measure learning outcomes and cannot be the basis for determining learning success. This is in accordance with research conducted by Eva D.K.S. that if the quality of the test is low, then this will be very detrimental to students in measuring ability (E. D. K. Sari 2017).

This can happen because the analytical tool used to analyze the items still uses classical measurements, so that it still depends on the number of samples available. However, if the item analysis is done with a more modern analysis, it can be anticipated that this is in accordance with the results of the study, which found that the PAI test made by the teacher is categorized as Good, although there are some questions that must be corrected and replaced. (E. D. kumala Sari, Rustam, and Yunita 2021; Rustam, Sari, and Yunita 2018; eva dwi kumala Sari, Tolla, and Margono 2018; 2019; eva dwi kumala Sari and Falani 2021).

This can happen because the analytical tool used to analyze question items still uses classical measurements, so that it still depends on the number of samples available. But if the item analysis is carried out with a more modern analysis, it can be anticipated that this is in accordance with the results of the study, which found that the PAI test made by the teacher is categorized as Good, although there are some questions that must be corrected and replaced.

The results showed that the test fit the model or item fit by 83%, and item and person reliability was also good. Pearson fit was also categorized as high at 71%. Pearson fit was 93%. It can be categorized that the teacher-made test is good and can be used to make measurements. Item analysis conducted in this study can also be used as evaluation material for improving the learning process.

CONCLUSIONS

This study has produced an analysis of PAI test instruments and can be used further to be used as an evaluation tool for PAI learning success. The item analysis conducted has resulted in a valid and reliable instrument. This research has gone through the item analysis procedure correctly with Rasch modeling and assisted by Winsteps software. The findings of the results of this study are as follows: 1) The PAI test instrument made by the teacher already has the feasibility to be used as a test to measure PAI learning outcomes. 2) The instrument has been reviewed from the level of difficulty, student ability, DIF, and Item Fit, and the results of the study have shown the suitability between the items and the students' abilities. The results of this study are expected to be an example for teachers to analyze the items developed by teachers, and teachers can continue to improve the quality of writing items for learning evaluation.

REFERENCES

Arikunto. S. (2016). Dasar-dasar Evaluasi Pendidikan. Jakarta : Bumi Aksara.

Azwar, S. (2012). Penyusunan Skala Psikologi edisi 2. Yogyakarta: Pustaka Pelajar

Azwar, S. (1993). "Kelompok subjek ini memiliki harga diri yang rendah"; kok, tahu...? Buletin Psikologi, I (2), 13-17

Bond, T. G., & Fox, C. M. (2013). Applying the Rasch model: Fundamental measurement in the human sciences. Psychology Press.

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). Rasch analysis in the human sciences. Springer Science & Business Media.

Cahyati. S. (2018) “Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Pendidikan Agama Islam”, Skripsi Sada Program PAI dan Ilmu Keguruan Purwokerto.

Djaali. 2017. Psikologi Pendidikan. Jakarta: PT Bumi Aksara.

Groves, Robert M., Survey Methodology (2010), Second edition of the (2004) first edition ISBN 0-471-48348-6

Halik, A.S. ettall. (2019) “Analisis Butir Soal Ujian Akhir Sekolah Mata Pelajaran Matematika,” Makassar: Al-Asma journal of Islamic Education, Vol. 1, No. 1.

Hamzah, A. (2014) Evaluasi Pembelajaran Matematika. Jakarta: PT Raja Grafindo Persada.

I Wayan Eka Mahendra. I. W.E. (2019). “Analisis Butir Soal,” Bali: Workshop Peningkatan Kompetensi Evaluasi Pembelajaran Guru.

Kristiana, I.F. ettall. (2018) Analisis Rasch dalam UtrechtWorkEngagemen (UWES-9) Versi Bahasa Indonesia. Jakarta: Jurnal Psikologi, Vol. 17 No.2.

Purwanto. (2009).Evaluasi Hasil Belajar, Yogyakarta: Pustaka Pelajar.

Purwanto. (2009). Evaluasi Hasil Belajar. Yogyakarta: Pustaka Pelajar.

Putra, Z.H. (2021). “Analisis Pengetahuan Matematika, Didaktika, dan

Teknologi Calon Guru Sekolah Dasar Menggunakan Rasch Model", Garut, Mosharafa: Jurnal Pendidikan Matematika, Vol.10, No. 3

Pratama, D. (2020). "Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch". Tarbawi: Jurnal Pendidikan Islam, vol.7, No.1.`

Rahmasari, D. & Ismiyati. (2016). "Analisis Butir Soal Mata Pelajaran Pengantar Administrasi Perkantoran," Semarang: Economic Education Analysis Journal, Vol. 5, No.

Rustam, A, Sari, E.D.K., Yunita, L. (2018). Statistika & Pengukuran Pendidikan (Analisis Menggunakan SPSS, Iteman dan Liserel).

Sari, E. D. K., Rustam, A., & Yunita, L. (2021). Pengembangan Instrumen Penelitian Sosial (Konsep, Tahapan dan contoh instrument, Analisis data menggunakan SPSS dan M-Plus, dan Winsteps) (1st ed.). Kun FayakunGenjong Kidul SidowarekNgoro Jombang Jawa Timur 61473.

Sari, E.D.K. ettall. (2021). " Kualitas Butir Soal Penilaian Akhir Semester (PAS) yang disusun Guru Madrasah". Jakarta: Jurnal Pendidikan Islam dan Ilmu Sosial, Vol. 30 No. 1.

Sari, E.D.K. (2017). " Isu-Isu Kritis dalam Pendidika, (Lemahnya Kualitas Tes Ujian Akhir sekolah berstandar Nasional (UASBN) ditingkat Sekolah Dasar ". Bekasi: Jurnal Al Marhalah, Vol. 1 No. 1.

Sari, E.D.K., Falani, I,. (2021) Developing Instrumen to Measure Elelmentary School Teacher 'Profesional Ethics in Indonesia'. Turki: elementary education online, 20 (2).

Sudijono. A. (2001). Pengantar Evaluasi Pendidikan. Jakarta: Raja Grafindo Persada.

Sugiyono. (2016). Statistika untuk penelitian. Alfabeta.

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

Vockell, E. L. (1983). Educational Research. Mac Millan Publishing Co., Inc.

https://www.semestapsikometrika.com/2018/07/membuat-kategori-skor-skala-dengan-spss.html diakses pada January 12, 2021

Wahyudin, W. (2016) "Pendidikan Sepanjang Hayat Menurut Perspektif Islam." Banten: Jurnal Kajian Keislaman, Vol. 3, No. 2.

Wahyudi, D. (2016). "Konsep al-Qur'an tentang Hakikat Evaluasi dalam Pendidikan Islam," Lampung: Hikmah, Vol. 12, No. 2.