Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

ANALYSIS OF MATHEMATICS QUESTIONS ON THEME 9 AT GRADE V OF RAUDHATUL RAHMAH ISLAMIC ELEMENTARY SCHOOL PEKANBARU

Eni Siskowati[1], Syahrul Ramadhan[2]

[1, 2] **Masters Program in Madrasah Ibtidaiyah Teacher Education at UIN Sunan Kalijaga, Indonesia**
[1]*enisiskowati@gmail.com*, [2]*syahrul.ramadhan@uin-suka.ac.id*

ABSTRACT

This paper discusses learning evaluation through test techniques, which are conducted on fifth-grade elementary school students on Mathematics learning subjects. The research is a type of quantitative descriptive research using the ex post facto method. The instrument used was 20 items of math subject matter questions, which were collected through a written test technique on students. The data obtained were analyzed using the Rasch Model assisted by the Winstep program through an empirical validity process, which includes validity, reliability, and difficulty level tests. The results indicate that in the validity test, items were invalid. In the reliability test, there were 2 reliabilities, Person Reliability was categorized as very good and Item Reliability was categorized as weak. Whereas on the results of the question difficulty level was categorized as very difficult. The implication of the research is that the research can be a reference for other researchers in conducting research on identifying and analyzing question items. When designing the questions, it is expected that a review will be carried out regarding the material that will be discussed on the test.

Keywords*: *rasch model, winstep, mathematics

ANALISIS 9 TEMA MATEMATIKA KELAS 5 SOAL MATEMATIKA DI SDIT RAUDHATUL RAHMAH PEKANBARU

ABSTRAK

Tulisan ini membahas evaluasi pembelajaran melalui teknik tes, yang dilakukan pada siswa kelas 5 SD pada mata pelajaran Matematika. Penelitian merupakan jenis penelitian deskriptif kuantitatif menggunakan metode *expost facto*. Instrumen yang digunakan adalah berupa 20 butir soal materi pelajaran matematika, yang dikumpulkan melalui teknik tes tertulis pada siswa. Data yang diperoleh dianalisis dengan menggunakan Model *Rasch* dengan dibantu program *Winstep* melalui proses validitas empiris, yang meliputi uji validitas, reliabilitas, dan tingkat kesukaran. Hasil penelitian menunjukkan bahwa pada uji validitas, terdapat item yang tidak valid. Pada uji reliabilitasnya terdiri dari 2 jenis, *Person Reliability* yang dikategorikan bagus sekali, sedangkan *Item Reliability* dikategorikan lemah. Sedangkan pada hasil tingkat kesukaran soal dikategorikan sangat sulit. Implikasi dari penelitian adalah penelitian dapat dijadikan sebagai referensi untuk peneliti lain dalam melakukan penelitian tentang penyusunan dan analisis butir soal. Pada saat perancangan soal, diharapkan dilakukan pengecekan kembali mengenai materi yang akan dibahas pada tes.

Kata Kunci: model rasch, winstep, matematika

Submitted	Accepted	Published
02 January 2023	25 March 2023	29 March 2023

Citation	:	Siskowati, E., & Ramadhan, S. (2023). Analysis Of Mathematics Questions On Theme 9 At Grade V Of Raudhatul Rahmah Islamic Elementary School Pekanbaru. *Jurnal PAJAR (Pendidikan dan Pengajaran), 7*(2), 397-406. DOI: http://dx.doi.org/10.33578/pjr.v7i2.9210.

INTRODUCTION

The purpose of education, based on Constitution No. 2 of 1985, is to educate the nation's youth. In addition, education is a means to assist a country in preparing human resources so that they have the competencies expected in the international market. This can be reviewed through PISA (Program for International Student Assessment) and TIMSS (Program for International Student Assessment) which evaluate the education system followed by more than 70 countries around the world. Sjoberg & Jenskin (2022) define PISA as one of two large-scale international student assessment comparison projects which are conducted every three years to test the academic performance of 15-year-old schoolchildren. Furthermore, Wiberg (2019) describes TIMSS as an international mathematics and science assessment, and the evaluation is

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

based on a framework created by participating nations for each curricular area and class.

The results of TIMSS and PISA have over the years been a snapshot of student learning outcomes around the world for improving education systems towards greater teacher academic capacity and student achievement. Fenanlampir, Batlolona, & Imelda (2019), explained the PISA results and TIMSS results in 2015, where the 2015 PISA results showed that 23% of students did not reach level 2 on the mathematics evaluation. Indonesia earned 399, 394 and 385 points. This shows that Indonesian students still have many shortcomings, so much improvement is needed to improve students' mathematical abilities. Meanwhile, the 2015 PISA results show that Indonesia still has great hopes to improve its ranking.

The quality of education which is often inappropriate and ineffective has become a problem in the implementation of education in this country, so a learning evaluation is carried out. Through this evaluation process it will affect changes in the quality of education in the future, as well as find out how students are capable, as well as find out whether the learning system planned by the teacher is appropriate or not. Like research conducted by Wiberg (2019) where, Sweden uses the TIMSS evaluation as a trend indicator to see how students perform over time at school, and helps in explaining the mathematics achievement of TIMSS students in Sweden. This was also explained by Aqmarani, Magdalen, & Ayudhiya (2020), where evaluation can be used as a basic component of the education system which is also used as a tool to measure the success of learning in schools.

The quality of the items can be seen from the ability of students to answer these questions, but it does not rule out that there are other factors that affect students' ability to answer math questions. As previous research by Wibowo and Agia (2020) found difficulties experienced by students and factors that causing students difficulty answering math questions caused by difficulty understanding concepts, difficulty in skills, and trouble solving problems. It is supported by research by Kumalasari et al. (2021) which found that students with difficulties in solving math problems are students who lack an understanding of a mathematical concept. It is common for students who lack this understanding to be lazy to work on issues, causing the value of the lesson to be less. The research of Destiniar et al. (2019) also explained that the ability to understand mathematical concepts is one of the factors in achieving proficiency or proficiency in working on math problems, namely by interpreting, classifying, explaining, formulating and calculating math problems accurately, efficiently and precisely. Furthermore, research by Azizah & Fitria (2019) shows that mathematical connection skills in students have a positive and significant influence. It means that students with high mathematical connection skills will indirectly have no difficulty solving math problems and relating them to other fields of science. In addition to students' understanding of concepts and connections regarding mathematics, one of the other factors that factor into students' difficulties in answering math problems is numerical skills. Research by Ramadhan et al. (2021) shows that students' numerical skills or abilities can affect students ability to solve math problems. Students with numerical skills will generally have an organized problem-solving mindset.

The measurement of students' abilities during evaluation requires a tool, and written and non-written tests are one of the instruments employed. In addition, the quality of the assessment instrument is an important component in creating a good testing system. Herwin, Tenriawaru, and Fane (2019) explain, the quality of the assessment instrument is an important component in creating a good testing system. If the test instrument has good quality, the measurement function on the test will run well and obtain the right test results and decisions. However, it does not rule out the possibility that a test kit can have weaknesses and shortcomings, as research conducted by Izmaimusah & Idris (2021) states that the test instrument used by teachers in Elementary Schools in Palu City is still very low in terms of achievement of cognitive abilities and does not use criteria. in compiling tests so that the instruments used are not able to reveal student competence. Therefore,

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

item analysis needs to be done to analyze students' abilities which can help teachers to be more effective in assisting the learning process.

Traditional and contemporary test theory can be used for item analysis. The difficulty and distinguishing index of the items based on the group of students working on them are two flaws in the traditional item analysis method. While standard error measurement (SEM) is used for all test takers, there is no fundamental theory to predict how participants will perform on tests that are pertinent to their aptitudes (Rizbudiani et al., 2021). As a result, the Rasch model and contemporary test theory's Item Response Theory were used to conduct the item analysis. Rasch model analysis enhances test and survey reliability and quality, making a variety of measuring tools possible (Al Ali & Shehab, 2020). Based on the descriptions above, this study aims to analyze the items by taking learning evaluations through test techniques for 5th grade students at SDIT Raudhaturrahmah Pekanbaru in the Theme 9 Examination in Mathematics. This item analysis research uses the Rasch model with the help of the Winstep program.

LITERATURE REVIEW

1. Evaluation on Mathematics Learning

In the process of learning mathematics, evaluation is needed by a teacher to find out the level of understanding of students in solving mathematical problems. Several things need to be explained regarding evaluation in learning mathematics, namely: (1) Evaluation is a process, not a product; (2) The goal of evaluation is to assess something's quality, particularly in terms of worth and significance; (3) A judgment must be made during the evaluation process; (4) Value and significance must be taken into account based on specific criteria (Yusuf, 2021).

One characteristic of evaluation is that it culminates in a decision. This choice is based on the value and benefits of the evaluation. While assessments have a wider reach than evaluations, the assessments seen only have a limited reach. The assessment is qualitative and quantitative. Since evaluation is an integral aspect of learning, it occupies a critical and strategic role. Almost all instructional system method professionals consider evaluation as one of the most important processes in the process (Hidayat & Asyafah, 2019).

Mathematics is a unique science that explores numbers, logic, space, shape, computation, and reasoning in a systematic, ordered manner. According to Waluyo, Muchyidin, and Kusmanto (2019), there are frequently several types of impediments that interfere with teaching and learning activities. The teacher's concepts may not be well understood by the students or are frequently referred to as misunderstandings, which is one of the challenges that students face during the learning process. unfocused learning Misconceptions impair learning and ultimately result in poor academic performance.

Things that cause changes in the concept according to Hashweh in Waluyo, Muchyidin, & Kusmanto (2019), among others; (1) the teacher does not know the student's preconceptions, (2) The teacher's evaluation strategy fails to put the student's understanding to the test, resulting in the incorrect response, (3) The teacher typically does not object to students' incorrectly prejudiced answers. Due to the necessity for learning evaluation, a teacher might offer a variety of exams in the form of written examinations or non-written tests to help students uncover their misconceptions. Furthermore, Wijaya (2021) revealed, through evaluating school learning can measure the quality of the abilities of its students, teachers can evaluate learning plans and teaching methods, as well as the goals of improving the quality of education in Indonesia.

2. Item Analysis

The application of the assessment system for teachers is how to measure changes in students' cognitive aspects at every level of education which is the responsibility of the school, therefore a valid and reliable measurement tool is needed to measure it (Idris & Uzmaimusah, 2021). Test questions may reveal details that can be used to create questions of a better standard. Prior to usage, each item will be examined and reviewed using data gathered from student replies as part of the item analysis process (Hermita et al., 2021).

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

Item analysis is the process of evaluating the quality of objective test items (things on essay tests or performance tests that are rarely reviewed) after they have been tried out on test takers. The activity of checking items is one of the activities that must be carried out by the instructor or teacher in order to involve the quality of the questions that have been made. Evaluators are often tasked with reviewing measuring instruments for student learning success. This analysis can be considered as an educational tool that determines students' competencies in educational improvement as well as gaps between educational goals and learning levels.

Multiple choice questions is a type of written question that is usually used in theory-based learning activities (Doust et al., 2021). Although multiple choice usually assesses a low level of knowledge, multiple choice can be increased to assess the level of knowledge, understanding, perception, application, and problem solving provided it is constructed appropriately. There are several benefits to multiple choice as follows: 1) multiple choice are more flexible than other types of questions and in addition to level of knowledge and justification ability, they assess student judgment, 2) MCQs can assess more of the educational goals and context of the course in a limited time period, 3) multiple choice questions can be tagged easily and graded electronically, 4) Compared to true/false question types, it's impossible to mark the correct answer by chance, and 5) if wrong answers are written correctly, multiple choice questions can diagnose students' misunderstandings and educational problems (Uduafemhe et al., 2021).

Designing questions and administering exams is often considered the final stage of coursework. However, to fulfill the educational process during the course, it is necessary to study and analyze quality items (questions). Therefore, item analysis is an integral component of course assessment which helps to observe item characteristics and improve the overall quality of the exam (Semiun et al., 2022). In other words, exam grading is a dynamic process aimed at improving questions and teaching.

Most studies have addressed the validity and reliability of exams, taxonomy of questions, index of difficulty and discrimination and, indeed, have suggested some suggestions for increasing exam levels. Finding the proper questions, reducing challenging questions, eliminating items with low discriminating power, boosting validity and reliability, and most significantly, applying the results for the upcoming semester are all ways to improve the quality of assessment based on item analysis. In addition, it provides a guide for improving teaching styles to improve student learning outcomes, because strengths and weaknesses as well as misunderstandings during teaching can be revealed by item analysis. Thus, item analysis is considered very important in the educational process.

3. Rasch Model

In the 1960s, Georg Rasch created the 1PL (one logistic parameter) analytical model of item response theory, also known as Item Response Theory (IRT) (Olsen in Marullo et al., 2022). Ben Wright later made this mathematical framework well-known. Rasch creates a model that links students and things using raw data that depicts students' abilities as dichotomous data (in the form of true and false).

a) Unidimensionality

The IRT model makes the basic assumption that a test's components measure only one ability, a concept known as unidimensionality (Hambleton, et al in Uzun & T, 2021). Because cognitive, personality, and test-taking factors constantly affect test results, at least somewhat, this premise cannot be strictly satisfied. These variables include degree of motivation, exam anxiety, quickness of work, propensity to guess when unsure of an answer, and cognitive abilities.

b) Reliability

There are two reliability indices for the Rasch Model measurement: person dependability and item reliability (Boone, et al in Brann et al., 2020). The value of person and item reliability is defined as follows: 0.67 is weak, 0.67-0.80 is sufficient, 0.81-0.90 is good, 0.91-0.94 is very good, and > 0 is regarded very good. 94 is special. Perfect consistency, that is, if the reliability value is 1, this cannot occur in measuring psychological and social aspects that use humans as subjects, because there are various sources of error in humans .

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

c) Fit Order Items

Fit order items explains whether the items function normally to measure or not about how accurately the data fits the ideal model with regard to suitability statistics, which consist of infit, outfit, mean-square, and standardized fit statistics (ZSTD). The sensitivity of the response pattern to the target item on the respondent, or vice versa, is known as infit (inlier-sensitive fit). Outfit (outlier-sensitive fit) assesses the responsiveness of 16 response patterns in respondents to items of varying degrees of difficulty, or vice versa. Items that do not fit indicate that there is a misconception among respondents about the items. Checking fit and misfit items can use the INFIT MNSQ value of each item; the mean and standard deviation are summed, then compared, a larger logit value indicates a misfit item .

d) Item Difficulty Index

The item difficulty index in the Rasch Model measurement can be seen through the logit value in the item measure table, and has been sorted from the highest to the lowest logit value . A high logit value indicates high item difficulty . Item difficulty index is symbolized by b. The value of b moves from -2 to +2 (Hambleton and Swaminathan in Al-zboon et al., 2021). The standard deviation value, when combined with the mean logit average in the item measure, allows the difficulty level of the item to be grouped. For instance, an item group with a 0.0 logit +1SD is considered challenging, one with a greater +1SD is considered extremely difficult, one with a 0.0 logit -1SD is considered easy, and one with a lower -1SD is considered very easy. Therefore, the objects are divided into four groups based on their difficulty.

e) *Differential Item Functioning* (DIF)

One measure to see whether a measurement is valid or not is the instruments and items used do not contain bias. Items and measurement instruments can be biased, if they are more in favor of individuals with certain characteristics, so that these items benefit certain parties and harm other parties. For example, an item that is easier to answer by a group of female respondents than a group of men means that there are items that are gender biased.

REASERCH METHOD

This study, which describes the caliber of the math items on 25 participants from SDIT Raudhaturrahmah Pekanbaru, is a sort of quantitative descriptive study. The method research used in this investigation is ex post facto method. As explained by Danuri, Maisaroh & Prosa (2021) ex post facto is research conducted after differences in the independent variables occur due to the development of a natural event. Further, the important thing in this approach is the absence manipulation of variables.

The subjects in this study were 25 fifth grade students at SDIT Raudhaturrahmah Pekanbaru. The instrument used in this study was in the form of math subject matter questions consisting of 20 items collected through a written test technique on students. Data analysis in this study used interactive methods because the activities in qualitative data analysis were carried out interactively and continued continuously until the data was saturated (Sugiyono in Alawiyah, Patmawati & Muhtadi, 2021). The data obtained were analyzed using the Rasch Model with the help of the Winstep program through an empirical validity process including testing validity, reliability, and level of difficulty. The item's difficulty level serves as the only input for the Rasch model. Meanwhile, according to Adi et al (2022) The analysis that may be performed using Winsteps, however includes Person Reliability, Item Reliability, Item Fit, Item Measure, and Person Fit.

RESULTS AND DISCUSSION

Validity results using the Rasch Model with the Winstep program. There are several categories if the question meets the criteria, including:

a) The Outfit Mean Square (MNSQ) received a value of 0.5 MNSQ 1
b) The value of the Z-Standard (ZSTD) outfit obtained was: -2.0 ZSTD +2.0
c) Point Measure Correlation Value (Pt Measure Corr): 0.4 to 0.85

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	PERSON
19	7	20	-.73	.82	2.96	3.7	9.90	3.3	A .51	.79	36.4	78.4	per19
24	8	20	-.03	.87	1.59	1.3	2.35	1.2	B .77	.83	72.7	83.1	per24
22	9	20	.80	.98	1.19	.5	2.10	1.1	C .84	.87	90.9	87.3	per22
16	8	20	-.03	.87	1.69	1.4	2.08	1.1	D .76	.83	72.7	83.1	per16
21	8	20	-.03	.87	1.51	1.2	1.91	1.0	E .78	.83	72.7	83.1	per21
23	12	20	4.54	1.27	1.80	1.3	.71	.2	F .82	.86	81.8	91.3	per23
17	9	20	.80	.98	1.67	1.1	1.14	.5	G .83	.87	72.7	87.3	per17
2	9	20	.80	.98	1.08	.3	.97	.4	H .86	.87	90.9	87.3	per02
4	9	20	.80	.98	1.08	.3	.97	.4	I .86	.87	90.9	87.3	per04
20	7	20	-.73	.82	.89	-.2	.53	.1	J .80	.79	72.7	78.4	per20
3	8	20	-.03	.87	.67	-.7	.39	-.2	K .86	.83	90.9	83.1	per03
10	8	20	-.03	.87	.67	-.7	.39	-.2	L .86	.83	90.9	83.1	per10
5	7	20	-.73	.82	.61	-1.1	.37	-.1	1 .83	.79	90.9	78.4	per05
6	7	20	-.73	.82	.61	-1.1	.37	-.1	k .83	.79	90.9	78.4	per06
15	7	20	-.73	.82	.61	-1.1	.37	-.1	j .83	.79	90.9	78.4	per15
18	6	20	-1.38	.81	.49	-1.6	.32	-.2	i .79	.74	90.9	79.4	per18
8	9	20	.80	.98	.41	-1.0	.21	-.5	h .91	.87	90.9	87.3	per08
1	10	20	1.88	1.11	.22	-1.3	.11	-.6	g .93	.89	100.0	91.0	per01
7	10	20	1.88	1.11	.22	-1.3	.11	-.6	f .93	.89	100.0	91.0	per07
9	10	20	1.88	1.11	.22	-1.3	.11	-.6	e .93	.89	100.0	91.0	per09
11	10	20	1.88	1.11	.22	-1.3	.11	-.6	d .93	.89	100.0	91.0	per11
12	10	20	1.88	1.11	.22	-1.3	.11	-.6	c .93	.89	100.0	91.0	per12
13	10	20	1.88	1.11	.22	-1.3	.11	-.6	b .93	.89	100.0	91.0	per13
25	10	20	1.88	1.11	.22	-1.3	.11	-.6	a .93	.89	100.0	91.0	per25
MEAN	8.3	20.0	.45	1.00	.88	-.2	1.08	.1			87.1	85.5	
S.D.	2.2	.0	1.74	.22	.69	1.3	1.97	.9			14.4	4.9	

Figure 1. Output Validity

Based on the above results, the Outfit Mean Square (MNSQ) score is above 1.5 in question numbers 19, 24, 22, 16, 21. Outfit Z-Standard score *(ZSTD)* there are values above 2 in question number 19. As for *the* Point Measure Correlation (Pt Measure Correlation) values in question numbers 2,4,3,18, 8, 1, 7, 9, 11-13 and number 25 so does not meet the criteria. For those items that do not meet the criteria, it is better to just delete them, which are used only for valid questions.

The measurement result information function of the instrument, which displays the normal curve's shape, attests to its high level of quality. Based on validity data using the Rasch model, information was obtained that in general the mathematics test instrument at SDIT Raudhaturrahmah Pekanbaru was dominated by good items. However, there are still some unsatisfactory items that need further attention and evaluation. Also, the validity analysis's findings revealed an SD value of 1.37 above 0.00, indicating that respondents tended to be able to provide accurate answers to questions. As explained by Herwin, Tenriawaru, & Fane (2019), analysis of item items using the *Rasch model* can support valid results and the ability of test takers to answer question items can be evaluated.

The criteria mentioned in the sentence are the criteria used to assess the validity of the questions given. The Outfit Mean Square (MNSQ) value is a measure that describes the score of each item. The accepted MNSQ score must be between 0.5 and 1.5. The Outfit Z-Standard (ZSTD) value is a value that measures how far an item's score is from the average value. The accepted ZSTD value must be between -2.0 and +2.0. Point Measure Correlation (Pt Measure Corr) is a value that measures consistency between item scores and overall scores. The accepted Pt Measure Corr value must be between 0.4 and 0.85. If these values meet the specified criteria, then the question can be said to be valid.

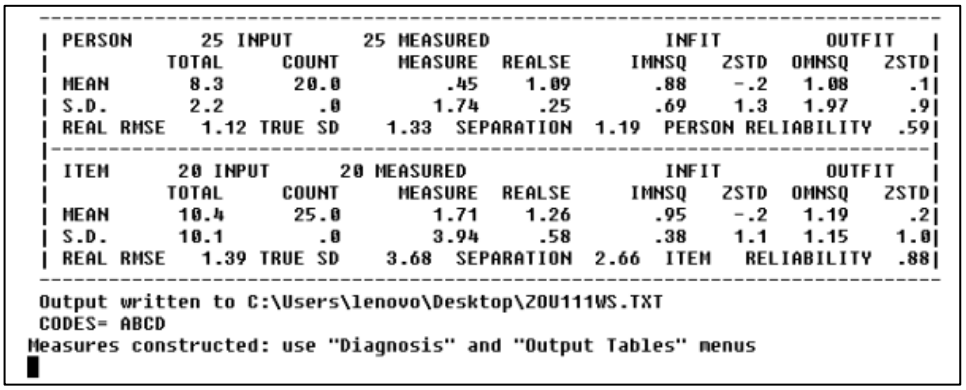
The comparison between the validity analysis and other research analyzes is that the validity analysis uses the Winstep Rasch Model Program to measure the level of validity of the questions given. The Outfit Mean Square (MNSQ) score is measured to reflect the score of each item and the accepted MNSQ score must be between 0.5 and 1.5. The Outfit Z-Standard (ZSTD) score is measured to measure how far an item's score is from the mean and the ZSTD value received must be between -2.0 and +2.0. The Point Measure Correlation (Pt Measure Corr) value is measured to measure the correlation between answers and questions and the Pt Measure Correlation received must be between 0.4 and 0.85.

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

Meanwhile, other research analysis, such as surveys or interviews, can measure social variables, behavior, and opinions using a Likert scale, numerical scale, interval scale, and other scales. This analysis can be used to determine the validity level of surveys or interviews using methods such as qualitative analysis, descriptive analysis, and regression analysis. This makes it possible to determine the variables that influence the behavior and opinions of respondents.

1. Reliability Test

The purpose of the reliability test is to evaluate the measuring instrument's consistency and overall quality. Person Reliability and Person Reliability provide the findings of this reliability test.



PERSON	25 INPUT	25 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	8.3	20.0	.45	1.09	.88	-.2	1.08	.1
S.D.	2.2	.0	1.74	.25	.69	1.3	1.97	.9
REAL RMSE	1.12	TRUE SD	1.33	SEPARATION	1.19	PERSON RELIABILITY		.59

ITEM	20 INPUT	20 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	10.4	25.0	1.71	1.26	.95	-.2	1.19	.2
S.D.	10.1	.0	3.94	.58	.38	1.1	1.15	1.0
REAL RMSE	1.39	TRUE SD	3.68	SEPARATION	2.66	ITEM RELIABILITY		.88

Output written to C:\Users\lenovo\Desktop\ZOU111WS.TXT
CODES= ABCD
Measures constructed: use "Diagnosis" and "Output Tables" menus

Figure 2. Output Reliability

The result of Person Reliability is 0.59, this means that this category is weak, for the suitability of students in answering questions. These results indicate that the ability of SDIT Raudhaturrahmah Pekanbaru students in answering math questions is relatively weak. Furthermore, the result of Item Reliability is 0.88 which means that the category is very good, this means the quality of the items, or the reliability of the items. As stated by Adi et al (2022) explained that, if the reliability value of the person is lower than the reliability of the item, this indicates that the consistency of student answers is weak, but the quality of the items on the instrument is very high.

2. Degree of Difficulty

The difficulty level can be seen from the item measure, the category of difficulty level using the Rasch Model, namely:

a) > +1.37 SD Very Hard category
b) 0.0 logit – 1.37 SD in the Difficult category
c) 0.0 logit 1.37 SD Medium category
d) < +1.37 Easy category

The following is the output of the Winstep program :

Table 1. Recapitulation of Problem Difficulty

Question Number	Measures	Category
22	0.82	Easy
4	1.62	Very difficult
15	2.16	Very difficult
24	1.19	Difficult
6	0.49	Easy
11	1.62	Very difficult
8	2.98	Very difficult

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

Question Number	Measures	Category
20	2.98	Very difficult
3	2.16	Very difficult
17	2.16	Very difficult
21	2.98	Very difficult
2	0.82	Easy
14	0.82	Easy
7	2.98	Very difficult
25	1.62	Very difficult
23	0.49	Easy
12	1.19	Difficult
16	1.62	Very difficult
9	1.62	Very difficult
10	2.16	Very difficult
19	1.62	Very difficult
18	2.98	Very difficult

The standard deviation in this study is 1.37. Value > +1.37 SD is a question with a very difficult category, so questions number 3, 4, 7, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, and 25 are questions with very difficult category. Values between 0.0 logit + 1.37 SD can be categorized as difficult questions, so questions 12 and 24 are difficult questions. Values between 0.0 logit 1.37 SD can be categorized as questions in the moderate category, but there are no questions with the moderate category. Furthermore, scores < +1.37 can be categorized as questions with an easy level, so questions number 2, 6, 14, 22, and 23 are easy questions.

Based on the output results above, the analysis of the items from the average difficulty level of the questions is categorized as very difficult because the results of the measure are more than 1.37. This shows that the math questions used are not in the category of good questions, because most of the items are at a very difficult level of difficulty. The results of the data analysis using the Rasch model provide information to teachers about the level of student ability and the different response patterns of each student in solving math problems. Knowing the skills of these students can make it easier for teachers to evaluate and plan further learning. In this case, the teacher can also improve the quality of the questions used in the assessment material based on the analysis results obtained. Good question items have a category of difficulty or difficulty level in the question items, namely 25% for easy questions, 25% for complex questions and 50% for moderate questions (Tyas et al., 2020). It is in accordance with the research of Hermita et al (2021) explained that good questions should have a moderate level of complexity and should not be either too easy or too difficult. Furthermore, Kurniasi et al. (2020) claimed that while difficult questions can demotivate students since they are more difficult to complete, easy questions can make it difficult for them to stimulate their critical thinking.

CONCLUSIONS AND RECOMMENDATION

Based on the results of analysis and data processing through the Winstep program, it can be concluded that in the validity test, there are items that are invalid, the reliability test consists of 2 types, namely Person Reliability which is categorized as very good, while Item Reliability is categorized as weak. Whereas on the results of the difficulty level of the questions, the category

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

is very difficult. So it is recommended, when designing questions, it is hoped that it will be checked again regarding**.**

REFERENCES

Adi, N. R. M., Amaruddin, H., Adi, H. M. M., & A`yun, L. Q. (2022). Validity and reliability analysis using the Rasch model to measure the quality of mathematics test items of vocational high schools. *Journal of Educational Research and Evaluation*, *11*(2), 103–113. http://journal.unnes.ac.id/sju/index.php/jere

Al-zboon, H. S., Alrekebat, A. F., & Bani Abdelrahman, M. S. (2021). The effect of multiple-choice test items' difficulty degree on the reliability coefficient and the standard error of measurement depending on the item response theory (IRT). *International Journal of Higher Education*, *10*(6), 22. https://doi.org/10.5430/ijhe.v10n6p22

Al Ali, R. M. A., & Shehab, R. T. (2020). Psychometric properties of social perception of mathematics: rasch model analysis. *International Education Studies*, *13*(12), 102. https://doi.org/10.5539/ies.v13n12p102

Andini Aqmarani, I. M. dan N. A. (2020). Evaluasi pembelajaran pada tingkat sekolah dasar andini. *Cerdika: Jurnal Ilmiah Indonesia*, *1*(2), 57–63. http://cerdika.publikasiindonesia.id/index.php/cerdika/indeks

Azizah, M & Fitria. (2019). Pengaruh kemampuan koneksi matematika terhadap hasil belajar siswa pada materi garis singgung lingkaran kelas VIII SMPN 2 Sumbergempol Tulungagung. *Jurnal Pendidikan Matematika RAFA*, *5*(1), 1-9.

Brann, K. L., Boone, W. J., Splett, J. W., Clemons, C., & Bidwell, S. (2020). Development of the school mental health self-efficacy teacher survey using rasch analysis. *Journal of Psychoeducational Assessment*, *39*(2). https://doi.org/10.1177/0734282920947504

Doust, A. R., Khan, W. A., & Al-Ghafri, M. (2021). An item analysis study on timss 2015 mathematics items of omani and iranian students comparison irt and cdm approaches. *International Journal of Mathematics Trends and Technology*, *67*(9), 87–95. https://doi.org/10.14445/22315373/ijmtt-v67i9p510

Destiniar, D., Jumroh, J., & Sari, D. M. (2019). Kemampuan Pemahaman Konsep Matematis ditinjau dari self efficacy siswa dan model pembelajaran Think Pair Share (TPS) di SMP Negeri 20 Palembang. *JPPM (Jurnal Penelitian dan Pembelajaran Matematika)*, *12*(1), 115-128.

Fenanlampir, A., Batlolona, J. R., & Imelda, I. (2019). The struggle of indonesian students in the context of timss and pisa has not ended. *International Journal of Civil Engineering and Technology*, *10*(2), 393–406.

Firdaus, F., Zulfadilla, Z., & Caniago, F. (2021). Research methodology : types in the new perspective. *Manazhim*, *3*(1), 1–16. https://doi.org/10.36088/manazhim.v3i1.903

Hermita, N., Sakinah, S., Wijaya, T. T., Vebrianto, R., Alim, J. A., Putra, Z. H., Fauza, N., Dipuja, D. A., Pereira, J., & Jihe, C. (2021). Item analysis of heat transfer concept using rasch model in elementary school. *Journal of Physics: Conference Series*, *2049*(1). https://doi.org/10.1088/1742-6596/2049/1/012058

Herwin, H., Tenriawaru, A., & Fane, A. (2019). Math elementary school exam analysis based on the Rasch model. *Jurnal Prima Edukasia*, *7*(2), 106–113. https://doi.org/10.21831/jpe.v7i2.24450

Hidayat, T., & Asyafah, A. (2019). Konsep dasar evaluasi dan implikasinya dalam evaluasi pembelajaran pendidikan agama islam di sekolah. *Al-Tadzkiyyah: Jurnal Pendidikan Islam*, *10*(1), 159–181. https://doi.org/10.24042/atjpi.v10i1.3729

Jurnal PAJAR (Pendidikan dan Pengajaran)
Volume 7 Nomor 2 March 2023 | ISSN Cetak : 2580 - 8435 | ISSN Online : 2614 - 1337
DOI : http://dx.doi.org/10.33578/pjr.v7i2.9210

Idris, M., & Uzmaimusah, D. (2021). Instrument development of mathematics learning outcomes by the Rasch model in elementary school to support the implementation of the 2013 curriculum. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, *7*(2), 88–102. https://doi.org/10.21009/jisae.v7i2.21830

Kurniasi, E. R., Y, Y., & Karennisa, F. (2020). Analisis Soal ulangan harian matematika kelas IX SMP negeri 1 toboali. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, *12*(1), 43–52. https://doi.org/10.37640/jip.v12i1.276

Kumalasari, M., Solfema, S., & Fauzan, A. (2021). Pengaruh kemampuan Membaca dan Motivasi Belajar terhadap Pemecahan Soal Matematika di Sekolah Dasar. *Jurnal Basicedu*, *5*(2), 997-1005.

Marullo, C., Ahn, J. M., Martelli, I., & Di Minin, A. (2022). Open for innovation: An improved measurement approach using item response theory. *Technovation*, *109*.

Ramadhan, M., Suaedi, S., & Ilyas, M. (2021). Pengaruh Kemampuan Numerik dan Kecerdasan Emosional Terhadap Pemecahan Masalah Matematika Siswa Kelas VIII SMP di Kecamatan Latimojong. Pedagogy: Jurnal Pendidikan Matematika, 6(2), 139-148.

Rizbudiani, A. D., Jaedun, A., Rahim, A., & Nurrahman, A. (2021). Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables (SLETV). *Al-Jabar : Jurnal Pendidikan Matematika*, *12*(2), 399–412. https://doi.org/10.24042/ajpm.v12i2.9939

Semiun, T. T., Wisrance, M. W., & Helentina, M. (2022). English summative tests : the quality of its items. *English Education: Journal of English Teaching and Research (JETAR)*, *7*(2), 119–127.

Sjøberg, S., & Jenkins, E. (2022). PISA: a political project and a research agenda. *Studies in Science Education*, *58*(1), 1–14. https://doi.org/10.1080/03057267.2020.1824473

Tyas, E. H., Hamdu, G., & Pranata, O. H. (2020). Analisis Soal Pilihan Ganda dengan Menggunakan Pemodelan RASCH untuk Mengukur Kemampuan Siswa dalam Mengurutkan Bilangan Pecahan di Sekolah Dasar. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 7(2), 1-12.

Uduafemhe, M. E., Uwelo, D., John, S. O., & Karfe, R. Y. (2021). Item analysis of the science and technology components of the 2019 basic education certificate examination conducted by national examinations council. *Universal Journal of Educational Research*, *9*(4), 862–869. https://doi.org/10.13189/ujer.2021.090420

Uzun, Z., & T, Ö. (2021). Test equating with the rasch model to compare pre-test and post-test measurements. *Journal of Measurement and Evaluation in Education and Psychology*, *12*(4). https://dergipark.org.tr/tr/pub/epod/issue/67388/957614

Waluyo, E. M., Muchyidin, A., & Kusmanto, H. (2019). Analysis of students misconception in completing mathematical questions using certainty of response index (cri). *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah*, *4*(1), 27–39. https://doi.org/10.24042/tadris.v4i1.2988

Wiberg, M. (2019). The relationship between TIMSS mathematics achievements, grades, and national test scores. *Education Inquiry*, *10*(4), 328–343. https://doi.org/10.1080/20004508.2019.1579626

Wijaya, T. T. (2021). An analysis of the online final examination items for ninth-graders in the mathematics course using the Rasch measurement model. *Preprints*, *September*, 2021090510. https://doi.org/10.20944/preprints202109.0510.v1

Yusuf, R. (2021). Analisis kualitas butir soal ujian nasional mata pelajaran matematika sekolah menengah pertama. *Journal of Didactic Mathematics*, *1*(3), 158–164. https://doi.org/10.34007/jdm.v1i3.417.