International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 16, Issue 2, 2022

Rasch Model To Analyze Item Quality And Ability Of Fourth Elementary School Students

Tatan Zenal Mutakin[1*], Burhanuddin Tola[2], Bahrul Hayat[3], [1]Indraprasta PGRI University, [2]State University of Jakarta, [3]Jakarta State Islamic University, Email: zmtatan74@gmail.com[*] and tatanzenalmutakin_7817158057@mhs.unj.ac.id

In 2019, Indonesian nation did not participate at Trends in International Mathematics and Science Study (TIMSS) because of not getting good achievements in several time. This study aims to analyze the fourth elementary school mathematics literacy items in the 2019 TIMSS framework. The study held in DKI Jakarta involving 407 public elementary school students. The test items analyze used the Rasch Model with Winstep program version 4.4.7. The results, from 47 test items answered by students, were 23 items fit the model and had moderate of difficulty levels. Besides, the level of students' ability in test items was medium and low level categories. This means that the fourth-level public elementary school students had good abilities in answering the items medium and low levels. While, test items high level could not be well-answered. To improve students' ability in answering high-order thinking skills (HOTs) test items very well, student shoud do regular exercises so that they get used to these HOTs test items.

Key Word**:** *Quality Item, Students' Ability, and Rasch Model*

Introduction

The development of science and technology (IPTEK) has impact on changes in the order of people's lives, such as geographical aspects and human resources (HR) aspects. Changes in geographical aspects have impact on openness between one country and another, regional boundaries make it easy to cooperation between countries. That was known as the global era. Meanwhile, changes in the human resources aspect have competition in various aspects of life, including in education.

Education has important role in producing quality human resources. Quality human resources have good reasoning skills in analyzing, synthesizing, solving problems and can provide solutions to life's problems. This reasoning ability known as higher order thinking skills (HOTs) part of the competencies needed in the global era or the 21st century.

In education, mathematics is one subjects that can foster higher order thinking skills (HOTs) (Dinni, 2018). Ressefendi ET said that mathematics is science that emphasizes reasoning. Mathematics is the result of the human mind related to ideas, processes, and reasoning (Rahmah, 2013). However, in the 21st century, mathematical ability not only ability to count or answer questions form of symbols or numbers. The expected mathematical ability is a person's skill to formulate, apply, and interpret mathematics in various life contexts by using appropriate reasoning with concepts, procedures, facts, and tools. It is intended to describe, explain, or predict phenomena that occur in the world in meet the needs of society. This mathematical ability is known as mathematical literacy (OECD, 2017), (Rizky & Priatna, 2018)

Mathematical literacy is a barometer of a country's educational success. Countries whose students have high mathematical literacy will produce excellent human resources and can compete in the global world. International studies such as PISA and TIMSS examine the world's mathematical literacy skills. Both of these studies examine mathematical literacy. Indonesia participated in the PISA study from 2000 to 2018, while for the TIMSS study, Indonesia participated from 1999 to 2015 (Sutrina, 2021), (Hadi & Novaliyosi, 2019).

The results two studies explain that Indonesian students have not been able to compete with developed countries and have low abilities. In 2018, the results of the PISA study explain the literacy ability of Indonesian students was at level 1. In 2015, the results of the TIMSS, Indonesian students obtained a score of 397. The results two studies explain that the mathematical literacy ability of Indonesian students was low level (OECD, 2019), and (IEA , 2015). There are many factors that cause the backwardness of Indonesian students' mathematical literacy skills compared to developed countries, including; weaknesses in content standards, process standards, and evaluation standards. In this study, the author only examines the aspects of evaluation standards.

Learning evaluation is part of the implementation of content standards and process standards. Evaluation is effort to improve the learning process and as a decision-making tool. Evaluation is carried out to assess learning outcomes. Assessment in evaluation is carried out in three forms: 1) assessment of learning (AoL), which aims to measure student achievement of predetermined competencies, such as: National Examinations, School Final Examinations, International Studies, etc., 2) assessment for learning (AfL), is assessment to improve the learning process, such as: assignments, presentations, projects, etc., and 3) assessment as learning (AaL), is assessments that involve students to be able to see the results of their

learning achievements and progress in determining their learning targets, such as peer assessment (Dirjen Dikdas, 2017: 8), (Setiawati at all, 120-121).

The teacher's ability to assess is the key to being able to measure objectively the success of students in following the lessons given. However, the reality is that there are still many teachers who have not been able to carry out the assessment process properly, especially in assessing higher order thinking skills. The results of the study explain that: 1) 91.43% of teachers have understood the concept of higher-order thinking, 2) 82.86% have implemented higher-order thinking-based learning, and 3) 79% of teachers have difficulty in designing and implementing higher-order thinking-based evaluations. (Neat, 2018). Another research result states that 66.4% of teachers know aspects of the content domain and cognitive domain in the curriculum, 78.49% of teachers design and test the questions to be tested. However, 64.52% of teachers did not apply cognitive domain analysis to the questions to be tested. This has an impact on the low quality of test questions, 88.45% are categorized as low-level thinking, 11.55% are categorized as intermediate-level thinking, and 0% are categorized as high-level thinking (Mutakin & Hakim, 2019).

Based on the background and theoretical study that has been described it encourages the author to take a research entitled “Rash Model for Analyzing Item Quality and Ability of Fourth Elementary School Students. The objectives of this study are: 1) to analyze the fourth grade elementary school mathematics literacy test questions based on the 2019 TIMSS framework, 2) to find out the actual abilities of public elementary school students in DKI Jakarta on the questions developed.

Research Method

This research is part of the research development process for the low grade elementary school mathematics literacy assessment was developed with reference to the 2019 TIMSS Framework. The research respondents were fourth and fifth grade public elementary school students or aged between 10 or 11 years located in 6 areas of DKI Jakarta: Central Jakarta, South Jakarta, East Jakarta, West Jakarta, North Jakarta, and the Thousand Islands. The total respondents was 407 students.

Students who become respondents directly answer items test given. The questions given to students were 47 items consisting of 42 multiple choice items and 5 simple description items. The items have been through content validation by 9 experts. The grid of questions given refers to the 2019 TIMSS framework as follows:

Table 1. Developed Questionnaire Grid

Domain Conten	Domain Cognitif			Total
	Knowing	*Applying*	*Reasoning*	
Number	11	12	4	27
Measurement & Geometry	5	8	1	14
Data	2	1	3	6
Total	18	20	8	47

After the items are done by students, then psychometric testing is carried out on the results of the student's answers. Psychometric testing uses the Rasch model (Sumintono & Widhiarso, 2015) with Winstep program version 4.4.7.

Rasch Model Item Analysis and Discussion
Rasch Model Item Analysis

The Rasch model is one of the most popular item response theory (IRT) models. For dichotomy data, Rasch modeling combines an algorithm that expresses the results of the probabilistic expectation of item "i" and respondent "n" which is mathematically expressed as follows:

$$P_n(x_n = 1|b_n, d_i) = \frac{e^{(bn-di)}}{1+ e^{(bn-di)}}$$

Where $P_n(x_n = 1|b_n, d_i)$ is the probability of response n in item i to produce a correct answer (x = 1) with the respondent's ability, βn and level of difficulty item δi. The item analysis developed in this study uses the Rasch model with Winstep program version 4.4.7. The stages of analysis are carried out as follows:

Person and Item Compatibility Level

The level of suitability of person and item is useful to see the accuracy of person and item with the model or person fit and item fit. The criteria used to see the level of suitability of persons and items by looking at the value of outfit means-square (MNSQ), outfit z-standard (ZSTD), and point measure correlation (Pt. Mean. Corr) with the following conditions: 1) Outfit Mean Value Square (MNSQ) accepted: 0.5 < MNSQ < 1.5, 2) Value of Outfit Z.Standard (ZSTD) accepted: -2.0 < ZSTD < +2.0, 3) Value of Point Measure Correlation (Pt .Mean Corr): 0.4 < Pt. mean. Corr < 0.85.

Boone, Stave & Yale mention that if a person or item does not meet these three criteria, it is better to discard the person or item (for the person) or repair it (for the item). Meanwhile, Sumintono & Widhiarso (2015: 72) state that if one of the three criteria does not meet, then the person or item does not need to be discarded or replaced.

Person Compatibility Level

The suitability level of the person is useful to see the suitability of the person with the model or person fit. To find out the level of suitability of the person, it can be seen from the summary of the output of Winstep version 4.4.7 table 6.1. The following is a summary of the results of the person suitability test:

Table 2. Recapitulation *Person Fit*

Person Fit	174
Person Mis Fit	233
Total	407

Summary table 2 explains that of the 407 students who worked on the test questions, 174 students did mis fit the Rasch model so they were deleted and 233 students fit the Rasch model.

Item Compatibility Level

The level of item suitability serves to determine the accuracy of the item with the model or item fit. Item fit also serves to determine whether the items can function normally in taking measurements or not. Questions that do not fit indicate a mis understanding of the subject in answering the question. To determine the level of suitability of items, it can be seen from the output of Winstep version 4.4.7 table 10.1. The following is a summary of the results of the item suitability test:

Table 3. Recapitulation *Item Fit*

Item Mis Fit	21
Item Fit	26
Total	47

Summary table 3 explains that of the 47 test items answered by the students, 21 questions did not fit the Rasch model so they were deleted and 26 items met the Rasch model.

Instrument Statistics Summary

The instrument statistical summary serves to determine the set of exam questions given to respondents. Summary statistics can be generated from the output of Winstep version 4.4.7 table 3 Summary Statistics are described in the following figure 1:

Figure 1. Instrument Statistical Summary

```
TABLE 3.1 DATA PAKET-2.xlsx                    ZOU039WS.TXT  May  4 2022  7:51
INPUT: 407 PERSON  47 ITEM  REPORTED: 233 PERSON  26 ITEM  2 CATS WINSTEPS 4.4.7

      SUMMARY OF 232 MEASURED (NON-EXTREME) PERSON
-------------------------------------------------------------------------------
|          TOTAL                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      11.8      26.0        -.12     .50      1.00    .04    .98    .01 |
|  SEM        .3       .0         .07     .01       .01    .05    .01    .05 |
| P.SD       4.2       .0        1.03     .08       .16    .82    .22    .78 |
| S.SD       4.2       .0        1.04     .08       .16    .82    .23    .78 |
| MAX.      23.0      26.0        3.44    1.04      1.47   2.19   1.53   1.94 |
| MIN.       3.0      26.0       -2.70     .46       .60  -2.21    .37  -2.03 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .52 TRUE SD     .89  SEPARATION  1.71  PERSON RELIABILITY  .74 |
|MODEL RMSE    .51 TRUE SD     .90  SEPARATION  1.78  PERSON RELIABILITY  .76 |
| S.E. OF PERSON MEAN = .07                                                   |
-------------------------------------------------------------------------------
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .75  SEM = 2.12

SUMMARY OF 23 MEASURED (NON-EXTREME) ITEM
-------------------------------------------------------------------------------
|          TOTAL                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     110.0     233.0         .00     .15      1.00    .06    .98    .00 |
|  SEM        .0       .0         .00     .00       .01    .17    .02    .19 |
| P.SD       .0       .0         .00     .00       .00    .00    .00    .00 |
| S.SD       .0       .0         .00     .00       .00    .00    .00    .00 |
| MAX.     110.0     233.0         .00     .15      1.00    .06    .98    .00 |
| MIN.     110.0     233.0         .00     .15      1.00    .06    .98    .00 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .00 TRUE SD     .00  SEPARATION  .00  ITEM   RELIABILITY  .00 |
|MODEL RMSE    .00 TRUE SD     .00  SEPARATION  .00  ITEM   RELIABILITY  .00 |
| S.E. OF ITEM MEAN = .00                                                     |
-------------------------------------------------------------------------------
```

Summary figure 1. explains that the number of people who answered the test items were 407 people with 47 test items. While those who fit the Rasch model are 233 people and 26 items. Person measure value = -0.12 indicates that the average value of all students in working on all the questions given. The average value is smaller than the logit value = 0.0 indicating a tendency for students' abilities to be smaller than the level of difficulty of the questions.

Cronbach's alpha value = 0.75 explains that the overall reliability of the interaction between person and item shows a good value. The value of person reliability = 0.74 and item reliability = 0.97. This shows that the consistency of students' answers is sufficient, while the quality of the items in the instrument has a special quality.

Item-Person Map (Wright Map)

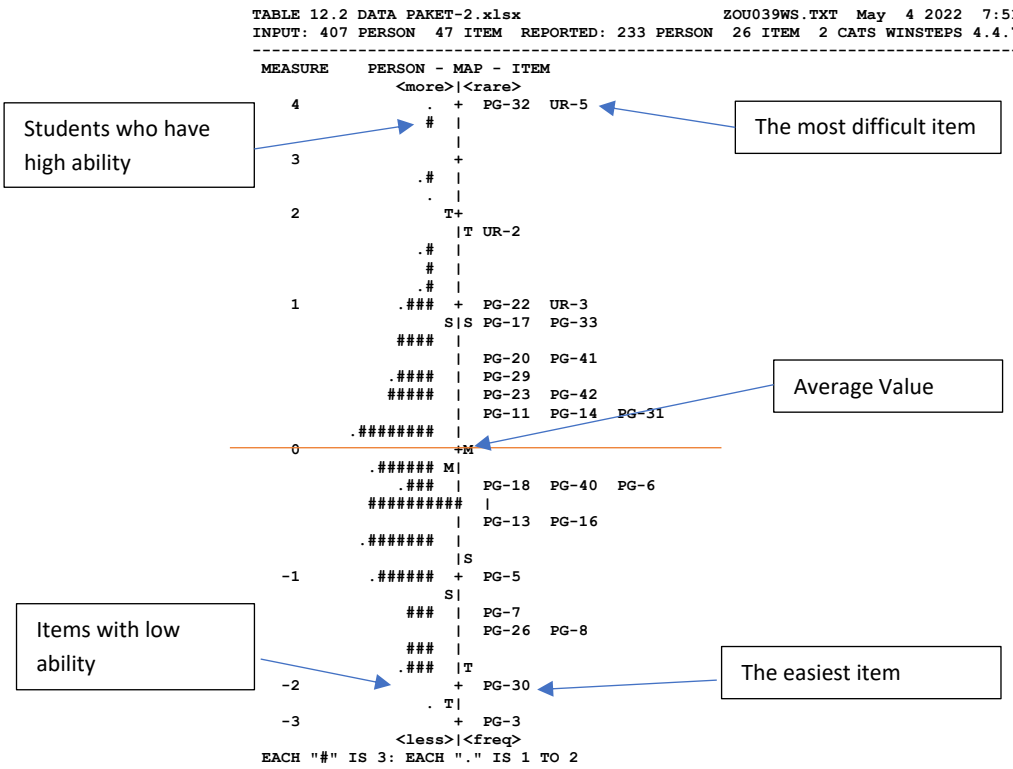
The item-person map consists of two parts, the left and the right. The left side describes the student's ability and the right describes the item's level of difficulty. In addition, there are horizontal lines that describe the same scale range or logit. The middle value is the average logit value with a range of -4 to +4.

The average logit value of the item is always set in 0.0 logit which indicates the initial reference point of the scale. The average logit person follows the starting point of the scale,

which is 0.0 logit. If the average logit person value is below 0.0 logit, it indicates the student's ability is below the average standard difficulty level of the question and vice versa, if the average logit person value is above 0.0 logit, it indicates the student's ability is above the average. standard difficulty level of the question. The item-person map in the Winstep program is depicted in table 12.2. Item-Person Map as follows:

The Item-Person Map table explains that the scale for persons and items is between -3 to +4. The higher the location of the item or person number, the more difficult it is (for the item) and the higher the ability (for the person) and vice versa. The lower the location of the item or person indicates the easier it is (for the item) and the lower the ability (for the person). The following is an explanation of the two parts of the Item-Person Map table

Figure 2. *Item-Person Map*



Item Difficulty Level

The item difficulty level analysis aims to determine the quality of the items from the easiest to the most difficult. In the Rasch model, the difficulty level of the item is obtained from the logit value (measure), the higher the measure value, the higher the difficulty level of the item and vice versa. Sumintono and Widhiarso (2015) provide guidance in assessing items into four categories, namely: 1) Measure value < -1 (very easy item), 2) Measure value -1 to d. 0 (easy item), 3) Measure value 0 to d. 1 (difficult item), 4) Measure value > 1 (very

difficult item). The following is a summary of the item difficulty levels obtained from the Winstep program table 13.1.

Table 4 explains that there are 2 items that have a very high level of difficulty (maximum measure), namely multiple choice questions number 32 and description questions number 5. The logit value of the two items is 7.24. Meanwhile, multiple choice question number 3 is a question that has the easiest level of difficulty (minimum measure) with a logit value of -7.22. The three items can be accepted by the Rasch model but are not good for use. The following is a summary of the difficulty levels of the items that have been sorted from the most difficult to the easiest.

Table 4. Summary of Item Difficulty Logit Value

Serial number	Item Number	Logit Value	Serial number	Item Number	Logit Value
1	32	7.24	14	14	0.24
2	Ur-5	7.24	15	31	0.24
3	Ur-2	1.88	16	6	-0.22
4	22	1.03	17	18	-0.22
5	Ur-3	1.03	18	40	-0.24
6	33	0.9	19	13	-0.66
7	17	0.88	20	16	-0.68
8	41	0.64	21	5	-1.01
9	20	0.59	22	7	-1.27
10	29	0.46	23	8	-1.35
11	23	0.43	24	26	-1.35
12	42	0.35	25	30	-1.98
13	11	0.3	26	3	-7.22

To find out the level of difficulty of the item, the logit values of the item can be entered into the Rasch model scaling. The following is a summary of the Rasch model scaling based on item difficulty level data.

Table 5 explains that the difficulty level of the majority of items is in the medium and low levels. There are 2 items in the very difficult level and 1 very easy item in the maximum and minimum measure categories. This corresponds to the view in the Item-Person Map.

Table 5. Item Difficulty Scale Summary

No	Scaling Model	Transformation Result Scale	Ability Level	Total
1.	X > M + 2SD	X > 5,18	Very difficult	2
2.	M + 2SD ≤ X < M + 1SD	5,18 ≤ X < 2,59	Difficult	0
3.	M + 1SD ≤ X < 0	2,59 ≤ X < 0	Medium	13
4.	0 ≤ X < M - 1SD	0 ≤ X < -2,59	Easy	10
5.	X < M - 1SD	X < -5,18	Very easy	1
	Total			26

Student Ability Level

The student's ability level aims to determine the student's ability to work on the items given. Similar to the analysis of the difficulty level of the item, in the Rasch model, the person's ability value is obtained from the logit value (measure), the higher the logit value, the probability level of students in answering the item correctly is very high and vice versa. The following is a summary of the person capabilities obtained from the Winstep program table 17.1:

Figure 3. Summary of Student Ability Level

TABLE 17.1 DATA PAKET-2.xlsx ZOU039WS.TXT May 4 2022 7:51
INPUT: 407 PERSON 47 ITEM REPORTED: 233 PERSON 26 ITEM 2 CATS WINSTEPS 4.4.7

PERSON: REAL SEP.: 1.75 REL.: .75 ... ITEM: REAL SEP.: 3.93 REL.: .94
PERSON STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	PERSON
42	24	26	4.70	1.85	MAXIMUM MEASURE				.77	.77	100.0	100.0	42 L IV 11
71	23	26	3.44	1.04	1.08	.38	1.12	.55	.64	.67	95.7	95.6	71 L IV 12
104	23	26	3.44	1.04	1.07	.37	1.05	.51	.65	.67	95.7	95.6	104 P IV 10
232	23	26	3.44	1.04	1.04	.34	.76	.29	.66	.67	95.7	95.6	232 L V 11
46	22	26	2.67	.76	1.04	.27	.76	.04	.60	.60	91.3	91.3	46 L IV 11
48	22	26	2.67	.76	1.00	.20	.66	-.09	.62	.60	91.3	91.3	48 P IV 11
103	22	26	2.67	.76	.86	-.07	.64	-.13	.64	.60	91.3	91.3	103 P IV 10
106	22	26	2.67	.76	1.17	.47	1.37	.67	.55	.60	91.3	91.3	106 P IV 10
105	21	26	2.18	.64	1.21	.60	1.22	.52	.50	.56	87.0	86.9	105 P IV 10
203	21	26	2.18	.64	.82	-.29	.58	-.44	.62	.56	87.0	86.9	203 P IV 11
217	5	26	-1.82	.59	1.06	.30	.79	-.21	.43	.43	78.3	82.9	217 L V 12
237	5	26	-1.82	.59	.94	-.06	.69	-.43	.46	.43	87.0	82.9	237 L V 12
239	5	26	-1.82	.59	.81	-.48	.58	-.70	.49	.43	87.0	82.9	239 P V 12
245	5	26	-1.82	.59	1.23	.74	1.21	.54	.36	.43	78.3	82.9	245 P V 12
300	5	26	-1.82	.59	.89	-.22	.78	-.23	.46	.43	87.0	82.9	300 L V 12
380	5	26	-1.82	.59	1.11	.41	1.12	.39	.39	.43	78.3	82.9	380 L IV 10
393	5	26	-1.82	.59	.71	-.83	.46	-1.02	.52	.43	87.0	82.9	393 L IV 11
391	4	26	-2.21	.65	.69	-.68	.37	-.96	.53	.43	87.0	86.9	391 L IV 10
399	4	26	-2.21	.65	1.13	.45	1.41	.75	.39	.43	87.0	86.9	399 L IV 10
120	3	26	-2.70	.77	1.04	.25	.78	.06	.45	.45	91.3	91.3	120 L IV 11
MEAN	11.9	26.0	-.10	.51	1.00	.0	.98	.0			72.5	72.4	
P.SD	4.3	.0	1.08	.12	.16	.8	.22	.8			10.2	6.3	

Figure 3 explains that there is 1 student who has a very high level of ability (maximum measure), namely student number 42. The logit value of the student is 4.70. These students can be accepted by the Rasch model but cannot be considered as students who have good abilities.

To determine the level of the student's ability level, the logitability values of the students can be entered into the Rasch model scaling. The following is a summary of the Rasch model scaling based on student ability level data:

Table 6. Summary of Student's Level of Ability Scale

No	Scaling Model	Student Logit Scale	Ability Level	Total
1.	X > M + 3SD	X > 3,24	Special	4
2.	M + 3SD ≤ X < M + 2SD	3,24 ≤ X < 2,16	Very high	6
3.	M + 2SD ≤ X < M + 1SD	2,16 ≤ X < 1,08	High	12
4.	M + 1SD ≤ X < 0	1,08 ≤ X < 0	Medium	77
5.	0 ≤ X < M - 1SD	0 ≤ X < -1,08	Low	102
6.	M - 1SD ≤ X < M - 2SD	-1,08 ≤ X < -2,16	Very Low	29
8.	X < M - 3SD	X < -2,16	Minim	3
	Total			233

Table 6 explains that the ability (ability) of students in working on the items worked on by the majority is in the medium and low ability levels.

Discussion

Mathematical literacy plays a very important role in improving students' abilities in applying mathematical facts, concepts, and procedures in everyday life, so that they can help solve problems that occur in society. However, based on data from international studies such as PISA and TIMSS, the mathematical literacy ability of Indonesian students is still low. Many factors cause low mathematical literacy skills, including the low ability of teachers to design good test questions, especially questions containing high-order thinking skills (HOTs). This is reinforced by several studies, such as the research of Rapih (2018) and Mutakin & Hakim (2019). Based on this research, the majority of teachers have not been able and are not accustomed to making HOTs questions, especially in the field of mathematics.

This research is a type of research that develops HOTs questions with reference to the TIMSS 2019 framework. The 2019 TIMSS framework was taken because after 2015, the Indonesian people no longer participated in studies conducted by TIMSS. In addition, the levels taken in the development of questions are carried out at level four (elementary school), as the foundation for the next level.

The questions developed refer to the abilities tested in the 2019 TIMSS framework for the fourth level of mathematics. The initial development of the questions consisted of 80 items

which were divided into 2 question packages. Each question package consists of 47 items with details of 42 multiple choice items and 5 essay questions. In this study, the writer examines the results of students' answers to package 2 questions with a grid of questions that have been explained in the research method.

Research analysis on the results of students' answers to questions was carried out using the Rasch model which was processed through the Winstep program version 4.4.7. The results of the analysis explain; First, of the 47 questions answered by the students, 26 items (55.32%) were accepted by the Rasch model (fit) and 23 items (48.94%) were questions that had a good category. Based on the results of the scaling, the fit items are in the medium and easy categories. This result means that public elementary school students in DKI Jakarta who are the sample in the study, are only able to answer questions that have moderate and easy levels of difficulty. This is in line with the average level of student ability based on the results of the scaling, which are at medium and low levels. The summary of the abilities tested from the items developed and included in the good category are as follows:

Table 7. Summary of Tested Abilities from Developed Test Items

Domain Conten	Domain Cognitif			Total
	Knowing	*Applying*	*Reasoning*	
Number	4	8	2	14
Measurement & Geometry	3	3	0	6
Data	0	1	2	3
Total	7	12	4	23

Table 7 above provides information that the abilities tested from the items developed are normally distributed. This means that the majority of the questions that the majority of students can answer are at a moderate level. These results are consistent with the data described by the Wright Map.

Conclusion

This study describes the use of the Rasch model in analyzing the developed mathematical literacy items. The results of the analysis explained that of the 47 test items carried out by students, there were 23 items that fit the model and had a moderate level of item difficulty. In addition, it was also explained that the level of ability of students in working on the items was included in the medium and low level categories. This means that the fourth and fifth-level public elementary school students had good abilities in answering the items medium and low levels. While, test items high level could not be well-answered. To improve students' ability in answering high-order thinking skills (HOTs) test items very well, student shoud do regular exercises so that they get used to these HOTs test items.

Acknowledgements

The author expresses his gratitude to Prof. Dr. Burhanuddin Tola and Bahrul Hayat, Ph.D as promoters and co-promoters in this research who have guided and provided input in the stages of completing this research.

REFERENCES

Dinni, Husna Nur, 2018, HOTS (High Order Thinking Skills) dan Kaitannya dengan Kemampuan Literasi Matematika, Prisma 1, Prosiding Seminar Nasional Matematika.

Hadi, Syamsul & Novaliyosi, 2019, TIMSS Indonesia (*Trends In International Mathematics And Science Study*), *Prosiding Seminar Nasional & Call For Papers* Program Studi Magister Pendidikan Matematika Universitas Siliwangi Tasikmalaya, Isbn: 978-602-9250-39-8

IEA, *TIMSS 2015, International Resultsin Mathematics*.

Kemdikbud, Dirjen Dikdasmen, 2017, *Panduan Penilaian oleh Pendidik dan Satuan Pendidikan*, Sekolah Menengah Pertama, Hal. 8

Mutakin, Tatan Zenal & Hakim, Arif Rahman, 2019, Teachers' Ability In Designing Test Assessments, *Advances In Social Science, Education And Humanities Research, Volume 512*, Proceedings Of The 1st International Conference On Folklore, Language, Education And Exhibition (ICOFLEX 2019).

OECD, 2017, *PISA for Development Assessment and Analytical Framework, Reading, Mathematics and Science, Preliminary Version*, p. 17

OECD, 2019, *PISA, Insights and Interpretations*.

Rahmah, Nur, 2013, Hakikat Pendidikan Matematika, *al-Khwarizmi*, Volume 2, Oktober 2013, halaman 1 – 10, https://ejournal.iainpalopo.ac.id/index.php/al-khwarizmi/article/download/88/75.

Rapih, Subroto & Sutaryadi, 2018, Perpektif Guru Sekolah Dasar Terhadap Higher Order Tinking Skills (HOTS): Pemahaman, Penerapan Dan Hambatan, Premiere Educandum: *Jurnal Pendidikan Dasar dan Pembelajaran*, Volume 8(1) 78 – 87 Juni 2018, ISSN: 2088-5350 (Print) / ISSN: 2528-5173 (Online), Http://E-Journal.Unipma.Ac.Id/Index.Php/PE/Article/Download/2560/Pdf

Rizky, L.M & Priatna, N, 2018, Mathematical literacy as the 21st century skill, *International Conference on Mathematics and Science Education* (ICMScE 2018), IOP Conf. Series: Journal of Physics: Conf. Series 1157 (2019) 042088

Setiawati, Wiwik at all, Modul Belajar Mandiri, Calon Guru, Pegawai Pemerintah dengan Perjanjian Kerja (PPPK), file:///F:/DATA%20PER%20PAKET/paket-2/wiwik.pdf.

Sumintono, Bambang & Widhiarso, Wahyu (2015), *Aplikasi Pemodelan Rasch Pada Asesment Pendidikan*, Bandung: Trim Komunikasi, Cet I.

Sumintono, Bambang & Widhiarso, Wahyu (2015), *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*, Bandung: Trim Komunikasi, Cet II.

Sutrina, Nana, 2021, Analisis Kemampuan Literasi Sains Peserta Didik SMA di Kota Sungai Penuh, *Jurnal Inovasi Penelitian*, Vol.1 No.12, ISSN 2722-9475 (Cetak), ISSN 2722-9467 (Online), f i l e : / / / F : / D A T A % 2 0 P E R % 2 0 P A K E T /paket-2/data%20PISA.pdf.