Check for updates

How does Rasch modeling reveal difficulty and suitability level the fraction test question?

Karlimah

Elementary School Education Study Program, Universitas Pendidikan Indonesia, West Java, Indonesia

[*] Correspondence: karlimah@upi.edu
© The Author 2022

Abstract

This article explains how to analyze test items in arithmetic operation with fractions to obtain the items' level of difficulty and fitness. Data were collected by using multiple-choice questions given to 50 fourth-grade students of an elementary school in Tasikmalaya city. The answers were then analyzed using the Rasch model and Winsteps 3.75 application, a combination of standard deviation (SD) and logit mean values (Mean). The score data of each person and question were used to estimate the pure score in the logit scale, indicating the level of difficulty of the test items. The categories were difficult (logit value $> +1$ SD); very difficult (0.0 logit $+1$ SD); easy (0.0 logit -1 SD); very easy (logit value $< -$SD). Three criteria were used to determine the level of difficulty and fitness of the questions: the Outfit Z-Standard/ZSTD value; Outfit Mean Square/MNSQ; and Point Measure Correlation. It resulted in a collection of test items suitable for use with several levels of difficulties, namely, difficult, very difficult, easy, and very easy, from the previous items, which had difficult, medium, and easy categories. Rasch model can help categorize questions and students' ability levels.

Keywords: difficulty analysis; fractions; item fitness; Rasch modeling; test questions

Introduction

Every learning process has a goal related to some knowledge or skills that students must obtain. In achieving the goal, measurement is used to determine the value, score, or percentage achieved by students related to the learning objectives. In the process or end of learning, it is necessary to measure the learning process and its results in the form of numbers that reflect the achievements of the learning process and results. According to Mardapi, measurement is basically an activity of systematically determining numbers on an object (Doran, 1980). Measurement is a process or activity to determine the quantity of something. The word "something" could mean students, teachers, school buildings, study tables, whiteboards, and others (Parker, 2000). Measurement is a process that describes the student's performance with a quantitative scale until qualitative characteristic is shown by numbers (Alwasilah, 1996). Thus, measurement in education means measuring student attributes or characteristics (Goddard et al., 2000).

Measurement in education is closely related to tests. One way that is often used to measure the results that students have achieved is by testing (Harris & Brown, 2010). In the measurement process, the teacher must use either test or non-test measurement tools (Parker, 2000). According to Zainul (2001), a test is a question or task or a set of tasks used to obtain information about an educational attribute. The test used is adjusted to the subject or field of science used as a test source (Zainul, 2001).

The results obtained from the test are data. This article analyzes the data using the Winsteps 3.75 application through the Rasch rating scale model. The determination of the assessment of the Rasch model was chosen because it has a good measurement approach and can model the relationship between item difficulty, problem-taking ability, and the probability of a given response (Andrich, 1981). In addition, Rasch modeling uses psychometric analysis techniques to be used to develop test items. Rasch modeling is also an important tool that can provide relevant information regarding student assessment after learning (Sumintono, 2018).

The analysis of instrument tests using the Rasch model includes response item measurement theory. This measurement describes of interaction between subject and test items. It will make the measurement result more appropriate and objective (Sumintono & Widhiarso, 2014). Meanwhile, according to Brodge (1977), the Rasch model is usually used for item measurement and the subject of people; in this context, examining the relationship between the comparison law and other additional combined measurements is discussed in this fraction test.

According to Masters, Rasch modeling can be used for various observation formats, including models for calculation analysis, repeated experiments, and rating scales (Masters, 1982). In addition, the statistical description of Rasch fitness can give a valuable framework for testing the correctness of a person's response, measuring the estimation of a person's response-ability, and being able to detect various disturbances to a person's response (Smith, 1986). The Rasch model is noted as a probability model from the individual response towards an item and, therefore, not explicitly become a response model for itself (Brodgen, 1977).

In the 1960s, Georg Rasch has developed an analytical model of Item Response Theory (IRT) and later was being popular by Ben Wright. The raw data is collected as dichotomy data

(true or false) that indicates students' ability. The Rasch model makes it a model that relates between students and items (Sumintono & Widhiarso, 2014). Besides dichotomy data, the Rasch model is used to analyze polytomy data that was developed by Andrich, based on two fundamental theorems, the ability of a person and the difficulty level of an item. The Rasch model assumed that the difficulty item is a disposition that is influenced by the response of respondence, and personability is a characteristic that influences by the difficulty level estimation (Linacre, 1999; Nur et al., 2020). The advantages of the Rasch model if compared with a classic theory that Rasch model can identify the false answer, identify the assessment that is not appropriate, and predict the loss data based on systematic response patterns (Fahmina et al., 2019; Goodwin & Leech, 2017; Ratna et al., 2017).

The topic of this research refers to previous research, where Rasch modeling is used for measuring critical thinking of student's skills in STEM learning in elementary schools (Hamdu et al., 2020). We know that STEM provides a learning experience of mathematics and science, technology, engineering that is suitable for students in elementary school. When students are given the task of making media or props for steam power in water ships, indirect measurement skills are needed. It requires math skills. However, there is an assumption that students have difficulty counting fractions.

Therefore, in this study, Rasch modeling is used to evaluate learning on rational numbers and describe how Rasch modeling can reveal the level of difficulty and fitness of test questions on fractions. Evaluation is an essential part of the learning process (Jackson et al., 2002). Generally, learning evaluation aims to widely determine the effectiveness, efficiency, and learning of the learning system, including knowing the quality of the questions used to evaluate. The quality of the questions are questions that can measure and describe students' thinking abilities, including questions that are appropriate to measure the level of cognition and difficulty.

Methods

To obtain information about the level of difficulty and the suitability of fractional number questions for elementary school students, it is necessary to conduct research using descriptive analysis, with a cross-sectional design, and data collection at one time. The research subjects were fourth-grade students from five elementary schools in Tasikmalaya City. Data were collected from 25 male students and 25 female students. The data studied were collected from the answers to 30 valid multiple-choice questions based on basic competencies in the fractional number material. The following are the basic mathematics competencies in fractions for fourth-grade elementary school students.

Table 1. Basic mathematics competencies in fractions in elementary school

No	Basic Competencies
1	Explain equivalent fractions with pictures and concrete models.
2	Describes the various forms of fractions (common, mixed, decimal, and percent) and the relationships between them.

No	Basic Competencies
3	Explain and perform estimations of the number, difference, product, and quotient of two whole numbers as well as fractions and decimals.
4	Identify equivalent fractions with pictures and concrete models.
5	Look for different forms of fractions (ordinary, mixed, decimal, and percent) and the relationships between them.
6	Solve problems estimating the number, difference, product, and quotient of two whole numbers as well as fractions and decimals.

Furthermore, the measurement of the quality of the test questions, that is, regarding the level of difficulty and appropriateness, was carried out. The data collection technique is shown in Figure 1. below, which was modified from Hamdu et al. (2020).
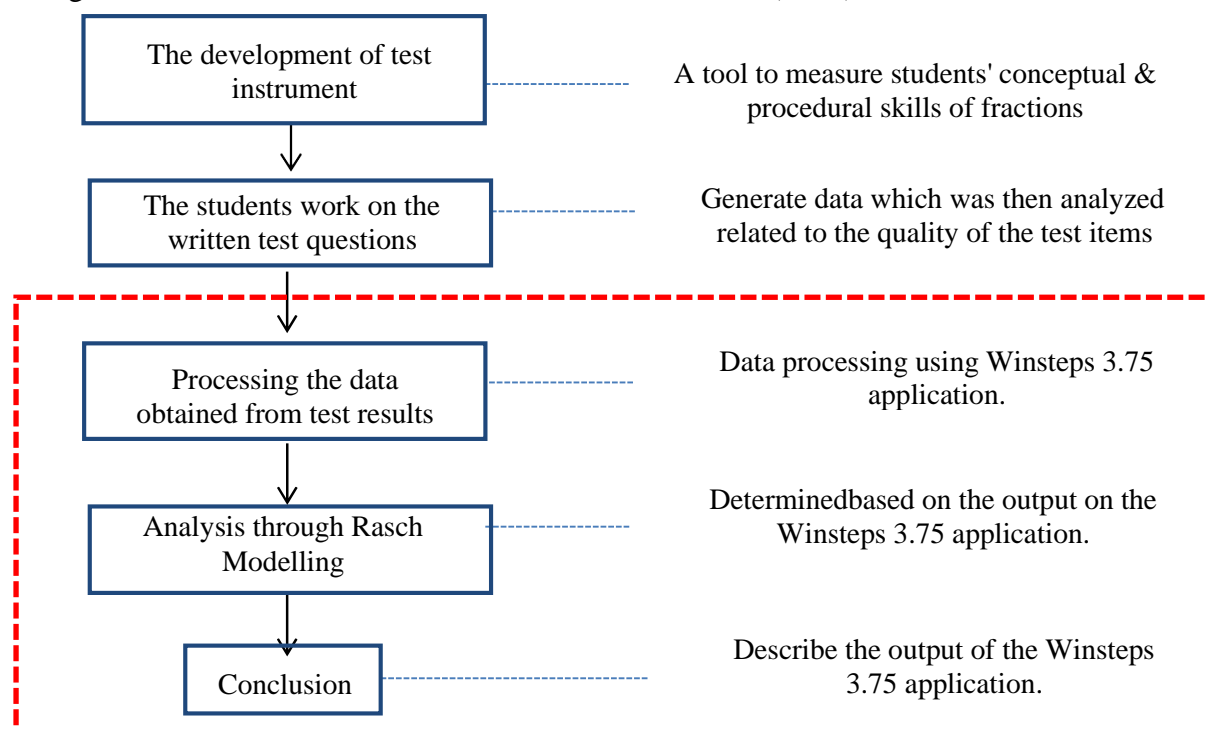

Figure 1. Data analysis stages

The stages of analysis in Figure 1. begin with preparing research instruments in the form of written test questions as a tool to obtain research data. Furthermore, the data is taken from the results of filling in the questions by students. This study focuses on the stages of data analysis in Figure 1. which are marked with a red box. The disclosure of the level of difficulty and suitability of the items was carried out using Rasch modeling assisted by the Winsteps 3.75 application. Rasch modeling uses the following categorization standards:

1. For the category of a very hard question, the logit value is greater than +1SD.
2. Hard question the value of 0.0 logit +1 SD.
3. Easy question the value is 0.0 logit -1 SD.
4. Very easy question the value is less than –SD.

Results

The following are the results of the analysis of test questions through Rasch modelling assisted by the Winsteps 3.75 application.

Analysis of Item Measure Difficulty Levels

The red squares in Table 2. provide information on the level of difficulty of each item of fractional number questions as follows.

Table 2. Item Statistics: Measure Order

```
|ENTRY   TOTAL   TOTAL            MODEL|   INFIT   |   OUTFIT   |PT-MEASURE |EXACT MATCH|         |
|NUMBER  SCORE   COUNT   MEASURE   S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP. | OBS%   EXP%| Item    |
|    29     13      50     2.48     .43|1.11   .5| .95    .1|  .60   .63| 79.2   84.1| S29     |
|    30     15      50     2.13     .40| .97  -.1|1.81   1.6|  .60   .63| 85.4   82.5| S30     |
|     5     17      50     1.82     .39|1.05   .3|1.20    .6|  .59   .62| 81.3   80.9| S5      |
|    27     18      50     1.67     .38|1.01   .1|1.06    .3|  .61   .62| 79.2   80.0| S27     |
|     9     21      50     1.25     .36| .91  -.4|1.12    .5|  .62   .60| 85.4   77.8| S9      |
|     4     22      50     1.12     .36|1.11   .6|1.03    .2|  .56   .60| 75.0   77.1| S4      |
|    22     22      50     1.12     .36|1.23  1.3|1.40   1.3|  .49   .60| 75.0   77.1| S22     |
|    16     23      50      .99     .36| .76 -1.5| .63  -1.3|  .69   .59| 85.4   76.3| S16     |
|    13     26      50      .62     .35| .82 -1.2| .80   -.6|  .64   .57| 81.3   74.1| S13     |
|    23     26      50      .62     .35| .96  -.2|1.50   1.5|  .55   .57| 81.3   74.1| S23     |
|    18     27      50      .50     .35| .88  -.7| .72   -.8|  .62   .56| 75.0   73.7| S18     |
|    19     27      50      .50     .35|1.17  1.2|1.24    .8|  .48   .56| 70.8   73.7| S19     |
|    20     27      50      .50     .35|1.16  1.1|1.45   1.3|  .47   .56| 70.8   73.7| S20     |
|    25     27      50      .50     .35|1.18  1.2|1.36   1.1|  .47   .56| 70.8   73.7| S25     |
|    15     28      50      .38     .35|1.37  2.3|1.58   1.6|  .38   .55| 62.5   73.5| S15     |
|    21     30      50      .14     .35| .97  -.2| .92   -.1|  .54   .53| 75.0   73.5| S21     |
|     7     32      50     -.11     .35| .94  -.3|1.28    .8|  .52   .51| 83.3   73.7| S7      |
|    17     33      50     -.23     .35| .80 -1.4| .63   -.8|  .59   .50| 75.0   74.4| S17     |
|    24     33      50     -.23     .35| .86  -.9| .68   -.7|  .57   .50| 79.2   74.4| S24     |
|    26     34      50     -.36     .36|1.24  1.5|1.23    .6|  .38   .49| 70.8   75.1| S26     |
|    28     35      50     -.49     .36|1.00   .0|1.27    .7|  .46   .48| 77.1   75.8| S28     |
|    11     37      50     -.76     .37| .82 -1.0| .67   -.5|  .53   .45| 83.3   77.8| S11     |
|     6     38      50     -.90     .38| .78 -1.3| .52   -.7|  .55   .44| 81.3   79.0| S6      |
|     2     41      50    -1.37     .42| .86  -.6| .85    .1|  .44   .39| 89.6   83.4| S2      |
|    10     41      50    -1.37     .42| .86  -.6| .78    .0|  .45   .39| 89.6   83.4| S10     |
|    12     41      50    -1.37     .42| .78 -1.0| .53   -.5|  .50   .39| 85.4   83.4| S12     |
|     3     42      50    -1.55     .43| .97   .0| .67   -.1|  .40   .37| 85.4   84.8| S3      |
|    14     42      50    -1.55     .43|1.13   .6|1.08    .4|  .31   .37| 85.4   84.8| S14     |
|     8     44      50    -1.96     .48| .90  -.2| .53   -.3|  .39   .33| 89.6   88.0| S8      |
|     1     49      50    -4.08    1.04| .95   .3| .22   -.4|  .21   .15| 97.9   97.9| S1      |
| MEAN     30.4    50.0      .00     .40| .99   .0| .99    .2|            | 80.2   78.7|         |
| S.D.      9.2     .0     1.37     .12| .16   .9| .37    .8|            |  7.3    5.6|         |
```

Table 2. above has provided information that all of the questions can be grouped as follows.
1. The group of items is very difficult, namely question no. 29, 30, 5 and 27
2. The group of difficult items, namely question no.9, 4, 22, 16, 13, 23, 18, 19, 20, 25, 15, and 21
3. The group of easy items, namely question no. 7, 17, 24, 26, 28, 11, and 6
4. The group of items is very easy, namely question no. 2, 10, 12, 3, 14, 8, and 1

     Question no. 29 and 30 are very difficult questions and question no. 1 is a very easy question. The question is as follows:

Table 3. Samples of Test Questions Based on Difficulty Levels

No	Difficulty Level	Test Question Form	Indicator
1	Very Hard	*(Question no. 30)* Bagas cut the watermelon into 8 parts. Bagas then ate two parts of the watermelon. 2 friends Bagas did likewise. So, the value of the watermelon that Bagas and his friends ate when written in fraction form is... a. $\frac{1}{2}$ c. $\frac{3}{4}$ b. $\frac{2}{3}$ d. $\frac{3}{2}$	Shows equivalent fractions with pictures and concrete models.

No	Difficulty Level	Test Question Form	Indicator
2	Hard	***(Question no. 22)*** There are 60 students in the fourth grade of Mekar elementary school. On this day there will be 4 students. Each of them will bring 1/60 of 50kg of rice to make *nasi liwet* (*liwet* rice) and eat it together. Therefore, the fourth grade students of SD Mekar, will cook *liwet* rice as much as ….. part of 50 kg of rice. a. $\frac{1}{15}$ c. $\frac{1}{30}$ b. $\frac{2}{15}$ d. $\frac{3}{30}$	Determine the sum, difference, product, and quotient of fractions.
3	Easy	***(Question no. 6)*** A fraction equivalent to $\frac{3}{5}$… a. $\frac{3}{10}$ c. $\frac{3}{100}$ b. $\frac{6}{10}$ d. $\frac{6}{100}$	Explain equivalent fractions.
4	Very Easy	***(Question no. 1)*** The fraction $\frac{1}{8}$ is pronounced as ..... a. One eight c. Eight one b. One eighths d. Eight by one	Pronounce the value of fractions.

The indicators for the form of test questions for each level of difficulty are based on table 3 above, namely; 1) Shows equivalent fractions with pictures and concrete models (very high); 2) Determine the sum, difference, product, and quotient of fractional numbers (high); 3) Explain equivalent fractions (medium); and 4) Pronounce the value of the fraction (low).
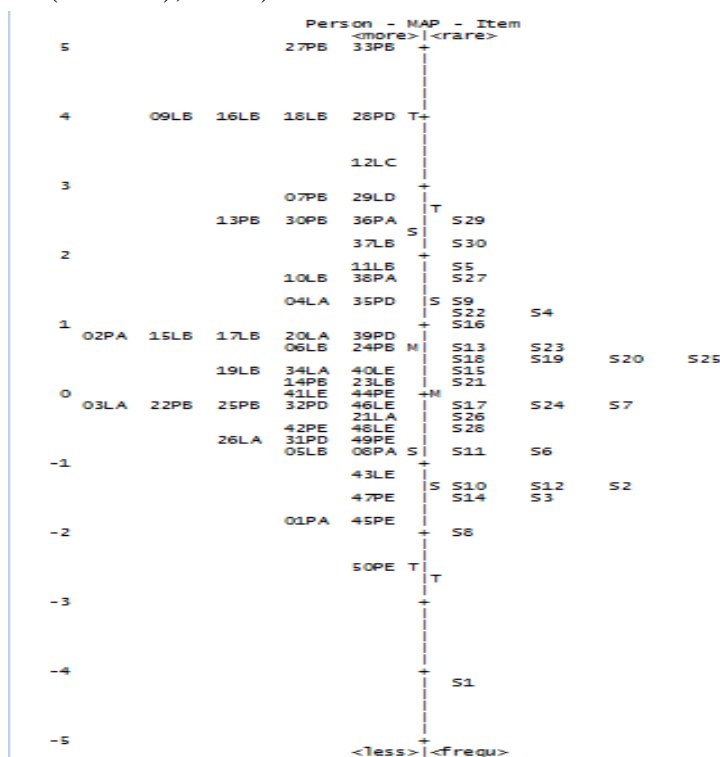

Figure 2. Wright map

Based on Wright's map, information was obtained that the very easy category questions (S1) are still considered very easy by students with a low ability (50PE). Nine students can easily answer very difficult questions with high abilities (07PB, 29LD, 12LC, 09LB, 16LB, 18LB, 28PD, 27PB, 33PB).

Analysis of Item Fit Order

The level of fitness of fractional numbers test questions (Item Fit) can be determined by using three criteria, namely the outfit mean-square value (Outfit MNSQ), Outfit Z-Standardized (Outfit ZSTD), and Point Measure Correlation (PT-Measure Corr) (Bond & Fox, 2015). The criteria used to check for non-conforming items (outlier or misfit) are:

1. The value of the Outfit square-mean (Outfit MNSQ) received: 0.5 < MNSQ < 1.5.
2. The value of Outfit Z-Standardized (Outfit ZSTD) received: -2.0 < ZSTD < +2.0.
3. Point Measure Correlation Value (PT-Measure Corr): 0.4 < PT-Measure Corr < 0.85

Table 4. Item statistics: Misfit order

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 OBS%	MATCH EXP%	Item
30	15	50	2.13	.40	.97	-.1	1.81	1.6	A .60	.63	85.4	82.5	S30
15	28	50	.38	.35	1.37	2.3	1.58	1.6	B .38	.55	62.5	73.5	S15
23	26	50	.62	.35	.96	-.2	1.50	1.5	C .55	.57	81.3	74.1	S23
20	27	50	.50	.35	1.16	1.1	1.45	1.3	D .47	.56	70.8	73.7	S20
22	22	50	1.12	.36	1.23	1.3	1.40	1.3	E .49	.60	75.0	77.1	S22
25	27	50	.50	.35	1.18	1.2	1.36	1.1	F .47	.56	70.8	73.7	S25
7	32	50	-.11	.35	.94	-.3	1.28	.8	G .52	.51	83.3	73.7	S7
28	35	50	-.49	.36	1.00	.0	1.27	.7	H .46	.48	77.1	75.8	S28
26	34	50	-.36	.36	1.24	1.5	1.23	.6	I .38	.49	70.8	75.1	S26
19	27	50	.50	.35	1.17	1.2	1.24	.8	J .48	.56	70.8	73.7	S19
5	17	50	1.82	.39	1.05	.3	1.20	.6	K .59	.62	81.3	80.9	S5
14	42	50	-1.55	.43	1.13	.6	1.08	.4	L .31	.37	85.4	84.8	S14
9	21	50	1.25	.36	.91	-.4	1.12	.5	M .62	.60	85.4	77.8	S9
29	13	50	2.48	.43	1.11	.5	.95	.1	N .60	.63	79.2	84.1	S29
4	22	50	1.12	.36	1.11	.6	1.03	.2	O .56	.60	75.0	77.1	S4
27	18	50	1.67	.38	1.01	.1	1.06	.3	o .61	.62	79.2	80.0	S27
21	30	50	.14	.35	.97	-.2	.92	-.1	n .54	.53	75.0	73.5	S21
3	42	50	-1.55	.43	.97	.0	.67	-.1	m .40	.37	85.4	84.8	S3
1	49	50	-4.08	1.04	.95	.3	.22	-.4	l .21	.15	97.9	97.9	S1
8	44	50	-1.96	.48	.90	-.2	.53	-.3	k .39	.33	89.6	88.0	S8
18	27	50	.50	.35	.88	-.7	.72	-.8	j .62	.56	75.0	73.7	S18
24	33	50	-.23	.35	.86	-.9	.68	-.7	i .57	.50	79.2	74.4	S24
2	41	50	-1.37	.42	.86	-.6	.85	.1	h .44	.39	89.6	83.4	S2
10	41	50	-1.37	.42	.86	-.6	.78	.0	g .45	.39	89.6	83.4	S10
11	37	50	-.76	.37	.82	-1.0	.67	-.5	f .53	.45	83.3	77.8	S11
13	26	50	.62	.35	.82	-1.2	.80	-.6	e .64	.57	81.3	74.1	S13
17	33	50	-.23	.35	.80	-1.4	.63	-.8	d .59	.50	75.0	74.4	S17
12	41	50	-1.37	.42	.78	-1.0	.53	-.5	c .50	.39	85.4	83.4	S12
6	38	50	-.90	.38	.78	-1.3	.52	-.7	b .55	.44	81.3	79.0	S6
16	23	50	.99	.36	.76	-1.5	.63	-1.3	a .69	.59	85.4	76.3	S16
MEAN	30.4	50.0	.00	.40	.99	.0	.99	.2			80.2	78.7	
S.D.	9.2	.0	1.37	.12	.16	.9	.37	.8			7.3	5.6	

The red square in Table 4. above shows that items 30, 15, and 1 do not meet the MNSQ value; all items meet the ZSTD value, and items 15, 26, 14, 1, and 8 do not meet the PT-Measure Corr. However, some items do not meet these three criteria. All questions about fractional fractions analyzed have an acceptable level of conformity and deserve to be maintained. As an additional explanation, the ZSTD value of all questions within the allowed limit is the cause of all items worth defending.

Discussion

Calibrated assessment instruments can verify student response patterns appropriately. In this case, Rasch modeling in principle precisely calibrates the measurement scale, respondents, and items. Rasch modeling works by processing score data based on per person and score data per item (question). The two scores become the basis for estimating the actual score on a logit scale which can indicate the level of difficulty of the item. Raw scores processed through Rasch modeling produce true/logit scores that have been measured using equal or equivalent intervals.

Assessment instruments processed through Rasch modeling can discuss assessment instruments more specifically. Efforts to know the competence and incompetence of students in a material can be obtained by focusing on discussing the material test instrument. In this article, the author analyzes the fractional number test instrument based on the basic competencies in the fractional number material curriculum used in learning to fourth-grade elementary school students.

The presentation of Rasch's modeling analysis results provides information that students' competence in explaining equivalent fractions with pictures and concrete models is considered very difficult. In skill competence, students find it difficult to solve the problem of estimating the number, difference, product, and quotient of two whole numbers and fractions and decimals. However, on the competence of skills to easily identify equivalent fractions with pictures and concrete models. Moreover, the knowledge competence explains various forms of fractions (ordinary, mixed, decimal, and percent), and students consider the relationship between them very easy.

Of all the test questions analyzed by Rasch modeling, it shows that based on the level of difficulty, the items are divided into four categories: very difficult, difficult, easy, and very easy. The items' difficulty level categorization was done by combining the standard deviation (SD) value, and the logit means value (Sumintono & Widhiarso, 2015). Based on the explanation and analysis of the difficulty level of measuring the items in the results section above. Problems can be used to measure students' mathematical knowledge and skills.

There are three criteria (MNSQ, ZSTD, and PT-Measure Corr) that can ensure that the question is not good enough so that it needs to be repaired or replaced (Biggs & Collis, 2014; Jackson et al., 2002). There were no items that did not meet the three criteria from these findings. Therefore, all the fractional number items analyzed have an acceptable level of conformity and deserve to be maintained. Besides that, information obtained from the processing results is shown in Table 2. Measure Order that questions that were previously made in the difficult, medium, and easy categories became very difficult, difficult, easy, and very easy categories. Then if you look at Figure 2, Wright's map, it was found that nine students (07PB, 29LD, 12LC, 09LB, 16LB, 18LB, 28PD, 27PB, 33PB) responded that the very difficult questions were still below their abilities, meaning that the very difficult questions were not considered very difficult for them difficult. Likewise, for questions categorized as very easy (no. 1), a student with the lowest ability (50PE) responded very easily, meaning that the question was very easy for students with the lowest ability, which was considered to be much easier. However, the questions proportional to the distribution of questions in the difficult,

medium, and easy categories, found on the Wright map are still proportional. The questions are more widely distributed in the difficult and easy categories, and it is shown that there are two very difficult questions (no. 29 and no. 30) and one very easy question (no. 1).

The results of the Rasch modeling analysis show that the level of difficulty of the fractional number test questions is appropriate for each category, namely very difficult, difficult, easy, and very easy. Various levels of item difficulty are needed to identify students' various abilities. In very difficult category questions, the percentage of students with correct answers tends to be the lowest compared to other question categories. However, in the categories of difficult, easy, and very easy questions, the percentage of students who answered correctly tended to vary with several possibilities, such as students answering questions by guessing or cheating. These possibilities can occur in some questions that are classified as difficult, easy, and very easy because of the pattern of presentation of questions or different levels of student ability.

Conclusion

Rasch modeling assisted by the Winsteps 3.75 application makes it easy for users to analyze the assessment instrument (test questions) where the raw scores are processed and provides explicit analysis on the design of the items used and whether the score pattern used is appropriate. Based on the results, it can be concluded that several test questions showed various levels of difficulty with an acceptable level of fitness. The difficulty level related to the fractional numbers questions, namely, 1) the majority of hard items were found in the indicator of the ability to show equivalent fractions with pictures and concrete models; 2) the majority of very hard items were found in the indicator of the ability to determine fractions based on pictures, determine the product and quotient of the fractional numbers and explain the form of mixed fractions; 3) the majority of easy questions were found in the indicator of the ability to determine the sum and difference of fractional numbers; 4) the majority of very easy questions were found in the indicator of the ability to pronounce equivalent fractions.

The analysis results through Rasch modeling specifically provide a comprehensive description of the quality and categories of questions. It can give an idea of students' ability to the material that has been studied. Furthermore, it can provide an overview of how mathematics learning is carried out on specific materials and at certain times. Therefore, the results of the analysis of Rasch modeling can provide a description of the conditions and learning situations, such as those relating to student characteristics and the implementation of mathematics learning in certain classes or schools. Thus, Rasch modeling analysis can identify the learning process and create comprehensive tests.

Conflicts of Interest

The author declares that no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely by the author.

References

Alwasilah, e. a. (1996). *Glossary of educational assesment term*. Ministry of Education and Culture.

Andrich, D. (1981). Book review: Probabilistic models for some intelligence and attainment test (expanded edition: Georg Rasch Chicago: The University of Chicago Press, 1980, 199 pp., 15 hardcover, 7 paperback). *Applied Psychological Measurement*, *5*(4), 545-550. https://doi.org/10.1177/014662168100500413

Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences, third edition*. Routledge. https://doi.org/10.4324/9781315814698

Brodgen, H. E. (1977). The Rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika, 42*(2), 631-634. https://doi.org/10.1007/BF02295985

Doran, R. L. (1980). *Basic measurement and evaluation of science instruction*. National Science Teachers Association, 1742 Connecticut ave., NW, Washington, D.C 20009.

Fahmina, S., Masykuri, M., Ramadhani, D., & Yamtinah, S. (2019). Content validity uses Rasch model on computerized testlet instrument to measure chemical literacy capabilities. *AIP Conference Proceeding, 2194*(1), 1-8, Article 020023. https://doi.org/10.1063/1.5139755

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: its meaning, measure, and impact on student achievement. *American education research journal, 37*(2), 479-507. https://doi.org/10.2307/1163531

Goodwin, L. D., & Leech, N. (2017). The meaning of validity in the new standards for educational and psychological testing. *Measurement and Evaluation in Counseling and Development,* 36(3), 181-191. https://doi.org/10.1080/07481756.2003.11909741

Hamdu, G., Fuadi, F., Yulianto, A., & Akhirani, Y. (2020). Items quality analysis using Rasch model to measure elementary school students' critical thinking skill on STEM learning. *JPI (Jurnal Pendidikan Indonesia), 9*(1), 61-74. https://doi.org/10.23887/jpi-undiksha.v9i1.20884

Harris, L. R., & Brown, G. T. (2010). Mixing interview and questionnaire methods: practical problems in aligning data. *Practical Assesment, Research, and Evaluation, 15*(1), 1-15. https://doi.org/10.7275/959j-ky83

Jackson, T., Draugalis, J., Slack, M., Zachry, W., & D'Agostino, J. (2002). Validation of authentic performance assessment: A process suited for Rasch modeling. *American Journal of Pharmaceutical Education, 66*(3), 233-242.

Linacre, J. (1999). Investigating rating scale category utility. *Journal of outcome measurement, 3*(2), 103-122.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174. https://doi.org/10.1007/BF02296272

Nur, L., Nurani, L. A., Suryana, D., & Ahmad, A. (2020). Rasch model application on character development instrument elementary school students. *International Journal of Learning,*

Teaching and Education Research, 19(3), 437-459. https://doi.org/10.26803/ijlter.19.3.24

Parker, C. (2000). Performance measurement. *Emerald Insight, 49*(2), 63-66. https://doi.org/10.1108/00438020010311197

Ratna, I., Yamtinah, S., Ashadi, Masykuri, M., & Shidiq, A. (2017). The implementation of testlet assessment instrument in solubility and solubility product material for measuring students' generic science skills. *Advances in Social Science, Education and Humanities Research (ASSEHR), 158*, 958-963. https://doi.org/10.2991/ictte-17.2017.110

Smith, R. M. (1986). Person fit in the Rasch model. *Educational and Psychological Measurement*, *46*(2)359-372. https://doi.org/10.1177/001316448604600210

Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017) Atlantis Press, 173*, 38-42. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Rasch model untuk penelitian ilmu-ilmu sosial edisi revisi* [*Rasch model app for social science research revised edition*]. Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan* [*Rasch modeling application in educational assessment*]. Trim Komunikata Publishing House.

Zainul, A. (2001). *Alternative assesment*. Dirjen Dikti.