Educational Administration: Theory and Practice
2024, 30(4), 767- 779
ISSN: 2148-2403
https://kuey.net/
Research Article

Cognitive Test Instruments to Measure Student Mathematics Ability in Elementary School: Rasch Analysis

Iva Sarifah[1], Eko Kuswati[2], Wayan Suniasih[3], Fahrurrozi Fahrurrozi[4], Ganjar Susilo[5]

[1,4]Jakarta State University, Jakarta, Indonesia
[2]Universitas Terbuka, Jakarta, Indonesia
[3]Ganesha Education University, Bali, Indonesia
[5]Balikpapan University, East Kalimantan, Indonesia

Citation: Iva Sarifah et al. (2024), Cognitive Test Instruments to Measure Student Mathematics Ability in Elementary School: Rasch Analysis, *Educational Administration: Theory And Practice*, *30*(2), 767-799,
Doi: 10.53555/kuey.v30i4.1557

ARTICLE INFO

ABSTRACT
The purpose of this study was to describe the quality of cognitive instruments to measure students' ability in mathematics at the elementary school level. Quantitative type research to validate the mathematical test instrument on fractional arithmetic operations material which has 22 items. The purposive sampling technique was obtained by respondents totaling 65 students from 3 elementary schools in Yogyakarta, Indonesia. The Rasch Model analysis reveals that item validity corresponds to the MNSQ outfit (0.59 to 1.57), ZSTD outfit (-1.04 to 1.65), and Pt Measure Corr. (0.50 to 0.70) are met. The reliability of the instrument was fulfilled based on the aspects of person reliability (0.83; 2.23), item reliability (0.82; 2.15), and CA (0.93). There are two questions with high difficulty Q7 (δ = 1.42) and Q12 (δ = 1.09), 17 questions in the medium category, two questions in the easy category, namely Q22 (δ = -1.20) and Q8 (δ = -1.72), and one question very easy category, namely Q12 (δ = -2.14). The DIF analysis showed that all items did not contain a gender bias (ρ = 0.057-1.00), so there was no difference in the level of difficulty between men and women working on cognitive test questions. Cognitive test instruments are prepared according to national competency standards for grade five elementary school material levels that can be used as good quality question banks.

Keywords: Cognitive Test; Test Instruments; Mathematics Ability; Elementary School; Rasch Analysis

Introduction

Learning mathematics is very important for students' cognitive abilities (Ding et al., 2021; Layne et al., 2021). The purpose of understanding in this cognitive area is processing, structuring, and using knowledge. Meanwhile, according to Anderson et al. (2020) cognitive is an ability that prioritizes brain-based skills and is needed to perform any task ranging from simple to complex. Cognitive skills are brain-based skills and are used to complete any task in stages (Shen et al., 2019; Stelzer et al., 2021; Trinidad, 2020). Cognitive skills refer to the mental processes involved in acquiring, retaining, and using information. Among them are perception, attention, memory, problem solving, reasoning, decision making, and mental processes needed to solve problems in everyday life. Cognitive skills are important because they diagnose students' learning difficulties, improve learning outcomes, measure student progress and development and evaluate brain function or performance. Overall, cognitive skills assessment provides valuable information in improving educational and career outcomes and managing learning difficulties.

With regard to academic qualifications and teacher competency standards, conducting assessments of learning outcomes is one of the core competencies that must be met (Kholifah et al., 2023; Klaar & Wank, 2022). A teacher must be able to carry out evaluations to find out whether the material provided can be understood by students (Calderon & Sood, 2020; Nurtanto, Sudira, Sofyan, Kholifah, et al., 2022). Valid and objective assessments can be obtained by using tests that contain questions or questions that can reflect the ability being measured (Charalambous et al., 2019; Klaar & Wank, 2022). A test used in assessment must be of good quality and be able to measure students' actual abilities (Barber et al., 2022; Lyons et al., 2021; Stolt et al., 2022).

In the fifth-grade mathematics subject, there are indicator competencies that must be achieved, namely solving problems related to fractional arithmetic operations material. Based on the results of the interviews and sample questions given by the mathematics teacher at several elementary schools in Yogyakarta, Indonesia, the questions used for the end of semester assessment were questions made by the teacher himself in accordance with the material that had been taught. Each teacher has its own criteria for making daily test questions, according to the indicators and learning objectives set out in the learning program plan (Wei & Cheng, 2022).
After conducting a reasoning ability test for fractional arithmetic operations in fifth grade elementary school students, the answer results were measured using the Assessment Reference Criteria (ARC) and it was known that around 75% of students were included in the category of very poor reasoning ability, only 25% of students were included in the ability category. less reasoning. This is because students are not used to solving questions that are designed to use reasoning abilities in answering them. Reasoning ability has a significantly positive effect on learning mathematics, the higher the reasoning ability the better the student in completing the math test (Ding et al., 2021; Layne et al., 2021; Stelzer et al., 2021).
Educational assessment standards state that the principles underlying the assessment of learning outcomes are valid, comprehensive, fair, integrated, open, objective, and sustainable, economical, accountable, and educative (Ajjawi et al., 2021; Alonzo, 2018; Cogan et al., 2019; Ormond, 2019). Until now, many learning outcomes instruments have not met the requirements as good tests (Alonzo, 2018; Cogan et al., 2019). One of the things that might be the cause is the teacher's ability to make tests that are still low so that measurements become (Bassi et al., 2022; McCarron et al., 2021; Ozkale & Ozdemir Erdogan, 2022).
The results of the analysis of the end-of-semester test items for the 2022 class increase on the material for fraction arithmetic operations carried out by teachers in Yogyakarta Regency, Indonesia, obtained a reliability value of 0.67; questions that were completely answered by students were 25.40%, the number of students who did not complete was 81.70%; the percentage of absorption power is 50.10%, questions that have good quality distractors are 26.25%; questions that have good discriminating power as much as 26.88%; The difficulty level of the questions is 14.50% very easy, 7.50% easy, 33.50% moderate, 20.75% difficult and 23.75% very difficult. Analysis of the items through the correct proportion and biserial point coefficients obtained the results of 27% of the questions that were accepted, 43% of the questions that had to be revised and 30% of the questions that were rejected. Factors causing the low ability of students in fraction arithmetic operations are a lack of basic understanding, lack of practice, low motivation, ineffective learning approaches and external factors influenced by the teacher in compiling questions. The impact that arises from the preparation of good and correct questions is that students have difficulty understanding the material, errors occur in assessment, low motivation and self-confidence, and lack of skills.
Early indications are oriented toward the instrument that has been used. Based on the results of the item analysis, shows that the test instruments used so far have not been standardized. The analysis carried out by the teacher still does not show independence and is not biased from the questions that have been tested. Therefore, a standardized cognitive test instrument is needed that can be produced through the test development stages (Brandstetter et al., 2017; Loda et al., 2022). The educational assessment standards state that the assessment instruments used must meet the validity and produce scores that can be compared between schools, between regions (Ajjawi et al., 2021; Ormond, 2019). Standard assessment instruments must be valid, reliable and free from bias (Bassi et al., 2022; Leung et al., 2022; Omara et al., 2021).
Research conducted by Ding et al. (2021) states that the development of cognitive assessment instruments is very important to test for feasibility. Layne et al. (2021) explains that developing a test instrument by following systematic steps can provide optimal results. The application of the Rasch model in developing test items can be a tool that performs very well in evaluation and selection of good questions so that the results are good and valid (Jiraniramai et al., 2021; Stolt et al., 2022; Suryadi et al., 2021; Wallace, 2020). Through the development of tests and analysis of the items used, the questions that have been developed can really measure and find out how far students' abilities (Hu et al., 2022; Shi et al., 2019).
The purpose of this study was to describe the quality of cognitive instruments to measure students' abilities in mathematics on fractional arithmetic operations. The cognitive level of the test includes the dimensions of factual knowledge, conceptual knowledge and procedural according to the revised Bloom's cognitive level. The quality of the instrument consists of aspects of validity, reliability, level of difficulty, distractor analysis, individual ability analysis, and the level of item bias. The result of this study was the development of cognitive test instruments for mathematics subject on fractional arithmetic operations in the form of multiple-choice questions consisting of 4 answer choices. Cognitive test instruments are prepared according to national competency standards for grade five elementary school material levels that can be used as good quality question banks.

Research Methods

This research is descriptive quantitative research that aims to validate the measurement of mathematical test instruments on fractional arithmetic operations using the Rasch model. The data collection method used in this study was to use the method of documenting math test scores for fifth grade elementary school students. The sampling technique uses purposive sampling based on the school area, namely schools located in urban areas. The research subjects were fifth grade students at the elementary school level in Yogyakarta (N=65

students). This research was conducted in grade 3 elementary schools in Yogyakarta Indonesia, namely SDN 1 Yogyakarta, SDN 1 Sleman, and SDN 3 Sleman.
The test instrument developed in this study was multiple choice questions consisting of 4 answer choices according to the syllabus for fifth grade elementary school mathematics, especially on fractional arithmetic operations. There are 22 multiple choice questions in the subject of Mathematics on fractional arithmetic operations. The questions consist of easy, medium, and difficult difficulty levels. The correct answer is given a score of 1, and the wrong answer is given a score of 0. Therefore, the data obtained is dichotomous. The data collection techniques used in this study were (1) test techniques, carried out on students using standard cognitive test instrument question sheets, and (2) non-test techniques carried out on students and teachers using questionnaires and interviews.
The instrument used to collect assessment data is a question validation instrument and a developed cognitive instrument. The data analysis used in this study is Rasch Modeling with item response theory (Jiraniramai et al., 2021; Omara et al., 2021; Repo et al., 2019). Quantitative data analysis used the results of the Ministep program analysis to test the Rasch Item Response Theory (IRT) Model (Quansah, 2022; Stolt et al., 2022) on the material test material for fractional arithmetic operations. Table 1 shows the lattice instruments for measuring cognitive mathematical abilities for fifth grade elementary school students on fractional operations.

Table 1. Instruments for Measuring Cognitive Math Skills for Fifth Grade Students

Basic Competencies	Aspects	Subject	Indicator	Item
Solve problems related to exponents (squares of two and three) and taking roots (square roots) of whole numbers	Demonstrate an example of a square number	Numbers raised to the power of two	Recognize square numbers	1,2
			Look for the properties of square numbers	3,4
			Define forms and examples of square numbers	5,6,7,8
	Find the result of taking the square root of a square number	Determining a number to the power of two takes the square root	Perform addition and subtraction operations on square numbers with groups	9,10,11, 12
			Look for the results of drawing square roots	13,14,15 ,16
	Shows forms and examples of cube numbers	Cube number	Find cubes of numbers	17,18,19
			Find the cube root of a number	20,21,2 2

The test is carried out on 4 things, namely: First, whether an item is valid or not depends on the MNSQ value, ZTSD value, and the resulting measurement correlation value. A good MNSQ outfit value for measurement: 0.5< MNSQ <1.5. A good ZSTD outfit value for measurement: -2.0< ZSTD < +2.0. Expected measurement correlation value: 0.4< Pt Measure Corr. <0.85, (Bassi et al., 2022; Geramipour, 2021; McCarron et al., 2021; Omara et al., 2021). Second, testing the reliability of the instrument using summary statistics (Bassi et al., 2022; Ha, 2021; Leung et al., 2022; Wallace, 2020; Zoechling et al., 2022). Third, testing of instrument items that are difficult and easy to approve respondents used the item measure and the item: dimensionality (Leung et al., 2022; Quansah, 2022; Robershaw et al., 2022; Zoechling et al., 2022). Fourth, testing perceptions or bias based on the demographics of respondents using the Differential Item Functional (DIF) plot (Adams et al., 2022). The criterion for a significant difference in perception uses table 30.4 in the Winsteps software with the criterion that the difference in perception is said to be significant if the probability value shows less than 0.05 ($p < 0.05$) (Jiraniramai et al., 2021; Ning, 2018; Peabody & Wind, 2019). The difficulty level of the questions using Rasch modeling was analyzed using the logit numbers contained in the problem measurement column.
Based on the results (output) of the analysis of the Rasch model with the Ministep program, information on the estimated difficulty level values of the test items is obtained which is then categorized according to the range of values with a logits scale categorization. Items with item difficulty level values on the logits scale $\delta<-2.00$ are categorized as very easy, items with $-2.00\leq\delta\leq-1.00$ are categorized as easy, items with $-1.00\leq\delta\leq+1.00$ are categorized moderate, items with $+1.00\leq\delta\leq+2.00$ are categorized as difficult and $-\geq+2.00$ are categorized as very difficult (Mokshein et al., 2019; Pellicciari et al., 2019; Stolt et al., 2022; Suryadi et al., 2021). The differential power of the questions using the Rasch modeling can be analyzed using the Wright map by comparing the distribution of students' abilities and the difficulty of the items as well as the item logit and individual logit values contained in. The expected criterion is that the item logit value is greater than the lowest individual logit value and smaller than the highest individual logit value or is in the good discriminating power category.

Result

a) Instrument Validity and Reliability (Summary Statistics)

Testing the validity of the construct using dimensionality items in Ministep. The criterion for having the ability to measure a range of variables or measuring all respondents is if the Raw Variance Explained by measures is above 40%. The results of the construct validity test are shown in Table 2.

Table 2. Dimensionality Map: Cognitive Test Instruments of Mathematical Ability (Eigenvalue Units)

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	35.5861	100.0%		100.0%
Raw variance explained by measures	=	13.5861	38.2%		37.4%
Raw variance explained by persons	=	8.1929	23.0%		22.5%
Raw Variance explained by items	=	5.3932	15.2%		14.8%
Raw unexplained variance (total)	=	22.0000	61.8%	100.0%	62.6%
Unexplned variance in 1st contrast	=	5.0781	14.3%	23.1%	
Unexplned variance in 2nd contrast	=	3.1583	8.9%	14.4%	
Unexplned variance in 3rd contrast	=	2.3643	6.6%	10.7%	
Unexplned variance in 4th contrast	=	2.1549	6.1%	9.8%	
Unexplned variance in 5th contrast	=	1.7933	5.0%	8.2%	

Table 3 found that the value possessed by the Raw Variance Explained by measures was 38.20% of the raw variance explained by measures which exceeded the expected 37.40%. However, 38.20% of the variance of the data based on the dimensions of the Rasch measurement model is considered weak according to the rating scale instrument quality criteria by (Mokshein et al., 2019). According to (Shin et al., 2020; Stolt et al., 2022; Suryani et al., 2021) the minimum requirement of 20% instrument uniformity has been achieved, but the minimum requirement of 40% of the Rasch measurement model has not been met.

The raw variance explained by people at 23.0% indicates that there is more variation in people's abilities compared to 15.20% of the difficulty level of the items. This is due to the smaller standard deviation for people (6.70) compared to the standard deviation for items (7.0). The variance that cannot be explained by the first contrast is 5%, with an eigenvalue of 5.0781 (<3.0), indicating the presence of a second dimension and the test is dimensionality (Ning, 2018; Pellicciari et al., 2019). Based on this, the cognitive test instrument for measuring mathematical ability in fractional arithmetic operations for fifth grade elementary school students has construct validity or is able to measure a range of variables or measure all respondents. The results of the reliability test are shown in Table 3.

Table 3. Item and Person Reliability

SUMMARY OF 65 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	13.7	22.0	.82	.39				
SEM	.8	.0	.29	.06				
P.SD	6.7	.0	2.30	.48				
S.SD	6.7	.0	2.32	.48				
MAX.	22.0	22.0	4.58	1.85				
MIN.	.0	22.0	-4.70	.46				

REAL RMSE .94 TRUE SD 2.10 SEPARATION 2.23 PERSON RELIABILITY .83
MODEL RMSE .93 TRUE SD 2.10 SEPARATION 2.27 PERSON RELIABILITY .84
S.E. OF PERSON MEAN = .29

PERSON RAW SCORE-TO-MEASURE CORRELATION = .97 (approximate due to missing data)
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .93 SEM = 1.74 (approximate due to missing data)
STANDARDIZED (50 ITEM) RELIABILITY = .92

SUMMARY OF 22 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	40.4	65.0	.00	.35	.99	-.07	1.00	.03
SEM	1.5	.0	.19	.01	.05	.24	.07	.10
P.SD	7.0	.0	.88	.04	.21	1.09	.31	.81
S.SD	7.1	.0	.90	.04	.21	1.12	.31	.83
MAX.	55.0	65.0	1.42	.48	1.50	2.32	1.57	1.65
MIN.	20.0	65.0	-2.14	.33	.72	-1.50	.57	-1.19

REAL RMSE .37 TRUE SD .80 SEPARATION 2.15 ITEM RELIABILITY .82
MODEL RMSE .36 TRUE SD .81 SEPARATION 2.27 ITEM RELIABILITY .84
S.E. OF ITEM MEAN = .19

Instrument reliability with Rasch modeling was analyzed using individual separation values and item separation as well as Cronbach's Alpha values displayed in the output of the Rasch Model program. The expected criteria for individual and item separation values are at least 0.70 and Cronbach's Alpha is at least 0.70 or in the sufficient category (Bassi et al., 2022; Leung et al., 2022; Wallace, 2020; Zoechling et al., 2022). The result of person reliability is 0.83, item reliability is 0.82, and CA is 0.93. All three aspects have been fulfilled so that it can be concluded that the reliability level of the instrument for measuring the cognitive abilities of fifth grade students in the material of fractional arithmetic operations can be trusted and relied upon.

b) Item Measure and Fit Statistic (Item Validity)

The output fit order items in the Rasch model analysis are shown in Table 4. The first aspect there are 2 items, namely Q6 and Q7 with an Outfit MNSQ value of 1.57, while what is accepted is 0.5 < MNSQ < 1.5. The second aspect of the Outfit ZSTD value for all items is accepted with a value of -2.0 < ZSTD < +2.0. The third aspect is the value of Pt Mean Corr. ranging from 0.4 < Pt Measure Corr. < 0.85. Even though the MNSQ values in Q6 and Q7 were not met, but the ZSTD and Pt Measure Corr values were met, then the 22 items were considered fit and could be maintained. (Bassi et al., 2022; Geramipour, 2021) classifies the Pt Mean Corr. value as very good (>0.40), then the 22 items of the instrument were able to discriminate against the items to measure mathematical ability in fractional arithmetic operations material for fifth grade elementary school.

Table 4. Item (Column) Test Results: Fit Order

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD		PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
6	40	65	.[illegible]1	.34	1.4[illegible]	2.32	1.57	1.65	A	.50	.63	60.0	76.3	Q6
7	28	65	1.[illegible]2	.33	1.1[illegible]	[illegible]	1.57	1.35	B	.55	.62	70.9	75.5	Q7
12	55	65	-2.[illegible]8	.48	1.5[illegible]	1.56	1.12	.43	C	.50	.60	83.6	89.3	Q12
18	37	65	.[illegible]4	.33	1.2[illegible]	1.36	1.48	1.48	D	.56	.63	70.9	74.7	Q18
8	53	65	-1.[illegible]2	.44	.8[illegible]	-.54	1.47	.80	E	.62	.61	89.1	86.0	Q8
21	36	65	.[illegible]5	.33	1.1[illegible]	.76	1.26	.87	F	.59	.63	70.9	74.7	Q21
2	37	65	.[illegible]4	.33	1.2[illegible]	1.48	1.20	.72	G	.57	.63	63.6	74.7	Q2
22	50	65	-1.[illegible]0	.40	1.0[illegible]	.26	1.22	.56	H	.60	.62	85.5	84.2	Q22
19	36	65	.[illegible]5	.33	1.1[illegible]	.84	.96	-.01	I	.60	.63	70.9	74.7	Q19
16	33	65	.[illegible]7	.33	1.0[illegible]	.19	.95	-.06	J	.62	.63	74.5	74.3	Q16
15	46	65	-.[illegible]2	.36	.9[illegible]	-.21	.79	-.41	K	.65	.63	83.6	81.1	Q15
13	32	65	.[illegible]0	.33	.9[illegible]	-.47	.88	-.25	k	.65	.63	78.2	74.4	Q13
10	33	65	.[illegible]7	.33	.9[illegible]	-.70	.83	-.44	j	.66	.63	78.2	74.3	Q10
5	40	65	.[illegible]1	.34	.8[illegible]	-.71	.87	-.34	i	.66	.63	85.5	76.3	Q5
4	44	65	-.[illegible]7	.35	.8[illegible]	-.95	.87	-.24	h	.67	.63	87.9	79.6	Q4
1	43	65	-.[illegible]4	.35	.8[illegible]	-.75	.73	-.72	g	.67	.63	81.8	78.8	Q1
9	41	65	-.[illegible]1	.34	.8[illegible]	-.87	.78	-.62	f	.67	.63	80.0	77.0	Q9
20	31	65	1.[illegible]0	.33	.8[illegible]	-.93	.86	-.32	e	.66	.63	80.0	74.6	Q20
11	38	65	.[illegible]3	.33	.8[illegible]	-1.00	.82	-.51	d	.67	.63	87.3	75.2	Q11
17	44	65	-.[illegible]7	.35	.79	-1.17	.64	-1.01	c	.69	.63	87.3	79.6	Q17
3	45	65	-.[illegible]9	.36	.7[illegible]	-1.50	.57	-1.19	b	.71	.63	89.1	80.4	Q3
14	46	65	-.[illegible]2	.36	.7[illegible]	-1.51	.59	-1.04	a	.70	.63	90.9	81.1	Q14
MEAN	40.4	65.0	.00	.35	.99	-.07	1.00	.03				79.5	78.1	
P.SD	7.0	.0	.88	.04	.21	1.09	.31	.81				8.4	4.2	

c) Local Independence

The results of the validity test by looking at the results of the rating test (partial credit) scale in Table 4 found that each rating (0 and 1) has a separate peak; or it means that the probability of each rating is clearly different from the respondents (Chong et al., 2022; Suryadi et al., 2021). Based on Figure 1, it can be concluded that the test instrument measuring mathematical ability in fractional arithmetic operations material can be distinguished by the respondent's rating.

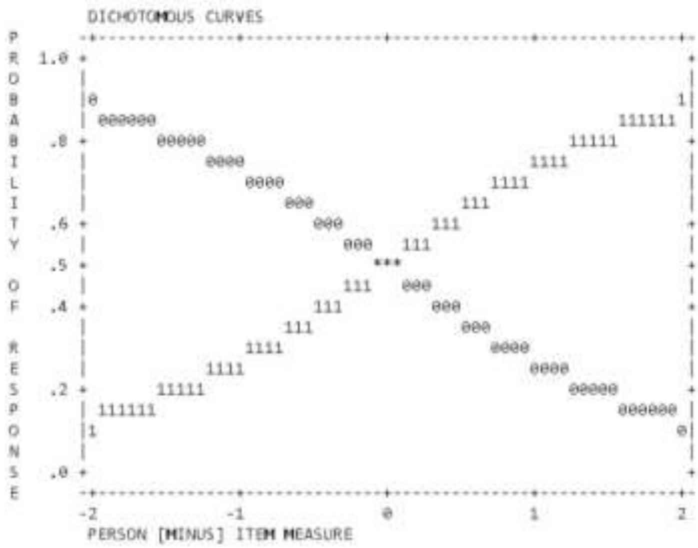

Figure 1. Rating Test Results (Partial-Credit) Scale

d) The Wright Map: Rating Scale Diagnostics

The output results on the measure are related to the wright map in Figure 2. The difficulty level of the items is sorted from high to low, from difficult to easy. Item Q7 is a very difficult item to be answered by fifth grade students on fractional arithmetic operations. It was proven from the number of respondents who answered correctly, there were 28 students and the easiest question, namely number Q12, could be answered by 55 students. The same logit value for each item showed that the level of difficulty of the item was not much different, as in item numbers Q19 and Q21, Q5 and Q6, in addition to Q4 and Q7. Items with item difficulty level values on the logits scale $\delta<-2.00$ are categorized as very easy, namely Q12, items with $-2.00\leq\delta\leq-1.00$ are categorized as easy Q22 and Q8, items with $-1.00\leq\delta\leq +1.00$ is categorized as moderate, namely 17 questions, items with +1.00≤δ≤+2.00 are categorized as difficult, namely Q7 and Q20 and -≥+2.00 no questions are categorized as very difficult.

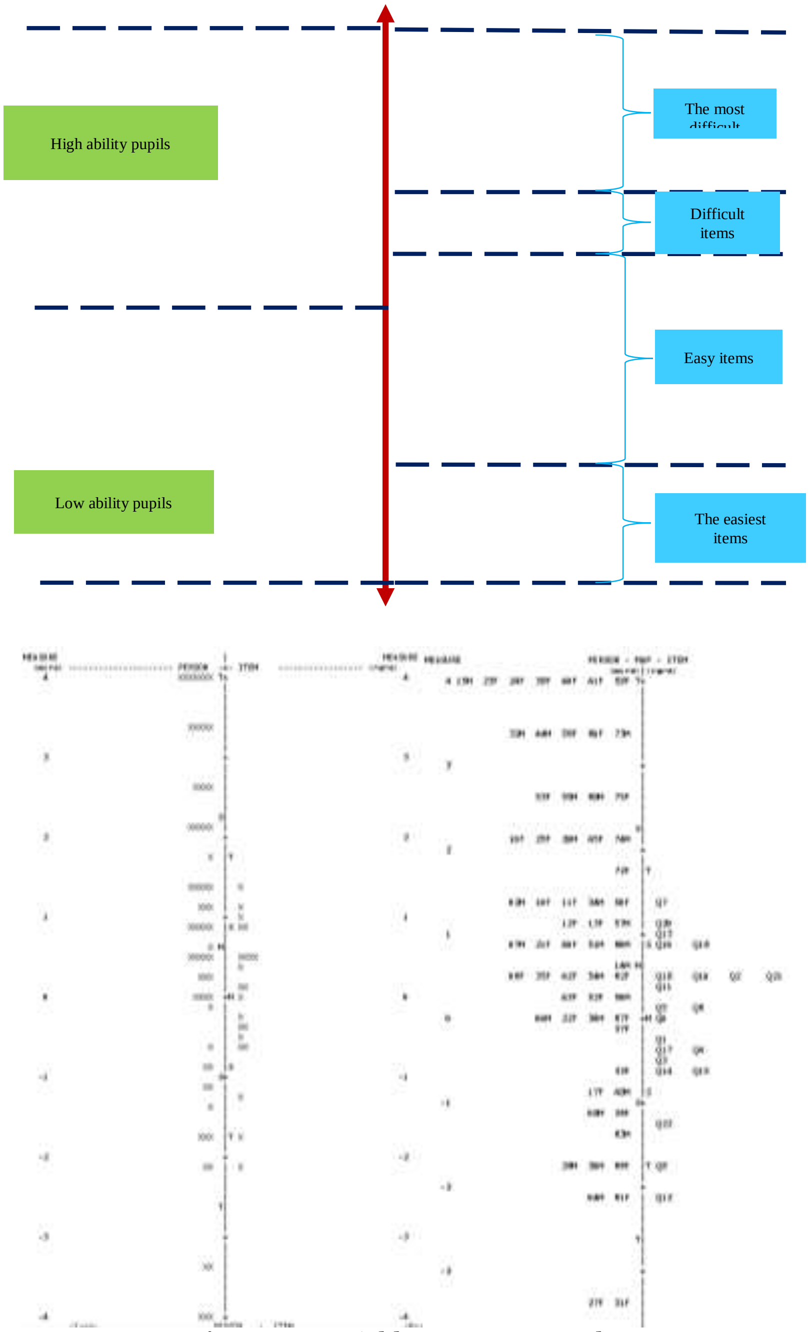

Figure 2. Variable Map Test Results

The most difficult item is item number 7 (Q7) which is at the very top. Meanwhile, item Q12 is the easiest for students to answer according to their abilities. Theoretically, with this question, there will be subjects who have the opportunity to answer the question correctly because they have a higher ability than the difficulty level of the question. The results of the item difficulty level analysis are shown in Table 4.

e) Person Ability Analysis

The meters are used as well as to determine the accuracy of items, namely the mean-square outfit, outfit-zstd and point measure correlation (Leung et al., 2022; Quansah, 2022; Robershaw et al., 2022; Zoechling et al., 2022). Values that are outside the limits of statistical precision indicate response patterns that need to be known further. In the winstep program, this information table can be displayed using the person first order function (Table 5) which sorts those that do not fit to the top. Table 5 shows that the 29M test participants (at the top) have a very disproportionate response pattern compared to the others, as well as for 02M and 63M.

Table 5. Function of Person First Order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE 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
25	19	22	2.09	.64	1.31	.39	3.75	2.00	A -.42	.23	86.4	86.3	29M
1	17	22	1.41	.53	1.15	.62	2.15	2.02	B -.08	.29	72.7	77.2	02M
57	5	22	-1.48	.55	1.38	1.21	2.07	2.08	C -.15	.37	77.3	79.5	63M
32	6	22	-1.12	.52	1.43	1.58	1.89	2.13	D -.14	.38	68.2	76.3	36F
35	4	22	-1.73	.59	1.11	.48	1.70	1.39	E .14	.35	77.3	82.7	39M
30	17	22	1.41	.53	1.27	1.09	1.58	1.21	F -.08	.29	81.8	77.2	34M
45	15	22	.90	.49	1.45	2.10	1.48	1.37	G -.15	.33	54.5	69.9	51M
55	3	22	-2.12	.66	1.36	.83	1.48	.85	H -.05	.33	81.8	86.4	61F
39	12	22	.24	.46	1.41	2.40	1.33	1.44	I -.03	.37	36.4	65.7	43F
16	19	22	2.09	.64	1.10	.50	1.26	.57	J .03	.23	86.4	86.3	19F
64	19	22	2.09	.64	1.16	.59	1.26	.57	K .03	.23	86.4	86.3	74M
15	7	22	-.86	.50	1.19	.80	1.25	.80	L .17	.39	72.7	73.3	17F
29	8	22	-.63	.48	1.09	.43	1.21	.86	M .17	.39	77.3	70.1	33F
38	13	22	.45	.46	.85	-.96	1.21	.87	N .44	.36	81.8	66.0	42F
48	13	22	.45	.46	.85	-.96	1.21	.87	O .44	.36	81.8	66.0	54M
56	13	22	.45	.46	.85	-.96	1.21	.87	P .44	.36	81.8	66.0	62F
4	13	22	.45	.46	1.17	1.13	1.18	.70	Q .17	.36	54.5	66.0	06F
23	4	22	-1.73	.59	1.12	.43	1.15	.43	R .21	.35	86.4	82.7	26M
18	15	22	.90	.49	1.12	.60	1.01	.14	S .23	.33	63.6	69.9	21F
46	12	22	.24	.46	1.12	.83	1.11	.53	T .25	.37	54.5	65.7	52F
11	16	22	1.15	.50	1.11	.55	.98	.08	U .22	.31	68.2	73.3	13F
44	17	22	1.41	.53	1.10	.48	.96	.07	V .21	.29	72.7	77.2	50F
65	20	22	2.57	.76	1.09	.36	1.05	.36	W .08	.19	90.9	90.9	75F
42	7	22	-.86	.50	1.02	.13	1.00	.36	X .34	.39	81.8	73.3	40M
49	20	22	2.57	.76	1.07	.36	.97	.27	Y .12	.19	90.9	90.9	55M
8	17	22	1.41	.53	1.04	.25	.85	-.18	Z .28	.29	72.7	77.2	10F
40	21	22	3.33	1.09	1.02	.33	.70	.20	.14	.14	95.5	95.4	44M
41	19	22	2.09	.64	1.00	.19	.74	-.18	z .28	.23	86.4	86.3	45F
62	18	22	1.72	.57	1.00	.13	.77	-.26	y .31	.26	81.8	81.8	72F
2	3	22	-2.12	.66	.96	.06	.97	.17	x .35	.33	90.9	86.4	04M
47	20	22	2.57	.76	.93	.06	.76	.02	w .27	.19	90.9	90.9	53F

BETTER FITTING NOT SHOWN

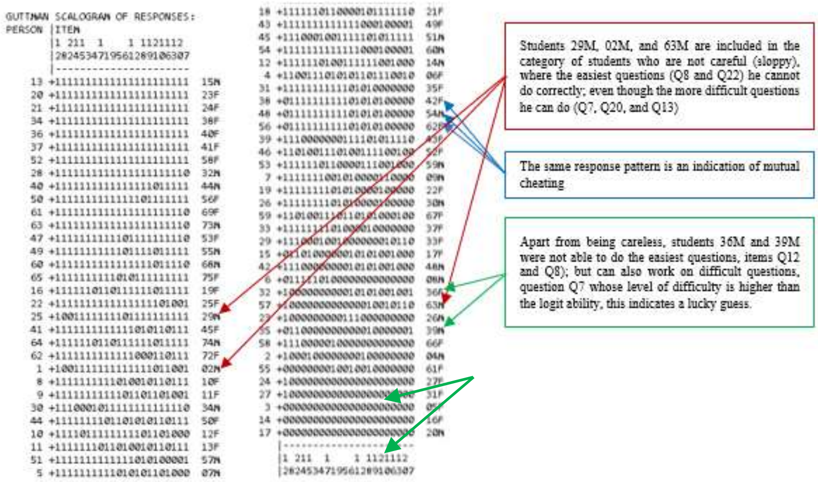

Figure 3. Individual Ability Analysis (Guttman Scalogram)

Information on unusual response patterns from the Pearson measure can be seen in more detail by looking at the scalogram (Figure 3). Through the Gutman matrix, it can be directly identified the cause why the response pattern does not match the model (Pellicciari et al., 2019). Another advantage of the scalogram is that it can detect cheating, namely students cheating and lucky guessing. The initial indication is to see if the same person value is found.

f) The Differential Item Functioning (DIF) Analysis

The test results use the Differential Item Functional (DIF) on the probability value of Table 6, where the difference in perception is stated to be significant if the p value <0.05 (Jiraniramai et al., 2021; Ning, 2018; Peabody & Wind, 2019), then as shown in Table 5 that between male and female students did not have a significant difference in perception on all 22 items. All probability values obtain a value of ρ => 0.05 (ρ ranges from 0.0577 to 1.000). This means that the 22 items are not perceived differently and there are no items that need to be corrected.

Table 6. The DF Test Results Based on Gender

DIF class/group specification is: DIF=$S3W1

PERSON CLASSES	SUMMARY DIF CHI-SQUARED	D.F.	PROB.	BETWEEN-CLASS/GROUP UNWTD MNSQ	ZSTD	ITEM Number	Name
2	1.7311	1	.1883	1.8096	.94	1	Q1
2	1.1525	1	.2880	1.1951	.60	2	Q2
2	.0289	1	.8650	.0288	-1.00	3	Q3
2	.8819	1	.3477	.9105	.41	4	Q4
2	.0619	1	.8035	.0623	-.81	5	Q5
2	.8433	1	.3585	.8704	.38	6	Q6
2	.1448	1	.7035	.1471	-.53	7	Q7
2	1.2789	1	.2581	1.3259	.68	8	Q8
2	.0000	1	1.0000	.0016	-1.40	9	Q9
2	.0443	1	.8334	.0440	-.90	10	Q10
2	.6452	1	.4218	.6642	.20	11	Q11
2	.0444	1	.8332	.0435	-.90	12	Q12
2	.6843	1	.4081	.7044	.24	13	Q13
2	.0225	1	.8807	.0214	-1.06	14	Q14
2	.3402	1	.5597	.3487	-.16	15	Q15
2	2.2458	1	.1340	2.3660	1.18	16	Q16
2	.0593	1	.8076	.0577	-.83	17	Q17
2	1.1525	1	.2880	1.1951	.60	18	Q18
2	.4785	1	.4891	.4932	.03	19	Q19
2	.3071	1	.5795	.3131	-.21	20	Q20
2	3.6016	1	.0577	3.8742	1.68	21	Q21
2	.0015	1	.9689	.0015	-1.41	22	Q22

The graph shows the relative item difficulty level for each group. The higher the graph point, the more difficult the item is for that group (Figure 4). There are three curves based on gender, namely M (male), F (female), and star (*) indicating the average value. From the graph, it can be roughly seen that the distance between the DIF measure values between L and P that is farthest is in items Q21 (ρ = 0.0577) and Q16 (ρ = 0.1340). Meanwhile, for other items, the distance between F and M is not too far. Even though Q21 and Q16 are long distance ranges, the value of ρ (probability) is more than 0.05. This shows that in all items there is no difference in difficulty level between men and women. In this case neither sex has an advantage because the items appear to be more difficult for men or women.

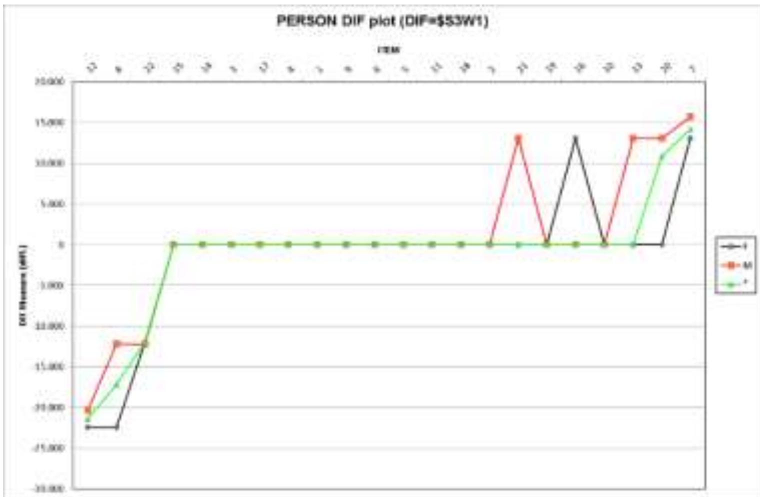

Figure 4. Graph of Differential Item Functioning (DIF) plot by Gender

g) Distractor Analysis

Question number Q7 with the highest logit value is 1.42 which indicates that the most difficult questions to choose are the most option 0 "zero". This can be interpreted those 37 students (57%) of all students answered question Q7 incorrectly, and so on. Based on the ability value, there is no option score value marked with an asterisk (*), this indicates that the 22 items on the cognitive test instrument for measuring mathematical ability in the material for counting operations have a good level of deception. So that there are no items that experience distractor problems that require items to require a repair process (McCarron et al., 2021; Shin et al., 2020). The distractor analysis is shown in Table 7.

Table 7. Distractor Analysis

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM
6 A	0	0	25	38	-.65	2.16	.44	1.4	1.6	-.50	Q6
	1	1	40	62	1.74	1.87	.30	1.5	1.6	.50	
7 B	0	0	37	57	-.28	2.08	.35	1.3	1.7	-.55	Q7
	1	1	28	43	2.28	1.68	.32	1.0	1.5	.55	
12 C	0	0	10	15	-1.86	2.05	.68	1.6	1.0	-.50	Q12
	1	1	55	85	1.31	1.99	.27	1.7	1.8	.50	
18 D	0	0	28	43	-.65	2.02	.39	1.1	1.3	-.56	Q18
	1	1	37	57	1.94	1.82	.30	1.3	1.7	.56	
8 E	0	0	12	18	-2.16	2.15	.65	.9	1.6	-.62	Q8
	1	1	53	82	1.50	1.72	.24	.7	.7	.62	
21 F	0	0	29	45	-.69	2.17	.41	1.4	1.7	-.59	Q21
	1	1	36	55	2.04	1.56	.26	.9	.7	.59	
2 G	0	0	28	43	-.68	1.99	.38	1.1	1.0	-.57	Q2
	1	1	37	57	1.96	1.82	.30	1.4	1.5	.57	
22 H	0	0	15	23	-1.68	2.15	.57	1.2	1.3	-.60	Q22
	1	1	50	77	1.57	1.75	.25	.9	.9	.60	
19 I	0	0	29	45	-.73	1.97	.37	1.1	.9	-.60	Q19
	1	1	36	55	2.07	1.71	.29	1.2	1.0	.60	
16 J	0	0	32	49	-.64	1.90	.34	.9	.8	-.62	Q16
	1	1	33	51	2.24	1.69	.30	1.2	1.1	.62	
15 K	0	0	19	29	-1.49	1.92	.45	.9	.7	-.65	Q15
	1	1	46	71	1.78	1.68	.25	.9	.9	.65	
13 k	0	0	33	51	-.65	1.85	.33	.8	.7	-.65	Q13
	1	1	32	49	2.34	1.64	.29	1.0	1.0	.65	
10 j	0	0	32	49	-.72	1.84	.33	.8	.7	-.66	Q10
	1	1	33	51	2.32	1.62	.29	1.0	1.0	.66	
5 i	0	0	25	38	-1.10	1.92	.39	.9	.9	-.66	Q5
	1	1	40	62	2.02	1.59	.25	.8	.8	.66	
4 h	0	0	21	32	-1.41	1.98	.44	1.0	1.0	-.67	Q4
	1	1	44	68	1.89	1.56	.24	.7	.6	.67	
1 g	0	0	22	34	-1.34	1.79	.39	.8	.6	-.67	Q1
	1	1	43	66	1.93	1.66	.26	1.0	1.0	.67	
9 f	0	0	24	37	-1.20	1.93	.40	.9	.9	-.67	Q9
	1	1	41	63	2.01	1.55	.25	.8	.6	.67	
20 e	0	0	34	52	-.63	1.82	.32	.8	.7	-.66	Q20
	1	1	31	48	2.42	1.60	.29	1.0	1.0	.66	
11 d	0	0	27	42	-1.01	1.88	.37	.9	.8	-.67	Q11
	1	1	38	58	2.12	1.57	.26	.8	.8	.67	
17 c	0	0	21	32	-1.48	1.84	.41	.8	.6	-.69	Q17
	1	1	44	68	1.92	1.57	.24	.7	.6	.69	
3 b	0	0	20	31	-1.62	1.78	.41	.7	.6	-.71	Q3
	1	1	45	69	1.91	1.55	.23	.7	.6	.71	
14 a	0	0	19	29	-1.70	1.80	.42	.8	.6	-.70	Q14
	1	1	46	71	1.86	1.56	.23	.7	.6	.70	

h) Test Information Function

From the Figure 5, we can conclude that of the 22 questions that we presented to 65 subjects, the items indicated were suitable for knowing the ability level of moderate and below average students. Based on the results of the information function test, it can be seen that the mathematical cognitive test to measure the reasoning ability of fraction arithmetic operations in fifth grade elementary school students is more suitable for end-of-semester or grade-grade assessment tests. In addition, the item information function also shows the reliability of measurement (Jiraniramai et al., 2021). This is supported by the results of the analysis of the reliability coefficient of the cognitive math test in table 1, which is 0.93 with a measure or logit value = 0, indicating that cognitive test tests are suitable for end-of-semester tests, so the measurement information obtained is very high.

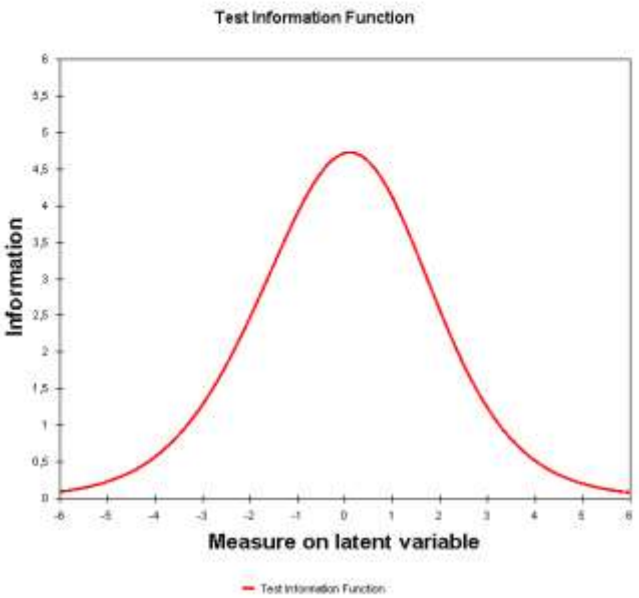

Figure 5. Graph of Measurement Information Functions

Discussion

The aim of developing a mathematics learning outcomes instrument for fifth grade elementary school students was to determine the level of validity, reliability, item difficulty, deception, and bias analysis in fractional arithmetic operations. Based on the results of item difficulty level analysis using the Rasch Model item response theory (IRT) method using Ministep software, it was found that the 22 items fit to measure the mathematical cognitive abilities of fifth grade students based on fractional arithmetic operations. There are 3 items that are included in the difficult category, 16 items in the moderate category, and 3 items in the easy category. This shows that the items in this category are mostly used by teachers for test takers. Likewise, research Mokshein et al. (2019) the level of difficulty of the test items obtained results, namely the difficult category of 50.0%, the medium category of 42.5%, and the easy category of 7.5%. Thus, questions that measure mathematical cognitive abilities are good items that are mostly used in measuring students' mathematics learning outcomes (Anderson et al., 2020; Shen et al., 2019; Trinidad, 2020; Yeganeh et al., 2022).

A good instrument must also have the ability to cover all aspects being measured. The Rasch model is software that can be used to evaluate whether an instrument has the ability to measure (construct validity) (Bassi et al., 2022; Suryadi et al., 2021; Wallace, 2020), because the Rasch model is able to show instrument items that are difficult to agree on or carried out by the respondent (Leung et al., 2022; Quansah, 2022) as well as the ability of the respondent (Geramipour, 2021; Leung et al., 2022; Robershaw et al., 2022; Zoechling et al., 2022). This study found 3 items that were difficult for respondents to agree on, namely determining the shape and example of squared numbers (Q7), looking for the results of taking square roots (Q13), and taking triple roots (Q20). While the ability of students who can answer the three difficult questions totaled 30 out of 65 students. This identifies that the level of student ability is equivalent to the level of difficulty of the item being developed. This research is consistent with the results of studies by Bassi et al. (2022); Omara et al. (2021); Robershaw et al. (2022), that the development of instruments recommends that questions can be reached evenly where questions are difficult to answer by high student competence, and vice versa.

Based on the results of the analysis of the test results of the students, it was found that out of all 22 items, they were included in the valid category (Table 4). Even though 2 items (Q6 and Q7) on the MNSQ Outfit value were not met, namely 1.57 (> 1.50), the ZSTD and Pt Measure Corr outfit aspects were fulfilled. These results support research by Bassi et al. (2022); Mokshein et al. (2019); Wiyarsi et al. (2019), that the 2 items can still be tolerated because 2 of 3 aspects of validity measurement are met. This indicates that students do not understand the instructions because they make mistakes that should not have happened. So that the sentence editor is simplified again according to the characteristics of fifth grade elementary school students. Approximately 70% of the 30 test takers made mistakes, especially in the forms and examples of square numbers. Based on these findings it is evident that the advantages of the Rasch Model in analyzing the validity of the instrument can be done from several aspects so that the resulting instrument can be more reliable which supports research by Barber et al. (2022); Geramipour (2021). The validity analysis using the Rasch Model can be said to be better because of its consistency (Jiraniramai et al., 2021; Stolt et al., 2022). Thus, Rasch modeling can help to tackle item measurement in the right way.

Assessment in the form of tests can be used to hone students' cognitive thinking skills, and have an effect on determining students' thinking skills (Brandstetter et al., 2017; Loda et al., 2022; Nurtanto, Sudira, Sofyan, & Jatmoko, 2022). Likewise, research by Cogan et al. (2019); Ozkale & Ozdemir Erdogan (2022) related to the development of PISA model math questions to find out student arguments, concluded that a prototype of the PISA model math problem set was produced to find out student arguments as many as 6 questions in the form non-objective description (open construct response). In addition, research by (Cantley, 2019; Teng, 2020) recommends a prototype of the PISA model mathematical problem set to determine student arguments produced have a potential effect on student arguments. This shows that in measuring student learning outcomes, especially in learning mathematics in elementary schools, it is customary to use test questions that measure higher-order thinking to hone mathematical problem-solving skills.

A good instrument must also avoid ambiguous items (Omara et al., 2021; Stolt et al., 2022). Ambiguous instruments must be avoided because they will be interpreted differently so the results will be different (Chi et al., 2021; Omara et al., 2021; Suryadi et al., 2021), using the Rasch model properly can help to avoid ambiguous or incapable instruments well understood by respondents (Chi et al., 2021). In this study, the Rasch model showed that of the 22 items, there were no instrument items that were perceived significantly differently by male and female respondents (students). So that all items do not identify bias questions that only benefit one party. This research supports and is consistent with research (Purwanto et al., 2020) that in developing the instrument, items that contain bias should not be found.

Cognitive is the thinking ability possessed by each individual in doing something or solving a problem (Brandstetter et al., 2017; Stelzer et al., 2021). Students' cognitive understanding in learning mathematics in fractional arithmetic operations material during online learning is still not very good. Based on the results of observations, interviews, and tests on fifth grade students, students' cognitive understanding is still not good because students have different levels of understanding. The research results are consistent with research by (Lyons et al., 2021) and Shen et al. (2019) that this is one of the reasons for the lack of ability and activity of students while studying at home.

There are students' cognitive understanding with high, medium, and low cognitive understanding. Students who have high cognitive understanding are able to solve questions properly and correctly and these students answer and explain the questions posed by researchers. The results of the research also support the statement of the results of the studies (Ding et al., 2021; Layne et al., 2021) that students whose cognitive understanding is moderate are still not precise in solving problems well, and less precise in answering and explaining questions from researchers. While the results of studies by Trinidad (2020); Yeganeh et al. (2022) support this research that students' low cognitive understanding is unable to solve problems properly, and cannot answer and explain questions from researchers.

In addition, the role of parents as facilitators is considered to be unable to guide students during online learning. Because during online learning students still do not give maximum results. Research by Pakaja & Wafa (2021) confirms that the educational background of parents also influences student learning. On the other hand, research by Reilly & Levintova (2021) is in line with the results of interviews. This research identified that as a result, material that was not understood by students was not fully able to be explained back by students. Learning mathematics is something that is difficult for students to understand. So this research is in line with Lyons et al. (2021), recommending that students need more guidance in understanding learning mathematics.

Students' cognitive abilities during online learning are very different from face-to-face learning. Student learning outcomes that are different from face-to-face learning are a consideration for the teacher in determining student scores. Students who usually have low cognitive abilities when learning online have very high scores (Anderson et al., 2020; Yeganeh et al., 2022). In contrast, research by Loda et al. (2022); Wang & Stein (2021) is inconsistent confirming that students with high cognitive abilities have mediocre scores during online learning which can be compared with this study. Based on the findings obtained, there is a student's ability to work on different fraction arithmetic operations. As explained by the researcher, fifth grade students have different levels of cognitive understanding during online learning. Students' cognitive abilities during online learning are very different when students carry out face-to-face learning. There is a high, medium, and low cognitive understanding of students. This is in line with Bloom's Taxonomy theory of the cognitive realm from the simple (knowing) to the more complex (evaluating).

Conclusion

Based on the results of the Rasch Model analysis on 22 cognitive instruments to measure math skills in fifth grade students, it was revealed that the test instruments were valid, reliable and could measure according to the capacities of elementary school students. This is indicated by the item reliability of 0.82, and the separation of 2.15. Meanwhile, person reliability was 0.83, and separation was 2.23. There are two questions that have a difficult category number Q7 and Q20, 17 questions with a moderate level of difficulty, while easy questions are Q22 and Q8, questions with the easiest difficulty are Q12. DIF analysis shows that all items do not contain gender bias, so there is no difference in the level of difficulty between men and women in working on cognitive test questions. So that the test instrument can be generalized to measure students' mathematical abilities in fifth grade elementary schools. Cognitive test instruments are prepared according to national competency standards for grade five elementary school material levels that can be used as good quality question banks.

This research is inseparable from the ability of elementary school teachers to teach mathematics interestingly. The activities carried out by the teacher include learning using visuals, interactively by involving various activities in learning, using fun materials, helping students create basic concepts, helping students make connections between mathematics and everyday life, and providing guidance and support to the student. In addition, teachers can use various methods to improve students' cognitive abilities including problem-solving, group discussions, practical activities, brain training methods, use of technology, learning through experience, and focusing on skills. The level of cognitive skills taught by the teacher takes into account the levels suggested by Bloom's Theory, namely understanding, application, synthesis, evaluation, and creation. We recommend skill-based cognitive and affective learning with several stages, namely (a) observing students in applying new concepts and skills to real situations; (b) skills tests to measure understanding and ability to apply new concepts and skills; (c) a survey to find out students' feelings and the understanding gained; (d) discuss with students to improve skills in the future; and (e) analysis of data to determine how effective what has been improved is and how to proceed.

References

1. Adams, D., Chuah, K. M., Sumintono, B., & Mohamed, A. (2022). Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: A Rasch analysis. *Asian Education and Development Studies*, *11*(2), 324–339. https://doi.org/10.1108/AEDS-05-2020-0100
2. Ajjawi, R., Bearman, M., & Boud, D. (2021). Performing standards: A critical perspective on the contemporary use of standards in assessment. *Teaching in Higher Education*, *26*(5), 728–741. https://doi.org/10.1080/13562517.2019.1678579
3. Alonzo, A. C. (2018). An argument for formative assessment with science learning progressions. *Applied Measurement in Education*, *31*(2), 104–112. https://doi.org/10.1080/08957347.2017.1408630

4. Anderson, K., Gong, X., Hong, K., & Zhang, X. (2020). The impacts of transition to middle school on student cognitive, non-cognitive and perceptual developments: Evidence from China. *Education Economics*, *28*(4), 384–402. https://doi.org/10.1080/09645292.2020.1749234
5. Barber, C. C., Middlemiss, W., & Medvedev, O. N. (2022). Applying Rasch methodology to examine and enhance precision of the Edinburgh Postnatal Depression Scale. *Journal of Affective Disorders*, *308*, 391–397. https://doi.org/10.1016/j.jad.2022.04.009
6. Bassi, I., Carzedda, M., Gori, E., & Iseppi, L. (2022). Rasch analysis of consumer attitudes towards the mountain product label. *Agricultural and Food Economics*, *10*(1), 1–13. https://doi.org/10.1186/s40100-022-00218-7
7. Brandstetter, M., Sandmann, A., & Florian, C. (2017). Understanding pictorial information in biology: Students' cognitive activities and visual reading strategies. *International Journal of Science Education*, *39*(9), 1218–1237. https://doi.org/10.1080/09500693.2017.1320454
8. Calderon, O., & Sood, C. (2020). Evaluating learning outcomes of an asynchronous online discussion assignment: A post-priori content analysis. *Interactive Learning Environments*, *28*(1), 3–17. https://doi.org/10.1080/10494820.2018.1510421
9. Cantley, I. (2019). PISA and policy-borrowing: A philosophical perspective on their interplay in mathematics education. *Educational Philosophy and Theory*, *51*(12), 1200–1215. https://doi.org/10.1080/00131857.2018.1523005
10. Charalambous, C. Y., Kyriakides, E., Kyriakides, L., & Tsangaridou, N. (2019). Are teachers consistently effective across subject matters? Revisiting the issue of differential teacher effectiveness. *School Effectiveness and School Improvement*, *30*(4), 353–379. https://doi.org/10.1080/09243453.2019.1618877
11. Chi, S., Liu, X., & Wang, Z. (2021). Comparing student science performance between hands-on and traditional item types: A many-facet Rasch analysis. *Studies in Educational Evaluation*, *70*, 1–9. https://doi.org/10.1016/j.stueduc.2021.100998
12. Chong, J., Mokshein, S. E., & Mustapha, R. (2022). Applying the Rasch Rating Scale Model (RSM) to investigate the rating scales function in survey research instrument. *Cakrawala Pendidikan*, *41*(1), 97–111. https://doi.org/10.21831/cp.v41i1.39130
13. Cogan, L. S., Schmidt, W. H., & Guo, S. (2019). The role that mathematics plays in college- and career-readiness: Evidence from PISA. *Journal of Curriculum Studies*, *51*(4), 530–553. https://doi.org/10.1080/00220272.2018.1533998
14. Ding, M., Hassler, R., & Li, X. (2021). Cognitive instructional principles in elementary mathematics classrooms: A case of teaching inverse relations. *International Journal of Mathematical Education in Science and Technology*, *52*(8), 1195–1224. https://doi.org/10.1080/0020739X.2020.1749319
15. Geramipour, M. (2021). Rasch testlet model and bifactor analysis: How do they assess the dimensionality of large-scale Iranian EFL reading comprehension tests? *Language Testing in Asia*, *11*(1), 1–10. https://doi.org/10.1186/s40468-021-00118-5
16. Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, *11*(1), 1–16. https://doi.org/10.1186/s40468-021-00132-7
17. Hu, X., Jiang, Y., & Bi, H. (2022). Measuring science self-efficacy with a focus on the perceived competence dimension: Using mixed methods to develop an instrument and explore changes through cross-sectional and longitudinal analyses in high school. *International Journal of STEM Education*, *9*(1), 1–47. https://doi.org/10.1186/s40594-022-00363-x
18. Jiraniramai, S., Wongpakaran, T., Angkurawaranon, C., Jiraporncharoen, W., & Wongpakaran, N. (2021). Construct validity and differential item functioning of the phq-9 among health care workers: Rasch analysis approach. *Neuropsychiatric Disease and Treatment*, *17*, 1035–1045. https://doi.org/10.2147/NDT.S271987
19. Kholifah, N., Kurdi, M. S., Nurtanto, M., Mutohhari, F., Fawaid, M., & Subramaniam, T. S. (2023). The role of teacher self-efficacy on the instructional quality in 21st century: A study on vocational teachers, Indonesia. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(2), Article 2. https://doi.org/10.11591/ijere.v12i2.23949
20. Klaar, S., & Wank, A. C. (2022). ECE as an educative and multifaceted practice for growth: To assess and evaluate teaching and learning by documenting children's actions and re-actions. *European Early Childhood Education Research Journal*, *30*(4), 557–571. https://doi.org/10.1080/1350293X.2022.2070649
21. Layne, T., Yli-Piipari, S., & Knox, T. (2021). Physical activity break program to improve elementary students' executive function and mathematics performance. *Education 3-13*, *49*(5), 583–591. https://doi.org/10.1080/03004279.2020.1746820
22. Leung, Y. Y., Thumboo, J., Yeo, S. J., Wylde, V., & Tennant, A. (2022). Pos1113 Construct Validity, Reliability, Responsiveness and Interval Scale Transformation of the Western Ontario and Mcmaster Universities Osteoarthritis Index (Womac) in Patients Undergoing Knee Arthroplasty. *Annals of the Rheumatic Diseases*, *81*(Suppl 1), 885–886. https://doi.org/10.1136/annrheumdis-2022-eular.1005

23. Loda, T., Berner, N., Erschens, R., Nikendei, C., Zipfel, S., & Herrmann-Werner, A. (2022). 'Student tutors go online'—Investigation of cognitive and social congruence in online student tutorials—A longitudinal study. *Medical Education Online*, *27*(1), 1–10. https://doi.org/10.1080/10872981.2022.2100038
24. Lyons, L. K., Dorsch, T. E., Lowe, K., Kaye, M. P., Arnett, J. J., Faherty, A., & Menendez, L. H. (2021). Parents' perceptions of parental involvement in emerging adults' intercollegiate athletic careers: Policy, education, and desired outcomes. *Journal for the Study of Sports and Athletes in Education*, *15*(2), 123–149. https://doi.org/10.1080/19357397.2021.1916303
25. McCarron, R. H., Gracey, F., & Bateman, A. (2021). Detecting mental health problems after paediatric acquired brain injury: A pilot Rasch analysis of the strengths and difficulties questionnaire. *Neuropsychological Rehabilitation*, *31*(7), 1048–1068. https://doi.org/10.1080/09602011.2020.1760111
26. Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). The use of rasch measurement model in English testing. *Cakrawala Pendidikan*, *38*(1), 16–32. https://doi.org/10.21831/cp.v38i1.22750
27. Ning, H. K. (2018). A Rasch Analysis of the Junior Metacognitive Awareness Inventory With Singapore Students. *Measurement and Evaluation in Counseling and Development*, *51*(2), 84–91. https://doi.org/10.1080/07481756.2017.1358061
28. Nurtanto, M., Sudira, P., Sofyan, H., & Jatmoko, D. (2022). The Meaning of Work for Indonesian Professional Teacher in Vocational Education: A Phenomenological Research. *Educational Administration: Theory and Practice*, *28*(3), 42–50. https://doi.org/10.17762/kuey.v28i03.549
29. Nurtanto, M., Sudira, P., Sofyan, H., Kholifah, N., & Triyanto, T. (2022). Professional Identity of Vocational Teachers in the 21st Century in Indonesia. *Journal of Engineering Education Transformations*, *35*(3), 30–36. https://doi.org/10.16920/jeet/2022/v35i3/22085
30. Omara, M., Salzberger, T., Boecker, M., Bekes, K., Steiner, G., Nell-Duxneuner, V., Ritschl, V., Mosor, E., Kloppenburg, M., Sautner, J., Steinecker-Frohnwieser, B., & Stamm, T. (2021). Improving the measurement of oral health-related quality of life: Rasch model of the oral health impact profile-14. *Journal of Dentistry*, *114*, 1–13. https://doi.org/10.1016/j.jdent.2021.103819
31. Ormond, B. M. (2019). The impact of standards-based assessment on knowledge for history education in New Zealand. *Assessment in Education: Principles, Policy and Practice*, *26*(2), 143–165. https://doi.org/10.1080/0969594X.2018.1432564
32. Ozkale, A., & Ozdemir Erdogan, E. (2022). An analysis of the interaction between mathematical literacy and financial literacy in PISA*. *International Journal of Mathematical Education in Science and Technology*, *53*(8), 1983–2003. https://doi.org/10.1080/0020739X.2020.1842526
33. Pakaja, F., & Wafa, M. (2021). Social family, parental involvement and intentions: Predicting the technology acceptance and interest students learning online. *Interactive Learning Environments*, *17*(2), 1–16. https://doi.org/10.1080/10494820.2021.2005105
34. Peabody, M. R., & Wind, S. A. (2019). Exploring the Stability of Differential Item Functioning Across Administrations and Critical Values Using the Rasch Separate Calibration t-test Method. *Measurement*, *17*(2), 78–92. https://doi.org/10.1080/15366367.2018.1533782
35. Pellicciari, L., Piscitelli, D., Caselli, S., & La Porta, F. (2019). A Rasch analysis of the Conley Scale in patients admitted to a general hospital. *Disability and Rehabilitation*, *41*(23), 2807–2816. https://doi.org/10.1080/09638288.2018.1478000
36. Purwanto, M. G., Suhandi, A., Coştu, B., Samsudin, A., & Nurtanto, M. (2020). Static fluid concept inventory (SFCI): A gender gap analysis using rasch model to promote a diagnostic test instrument on students' conception. *International Journal of Advanced Science and Technology*.
37. Quansah, F. (2022). Item and rater variabilities in students' evaluation of teaching in a university in Ghana: Application of Many-Facet Rasch Model. *Heliyon*, *8*(12), 1–12. https://doi.org/10.1016/j.heliyon.2022.e12548
38. Reilly, K. A., & Levintova, E. M. (2021). Student Parents and HIPs: Missing Out on High-Impact Practices. *Journal of Continuing Higher Education*, *21*(3), 1–13. https://doi.org/10.1080/07377363.2021.2001615
39. Repo, J. P., Tukiainen, E. J., Roine, R. P., Sampo, M., Sandelin, H., & Häkkinen, A. H. (2019). Rasch analysis of the Lower Extremity Functional Scale for foot and ankle patients. *Disability and Rehabilitation*, *41*(24), 2965–2971. https://doi.org/10.1080/09638288.2018.1483435
40. Robershaw, K. L., Bradley, K. D., & Waddington, R. J. (2022). Parents' Awareness and Perspectives of School Choice Scale: Psychometric Evidence Using Rasch Modelling. *Journal of School Choice*, *16*(2), 275–305. https://doi.org/10.1080/15582159.2021.2004493
41. Shen, Z., Tan, S., & Siau, K. (2019). Use of mental models and cognitive maps to understand students' learning challenges. *Journal of Education for Business*, *94*(5), 281–289. https://doi.org/10.1080/08832323.2018.1527748
42. Shi, Y., Gugiu, P. C., Crowe, R. P., & Way, D. P. (2019). A Rasch Analysis Validation of the Maslach Burnout Inventory–Student Survey with Preclinical Medical Students. *Teaching and Learning in Medicine*, *31*(2), 154–169. https://doi.org/10.1080/10401334.2018.1523010
43. Shin, J., Bulut, O., & Gierl, M. J. (2020). The Effect of the Most-Attractive-Distractor Location on Multiple-Choice Item Difficulty. *Journal of Experimental Education*, *88*(4), 643–659. https://doi.org/10.1080/00220973.2019.1629577

44. Stelzer, F., Richard's, M. M., Andrés, M. L., Vernucci, S., & Introzzi, I. (2021). Cognitive and maths-specific predictors of fraction conceptual knowledge. *Educational Psychology*, *41*(2), 172–190. https://doi.org/10.1080/01443410.2019.1693508
45. Stolt, M., Kottorp, A., & Suhonen, R. (2022). The use and quality of reporting of Rasch analysis in nursing research: A methodological scoping review. *International Journal of Nursing Studies*, *132*, 1–14. https://doi.org/10.1016/j.ijnurstu.2022.104244
46. Suryadi, B., Hayat, B., & Putra, M. D. K. (2021). The Indonesian version of the Life Orientation Test-Revised (LOT-R): Psychometric properties based on the Rasch model. *Cogent Psychology*, *8*(1), 1869375. https://doi.org/10.1080/23311908.2020.1869375
47. Suryani, D. I., Ramdani, S. D., Vitasari, M., & ... (2021). Pelatihan Pengelolaan Sampah Organik dengan Smart Sensor kepada Office Boy. In *Jurnal Pengabdian ...* (Vol. 6, Issue 3, pp. 911–920).
48. Teng, Y. (2020). The relationship between school climate and students' mathematics achievement gaps in Shanghai China: Evidence from PISA 2012. *Asia Pacific Journal of Education*, *40*(3), 356–372. https://doi.org/10.1080/02188791.2019.1682516
49. Trinidad, J. E. (2020). Understanding student-centred learning in higher education: Students' and teachers' perceptions, challenges, and cognitive gaps. *Journal of Further and Higher Education*, *44*(8), 1013–1023. https://doi.org/10.1080/0309877X.2019.1636214
50. Wallace, G. H. (2020). Improving Spanish classroom assessment via logistic regression: Lessons from the Rasch model. *Journal of Spanish Language Teaching*, *7*(1), 51–63. https://doi.org/10.1080/23247797.2020.1771009
51. Wang, Y., & Stein, D. (2021). Effects of online teaching presence on students' cognitive conflict and engagement. *Distance Education*, *42*(4), 547–566. https://doi.org/10.1080/01587919.2021.1987837
52. Wei, W., & Cheng, L. (2022). Exploring the relationships between teacher-led and learner-led mobile learning activities and their impacts on teacher evaluation results. *Technology, Pedagogy and Education*, *31*(2), 247–259. https://doi.org/10.1080/1475939X.2021.2010591
53. Wiyarsi, A., Fachriyah, A. R., Supriadi, D., & Bin Muhamad Damanhuri, M. I. (2019). A test of analytical thinking and chemical representation ability on 'rate of reaction' topic. *Cakrawala Pendidikan*, *38*(2), 228–242. https://doi.org/10.21831/cp.v38i2.23062
54. Yeganeh, N. M., King, R., Boyd, L. A., Rose, G. M., & Weber, R. C. (2022). Symbol relations training improves cognitive functioning in students with neurodevelopmental disorders. *Applied Neuropsychology: Child*, *11*(4), 789–796. https://doi.org/10.1080/21622965.2021.1967154
55. Zoechling, S., Hopf, M., Woithe, J., & Schmeling, S. (2022). Students' interest in particle physics: Conceptualisation, instrument development, and evaluation using Rasch theory and analysis. *International Journal of Science Education*, *44*(15), 2353–2380. https://doi.org/10.1080/09500693.2022.2122897