Item Characteristics Analysis Using the Rasch Model in the Development of Quantitative Literacy Instruments For Elementary Schools Students

Dameis Surya Anggara[1*], Kadir [2], Herwindo Haribowo[3]

[1]Universitas Negeri Jakarta, Indonesia and Universitas Pamulang, Indonesia

[2]Universitas Islam Negeri Syarif Hidayatullah, Indonesia

[3]Universitas Negeri Jakarta, Indonesia

*Email: dameis_surya@yahoo.com

ABSTRACT

This research is part of the development of quantitative literacy instruments for elementary school students. This study aims to analyze the characteristics of items from the device, namely the validity of the internal structure, the reliability of internal consistency, and the differential item function [DIF]. The instruments developed were 40 multiple-choice items constructed from interpretation, representation, calculation, assumptions, analysis, and communication skills. Furthermore, the device was tested on 480 elementary school students in Semarang City and Demak Regency, Indonesia. The collected data were analyzed using the Rasch Model approach to determine the characteristics of the items. The results showed 30 items out of 40 things that had valid criteria with Outfit ZSTD values between -2.0 and +2.0. The instrument's reliability has the right standards with a Cronbach Alpha (KR-20) value of 0.70, and item reliability has particular criteria with an item reliability value of 0.99. Then, from 30 valid items, free from DIF with the mantel-haenszel probability value for each item is more than 0.05.

Keywords

Quantitative Literacy, Rasch Model, Validity, Reliability, Differential Item Function [DIF].

Article Received: 18 October 2020, Revised: 3 November 2020, Accepted: 24 December 2020

Introduction

Quantitative literacy is a habit of thinking that involves understanding, beliefs, and mathematical dispositions in everyday life (Wilkins, 2010). Sweet and Strand (2006) also outlined quantitative literacy is the ability to understand and process statistical information. Furthermore, Hallet (2003) describes quantitative literacy as the ability to identify, understand, and use quantitative arguments in everyday contexts. Then, Mayes (2014) explained that quantitative literacy is the use of numbers and arithmetic to measure context to understand phenomena so that they can make decisions. So it can be concluded that quantitative literacy is the ability to reason in managing information in numbers and statistics in various everyday contexts.

Quantitative literacy is one part of mathematical ability. This is under Otanrio Education (2014), which explains that quantitative literacy is one of the dimensions of mathematical literacy, apart from spatial literacy and numerical literacy. Furthermore, Lange (2006) explains that mathematical literacy consists of spatial, numerical, and quantitative literacy. Then it is consistent with Ministry of Education Regulation No. 22/2006 described that the purpose of Mathematics in elementary schools includes: understanding mathematical concepts, applying concepts, using reasoning, explaining mathematical ideas and statements, solving problems, and communicating mathematical ideas (Irawan, 2019).

The Association of American Colleges and Universities [AACU] describes that quantitative literacy has six dimensions: 1) interpretation ability is the ability to examine graphs and table information, 2) representation is the ability to conceptualize information in geometric patterns, mathematical models, and charts, (3) calculation is the ability to use arithmetic operations in solving problems, 4) assumptions is the ability to interpret information in the form of a story, 5) analysis is the ability to analyze information based on quantitative data analysis in solving problems, and 6) communication skills is the ability to conceptualize problem-solving in the form of mathematical models. (Steen, 2001).

Quantitative literacy has an essential role in improving society's quality of life, Burdette and McLoughlin (2010) which states that quantitative skills are necessary for higher education, employment, and active and wise citizens. Then Wilkins (2010) describes that quantitative literacy enables a person to function in society. Then Irianto and Febrianti (2017) explained that one must be a literate person to improve the quality of self.

However, the quantitative literacy of Indonesian students has many problems. As the Trend results in International Mathematics and Science Study [TIMSS] in 2015 reported, the mathematical ability of Indonesia's fourth-grade students ranked 45th out of 50 countries with 397 points (TIMSS, 2016). The results of the Program for International Student Assessment [PISA] in 2015 reported that the mathematical abilities of Indonesian students at the age of 15 occupy 63 places in 69 countries. The power of Indonesian students at levels 5 and 6 is only 0.8%, and those at levels below 2 are 42.3% of the total number of Indonesian student participants (OECD, 2018). The evidence above confirms that Indonesian students' quantitative literacy is low, with low mathematical ability indicators.

By following up on these problems, it is necessary to increase quantitative literacy by creating effective mathematics learning. The learning process can run effectively if it considers individual differences between students (Martony, 2019). A quantitative literacy test is needed so that the teacher can know individual differences to create effective learning (Suwarsono, 1987)

A good test must have valid items, be reliable, and be free from DIF, as well as quantitative literacy tests. Cohen and Swerdlik (2010) describe that a good test is a valid and reliable test. Then Sumintono and Widhiarso (2015) explain that validity is one of the concepts determining an instrument's quality. Furthermore, Goodwin and Leech (2003) describe that fact is the most basic consideration in developing and evaluating tests. Besides, Dorans and Holland (1992) explain that DIF refers to differences in items between two groups of comparable examinees.

Schuhmann has researched quantitative literacy et al. (2005); Bookman et al. (2008); Burdette and McLoughlin (2010); and Rafianti et al. (2018) with research focusing on quantitative literacy skills assessment. Furthermore, Wilkin's (2010) research focuses on finding quantitative literacy measurement models with constructs of belief, cognition, and disposition dimensions using Exploratory Factor Analysis and Confirmatory Factor Analysis.

The studies related to quantitative literacy focus on aspects of assessment without analyzing the instruments used. Then also focus on parts of the analysis of the validity of measurement models using factor analysis. So the novelty of this study is to analyze the characteristics of quantitative literacy instrument items using the Rasch Model with different constructs consisting of interpretation, representation, calculation, assumptions, analysis, and communication skills.

Literature Review

Quantitative Literacy

Quantitative literacy is part of mathematical literacy, spatial literacy, and numerical literacy (Lange, 2006; Otanrio, 2014). Quantitative literacy requires understanding and mathematical skills, such as numbers, odds, change, and relationships. Wilkins (2010) quantitative literacy is a habit of thinking that involves cognitive, belief, and attitudes towards mathematics in everyday life. It can be said that quantitative literacy has a functional knowledge of mathematical content, reasoning skills, and a positive attitude towards mathematics. Furthermore, Steen (2001), quantitative literacy is a person's ability to reason and think mathematically to function in collecting quantitative content. Besides, Sweet and Strand (2006), quantitative literacy refers to

understanding and managing statistical information.

Meanwhile, Hallet (2003), quantitative literacy can identify, understand, and use quantitative arguments in everyday contexts. Meanwhile, Mayes (2014) states that quantitative literacy is the use of numbers and arithmetic to measure context to understand phenomena so that they can make decisions. Based on these experts' descriptions, it can be concluded that quantitative literacy is the ability to reason to solve problems in the form of numbers, arithmetic, and statistics from various contexts and everyday life. Individuals with good quantitative literacy skills will be able to reason and solve quantitative problems from multiple contexts and everyday life situations. They can understand, make arguments with quantitative evidence, and communicate these arguments.

Quantitative literacy consists of 4 essential elements: counting, measurement, proportional reasoning, descriptive statistics, and basic probability (Mayes, 2014). The Association of American Colleges and Universities (in Steen, 2001) states that quantitative literacy has six dimensions, namely 1) interpretation ability, 2) representation ability, 3) calculation ability, 4) assumption ability, 5) analysis ability and 6) communication skills. The descriptions of the six indicators are presented in the following table 1.

Table 1: Dimensions of Quantitative Literacy

No	Dimensions	Description
1	Interpretation	Ability to review the information presented in graphs and tables
2	Representation	Ability to conceptualize information into geometric patterns, mathematical models, and graphs
3	Calculation	Ability to use arithmetic operations to solve problems
4	Analysis	Ability to analyze information based on quantitative data analysis in solving problems.
5	Assumption	Ability to interpret problem-solving results from information in the form of story problems.
6	Communication	Ability to conceptualize problem-solving in the form of a mathematical model

Validity

Validity means measuring what you want to measure (Field, 2005). Wijanto (2008) states that reality is related to whether variable measures what is being measured. Furthermore, Azwar (1987) says that validity means the extent to which a measuring instrument's accuracy and accuracy (test) in performing its measuring function. Besides, Cohen and Swerdlik (2010) argue that test validity is an assessment or estimate of how well a test measures what is to be measured in a particular context.

The internal structure's validity is the accuracy of the item in measuring the defined measuring construct through theory. The concept of internal structure validity is the same as construct validity. That is, they both know the validity based on the measuring construct. This is reinforced by Goodwin and Leech (2003). They state that fact based on the internal structure is the accuracy of a test's internal components with a measuring construct, one of which can be analyzed by Confirmatory Factor Analysis. Then, Mardapi (2008) states that construct validity refers to the extent to which the test measures the concept of a theory, which is the basis for preparing the trial, which can be proven through the test's internal structure. The statistical analysis

used to determine this validity is the model item accuracy index specified by the infit coefficient and Outfit in Rasch modeling (Sumintono & Widhiarso, 2015).

Reliability

Crocker and Algina (2008) argue that reliability is the consistency of scores produced by tests. Mardapi (2008) states that the reliability principle shows that the measurement results are relatively the same if the measurement is repeated on the same object. According to Sumintono and Widhiarso (2015), reliability explains how far a measure made multiple times will produce the same information. Naga (2013) states that data reliability is the level of data confidence. Based on the expert opinion above, it can be concluded that reliability is the level of consistency/stability of the measurement data.

The method of estimating reliability in this study uses the internal consistency estimation method. This method is a way of knowing the consistency of an item by measuring one instrument, then calculating the reliability coefficient as a determinant of the reliability criteria. The reliability coefficient used is the KR 20 reliability coefficient. According to Naga (2013), Kuder Richardson's reliability coefficient was stated in 1937, which applies to dichotomous data. The general criteria for determining the reliability of data based on the KR reliability coefficient are presented in Table 2.

Table 2: General Criteria KR Reliability Coefficient

KR Reliability Coefficient	Criteria Reliability	Information
$\alpha < 0,5$	Not Acceptable	Not Reliable
$0,5 \leq \alpha < 0,6$	Less	Not Reliable
$0,6 \leq \alpha < 0,7$	Problematic	Not Reliable
$0,7 \leq \alpha < 0,8$	Acceptable	Reliable
$0,8 \leq \alpha < 0,9$	Good	Reliable
$0,9 \leq \alpha$	Very Good	Reliable

(Naga, 2013: 241)

Differensial Item Fungtion (DIF)

Dorans and Holland (1992) stated that DIF refers to differences in item performance between two comparable groups of examinees. DIF detection in this study refers to two different groups, namely the male and female genders. Suppose the probability value of the haenszel coat is more than 0.05. In that case, the item is not detected by DIF, which means that the item's performance for the person with the same ability who comes from two different groups (male and female sex) is the same.

Methods

Quantitative literacy instruments were developed as many as 40 multiple-choice items constructed from interpretation, representation, calculations, assumptions, analysis, and communication skills. Three mathematicians and 3 measurement experts then assessed the things to determine the validity of their contents. After the instrument was valid in range, it was tested on 480 elementary school students in Semarang City and Demak Regency, Indonesia. The collected data were analyzed using the Rasch Model approach with assumptions of fit items, fit person, unidimensional items, local independence, group invariance.

Data Analysis

The data analysis in this study is the Rasch Model Analysis. The Rasch model is a mathematical formula related to the probability of the outcome when one person answers an item

with the person's characteristics and item (McArthur, 1987). In full, the data analysis is presented in Table 3

Table 3: Data Analysis Techniques

No	Aspect of Analysis	Analytical Approach	Criteria
1	**Assumption of the Rasch Model**		
	Fit items and fit person	Rasch Model	Fit items and fit persons if -2.0 <ZSTD <+2.0 (Boone, et al, 2014)
	Unidimensional	Rasch Model	Unidimensional assumptions are met if the value of the test variant ≥ 20% (Reckase, 1979)
	Local Independence	Rasch Model	Assumptions of local independence are met if the residual correlation value ≤ 0.20 (Christensen et al., 2016).
	Group Invariance	Rasch Model	Examination of group invariance can be guided by an increase invalid scores along with ability levels (Kang et al. 2018)
2	**Characteristics of items**		
	The validity of Internal Structures	Rasch Model	items have valid internal structure criteria, if the outfit value is ZSTD: -2.0 <ZSTD <+2.0 (Boone, et al, 2014).
	Item and Instrument Reliability	Rasch Model	The instrument would have reliable criteria if the reliability coefficient KR-20 ≥ 0.70 (Naga, 2013) Items have reliable criteria, if the value of Item Reliability ≥ 0.67 (Fisher, 2007)
	Differential Item Function [DIF].	Rasch Model	The items are free from DIF if the Mantel-Haenszel probability is ≥ 0.05

Results

Before analyzing the item characteristics of quantitative literacy instruments using the Rasch Model, it must meet the Rasch Model's assumptions. The assumptions are 1) item fit and person fit, 2) unidimensional, 3) local independence, and 4) group invariance. The first assumption is checking the item fit and person fit. The purpose of this assumption is to determine the appropriateness/suitability of items and persons with the Rasch Model. Initially, the number of items was 40 items, and the number of people was 480. After analysis, 10 misfit items and 65 misfit people with ZSTD values outside the -2.0 <ZSTD <+2.0 criteria. Misfit items and person misfits are presented in Table 4.

Table 4: *Misfit* Item dan Person

Misfit	Total	Number
Item	10 Item	35, 28, 25, 22, 27, 39, 21, 2, 31, 9
Person	65 Person	216, 115, 179, 124, 169, 426, 110, 23, 20, 241, 287, 400, 371, 202, 118, 164, 105, 452, 163, 58, 455, 442, 347, 136, 257, 162, 327, 8, 270, 187, 247, 186, 14, 391, 325, 242, 249, 125, 213, 19, 324, 148, 181, 106, 201, 250, 129, 354, 245, 121, 271, 67, 350, 474, 104, 68, 368, 305, 383, 428, 256, 31, 183, 217, 57

Then the misfit items and the misfit person are eliminated from the data analysis. The remaining items and person are analyzed again until they show the fit items and person fit results.

The fit items' works are presented in table 3, and the relevant person results are presented in Table 5.

Table 5: Fit Item Results

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PT-MEASURE		EXACT OBS%	MATCH EXP%	ITEM	G
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
36	149	414	.22	.11	1.08	1.8	1.10	1.7	A .24	.34	67.4	69.0	ITEM36	0
24	160	415	.10	.11	1.07	1.8	1.09	1.7	B .25	.34	64.1	67.4	ITEM24	0
8	77	415	1.25	.13	1.09	1.0	1.08	.7	C .23	.32	80.2	82.6	ITEM8	0
17	150	414	.21	.11	1.04	1.0	1.07	1.2	D .28	.34	67.9	68.8	ITEM17	0
38	160	411	.08	.11	1.05	1.3	1.06	1.2	E .27	.34	63.7	67.1	ITEM38	0
33	99	415	.89	.12	1.03	.5	1.06	.7	F .29	.33	77.3	78.1	ITEM33	0
13	118	415	.62	.12	1.05	1.0	1.05	.7	G .28	.34	71.6	74.5	ITEM13	0
11	37	413	2.18	.18	1.05	.4	1.01	.1	H .24	.28	90.8	91.3	ITEM11	0
23	95	412	.94	.12	1.04	.6	1.01	.2	I .30	.33	76.7	78.8	ITEM23	0
26	83	415	1.14	.13	1.03	.4	.98	-.1	J .31	.33	80.2	81.3	ITEM26	0
4	152	414	.19	.11	1.02	.5	1.03	.6	K .31	.34	67.9	68.5	ITEM4	0
29	164	414	.05	.11	1.03	.8	1.02	.4	L .30	.34	63.8	66.8	ITEM29	0
30	226	409	-.67	.11	1.03	.9	1.00	.0	M .29	.31	58.9	63.2	ITEM30	0
18	70	415	1.38	.14	1.01	.2	1.00	.0	N .31	.32	83.6	84.0	ITEM18	0
40	153	413	.17	.11	.99	-.2	1.00	.0	O .35	.34	69.0	68.3	ITEM40	0
6	137	415	.37	.11	.99	-.2	.99	-.1	o .35	.34	72.3	71.0	ITEM6	0
32	262	414	-1.03	.11	.99	-.2	.95	-.8	n .32	.30	65.9	66.2	ITEM32	0
34	166	413	.02	.11	.98	-.6	.99	-.2	m .36	.34	67.1	66.5	ITEM34	0

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PT-MEASURE		EXACT OBS%	MATCH EXP%	ITEM	G
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
14	128	415	.49	.11	.99	-.2	.97	-.5	l .36	.34	73.7	72.6	ITEM14	0
16	197	415	-.31	.10	.98	-.5	.96	-.9	k .35	.33	64.6	63.6	ITEM16	0
20	354	415	-2.34	.14	.97	-.3	.87	-.9	j .26	.21	85.5	85.3	ITEM20	0
7	286	413	-1.33	.11	.97	-.7	.92	-1.0	i .33	.28	69.2	70.3	ITEM7	0
37	178	413	-.11	.11	.96	-1.0	.95	-1.1	h .38	.33	68.5	65.1	ITEM37	0
1	103	414	.82	.12	.96	-.6	.95	-.5	g .38	.34	77.8	77.3	ITEM1	0
19	193	414	-.28	.10	.96	-1.3	.94	-1.3	f .38	.33	65.0	63.8	ITEM19	0
3	180	414	-.13	.11	.96	-1.3	.93	-1.5	e .39	.33	67.9	64.9	ITEM3	0
15	107	415	.77	.12	.94	-.9	.95	-.6	d .39	.34	79.3	76.5	ITEM15	0
10	405	413	-4.57	.36	.95	.0	.40	-1.7	c .22	.08	98.1	98.1	ITEM10	0
5	235	415	-.72	.10	.95	-1.7	.90	-1.8	b .38	.31	64.1	63.4	ITEM5	0
12	202	413	-.37	.10	.95	-1.8	.92	-1.7	a .40	.33	65.9	63.3	ITEM12	0
MEAN	167.5	413.8	.00	.12	1.00	.0	.97	-.2			72.3	72.6		
S.D.	79.8	1.4	1.21	.05	.04	1.0	.12	1.0			8.9	8.8		

Table 6: Fit Person Result

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PT-MEASURE		EXACT OBS%	MATCH EXP%	PERSON
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.			
34	7	30	-1.37	.49	1.53	1.7	1.95	1.9	A .07	.45	70.0	81.3	34 L
326	7	30	-1.37	.49	1.45	1.5	1.86	1.8	B .10	.45	76.7	81.3	326 P

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PT-MEASURE		EXACT OBS %	MATCH EXP %	PERSON
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.			
48	6	30	-1.62	.52	1.43	1.3	1.86	1.6	C .12	.46	80.0	84.2	48 L
255	7	30	-1.37	.49	1.14	.6	1.82	1.7	D .29	.45	76.7	81.3	255 L
418	5	30	-1.91	.56	1.69	1.6	1.76	1.3	E .03	.47	76.7	86.9	418 L
333	16	30	.25	.40	1.35	2.6	1.71	1.6	F .06	.37	50.0	65.5	333 L
18	7	30	-1.37	.49	1.30	1.1	1.67	1.4	G .23	.45	70.0	81.3	18 L
399	8	30	-1.14	.46	1.05	.3	1.61	1.5	H .36	.44	76.7	78.6	399 L
154	12	30	-.39	.41	1.30	1.8	1.61	1.6	I .14	.41	60.0	69.2	154 P
75	8	30	-1.14	.46	1.44	1.7	1.58	1.4	J .11	.44	70.0	78.6	75 L
349	6	30	-1.62	.52	1.09	.4	1.56	1.2	K .34	.46	86.7	84.2	349 L
16	9	30	-.94	.44	1.03	.2	1.55	1.4	L .35	.43	83.3	76.0	16 P
13	9	30	-.94	.44	1.06	.4	1.53	1.4	M .34	.43	76.7	76.0	13 L
192	7	30	-1.37	.49	1.11	.5	1.53	1.2	N .34	.45	76.7	81.3	192 L
413	5	30	-1.91	.56	1.33	.9	1.53	1.0	O .23	.47	83.3	86.9	413 L
21	13	30	-.23	.40	1.27	1.8	1.51	1.4	P .16	.40	60.0	67.5	21 L
43	11	30	-.57	.42	1.46	2.4	1.46	1.3	Q .08	.42	50.0	71.3	43 L
191	14	30	-.07	.40	1.27	1.9	1.46	1.2	R .16	.39	70.0	66.2	191 P
172	9	30	-.94	.44	1.33	1.4	1.42	1.2	S .19	.43	70.0	76.0	172 P
54	13	30	-.23	.40	1.27	1.9	1.42	1.2	T .18	.40	53.3	67.5	54 P
36	11	30	-.57	.42	1.33	1.8	1.42	1.2	U .16	.42	56.7	71.3	36 P
127	9	30	-.94	.44	1.39	1.7	1.31	.9	V .17	.43	63.3	76.0	127 L
438	7	30	-1.37	.49	1.33	1.1	1.39	1.0	W .22	.45	76.7	81.3	438 L
361	12	30	-.39	.41	1.39	2.3	1.38	1.1	X .13	.41	46.7	69.2	361 L
373	12	29	-.29	.42	1.16	1.1	1.37	1.1	Y .25	.40	69.0	68.4	373 L

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PT-MEASURE		EXACT OBS %	MATCH EXP %	PERSON
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.			
441	5	30	-1.91	.56	1.37	1.0	1.34	.7	Z .25	.47	83.3	86.9	441 P
120	16	30	.25	.40	1.33	2.5	1.32	.9	.12	.37	50.0	65.5	120 L
88	14	30	-.07	.40	1.28	2.0	1.24	.7	.19	.39	50.0	66.2	88 P
88	14	30	-.07	.40	1.28	2.0	1.24	.7	.19	.39	50.0	66.2	88 P
45	7	30	-1.37	.49	.82	-.6	.70	-.6	z .58	.45	83.3	81.3	45 L
304	15	30	.09	.40	.82	-1.5	.73	-.6	y .52	.38	70.0	65.3	304 L
342	26	30	2.23	.56	.82	-.4	.60	-.5	x .38	.23	86.7	86.7	342 P
93	9	30	-.94	.44	.82	-.8	.71	-.8	w .57	.43	83.3	76.0	93 P
140	23	30	1.48	.46	.81	-.8	.66	-.4	v .45	.29	80.0	77.8	140 L
477	13	30	-.23	.40	.80	-1.5	.71	-.8	u .56	.40	73.3	67.5	477 P
292	10	30	-.75	.43	.80	-1.1	.78	-.6	t .56	.42	80.0	73.5	292 P
46	9	30	-.94	.44	.80	-.9	.67	-.9	s .59	.43	83.3	76.0	46 P
224	16	30	.25	.40	.80	-1.7	.71	-.6	r .53	.37	76.7	65.5	224 L
445	21	30	1.09	.43	.79	-1.2	.70	-.3	q .48	.32	83.3	72.5	445 P
233	13	30	-.23	.40	.79	-1.6	.72	-.8	p .56	.40	86.7	67.5	233 L
290	29	30	3.82	1.03	.79	.1	.20	-.5	o .34	.13	96.7	96.7	290 L
134	10	30	-.75	.43	.78	-1.2	.66	-1.0	n .60	.42	80.0	73.5	134 P
303	10	30	-.75	.43	.78	-1.2	.66	-1.0	m .60	.42	80.0	73.5	303 L
111	10	30	-.75	.43	.77	-1.2	.71	-.8	l .59	.42	80.0	73.5	111 P
457	16	30	.25	.40	.77	-2.0	.69	-.7	k .55	.37	76.7	65.5	457 P
133	8	30	-1.14	.46	.76	-1.0	.62	-1.0	j .62	.44	83.3	78.6	133 L
155	11	30	-.57	.42	.76	-1.5	.67	-1.0	i .60	.42	83.3	71.3	155 P
295	24	30	1.70	.48	.75	-.9	.55	-.7	h .47	.27	83.3	80.8	295 P
12	8	30	-1.14	.46	.75	-1.0	.62	-1.0	g .63	.44	83.3	78.6	12 L
274	10	30	-.75	.43	.75	-1.4	.63	-1.1	f .62	.42	80.0	73.5	274 L
432	7	30	-1.37	.49	.74	-.9	.59	-1.0	e .64	.45	83.3	81.3	432 P
83	6	30	-1.62	.52	.74	-.8	.58	-.8	d .65	.46	86.7	84.2	83 L
51	19	30	.74	.41	.73	-2.0	.64	-.6	c .55	.34	83.3	68.7	51 P
130	15	30	.09	.40	.71	-2.5	.64	-.9	b .60	.38	83.3	65.3	130 P
152	9	30	-.94	.44	.70	-1.5	.58	-1.3	a .66	.43	83.3	76.0	152 P
MEAN	12.1	29.9	-.40	.43	1.00	.0	.97	.0			72.3	72.6	
S.D.	4.3	.4	.79	.05	.15	.8	.24	.6			8.6	5.9	

Based on table 3, it can be explained that the number of items originally was 40 items. Ten things were eliminated from the data analysis, so there were still 30 items. All remaining items have a ZSTD Outfit value of -2.0 to +2.0. Furthermore, based on table 4, it can be explained that the initial number of people was 480, then 65 people were eliminated so that there were still 415 people. All remaining persons have a ZSTD Outfit value of -2.0 to +2.0. So it can be concluded that the 30

items and 415 people fit the Rasch model and qualify to be analyzed using the Rasch Model. In summary, the results of the fit item and fit person analysis are presented in table 7.

Table 7: Results of the Item Fit and Person Fit Tests

Category	Total	Misfit Total	Fit Total
Item	40 item	10 item	30 item
Person	480 person	65 person	415 person

The second assumption is the Unidimensional Item. This assumption explains that each item must measure just one ability. The unidimensional results of the items in this study are presented in table 8.

Table 8: Item Unidimensional Results

Table of RAW RESIDUAL Variance (in Eigenvalue units)

		-- Empirical --		Modeled
Total raw Variance in observations	= 39.8	100.0%		100.0 %
Raw Variance explained by measures	= 9.8	24.6 %		21.9 %
Raw Variance explained by persons	= 2.6	6.6 %		5.9%
Raw Variance explained by items	= 7.1	17.9 %		16.0 %
Raw unexplained Variance (total)	= 30.0	75.4 %	100.0%	78.1 %
Unexplained Variance in 1st contrast	= 1.7	4.2 %	5.6 %	
Unexplained Variance in 2nd contrast	= 1.6	4.0 %	5.3 %	
Unexplained Variance in 3rd contrast	= 1.4	3.6 %	4.8 %	
Unexplained Variance in 4th contrast	= 1.4	3.5 %	4.7 %	
Unexplained Variance in 5th contrast	= 1.3	3.3 %	4.4 %	

Based on table 6, it can be explained that the value of the Raw Variance Explained by Empirical Measures is 24.6%, exceeding 20%, so the Unidimensional item's assumptions are met. So it can be concluded that things only measure quantitative literacy and meet the requirements to be analyzed using the Rasch Model.

The third assumption is the local independence. This assumption explains that the answers generated between test takers and between items do not affect each other. The results of the regional autonomy of the things in this study are presented in Tables 9, and the results of the person's local independence shown in Table 10.

Table 9: Local Independence Results of Items

LARGEST OBSERVATION RESIDUAL CORRELATIONS USED TO IDENTIFY DEPENDENT ITEM				
CORREL-ATION	**ENTRY NUMBER**	**ITEM**	**ENTRY NUMBER**	**ITEM**
-.17	5	ITEM5	24	ITEM 24
-.17	17	ITEM17	20	ITEM 20
-.15	8	ITEM8	32	ITEM 32
-.15	1	ITEM1	18	ITEM 18
-.15	18	ITEM18	32	ITEM 32
-.14	6	ITEM6	32	ITEM 32
-.14	11	ITEM11	36	ITEM 36
-.14	14	ITEM14	18	ITEM 18
-.13	5	ITEM5	17	ITEM 17
-.13	13	ITEM13	24	ITEM 24

Table 10: Local Independence Results of Persons

LARGEST OBSERVATION RESIDUAL CORRELATIONS USED TO IDENTIFY DEPENDENT PERSON				
CORREL-ATION	**ENTRY NUMBER**	**PERSON**	**ENTRY NUMBER**	**PERSON**
-.73	243	243 L	387	387 P
-.67	103	103	448	448 P

		L		
-.66	102	102	243	243 L
		L		
-.66	204	204 P	315	315 P
-.65	27	27 P	206	206 P
-.64	120	120	446	446 L
		L		
-.63	285	285	425	425 P
		L		
-.63	215	215	339	339 L
		L		
-.63	55	55 L	379	379 L
-.63	77	77 L	462	462 P

Based on table 9, it can be explained that all residual correlations have a value of less than 0.20, so the assumption of independence between items is fulfilled. Furthermore, based on table 10, it can be explained that all the values of residual correlations have a value of less than 0.20, so the assumption of independence between persons is fulfilled. So it can be concluded that the answers between items and between-person do not influence each other and qualify to be analyzed using the Rasch Model.

The final assumption is group invariance. This assumption explains that all groups of person abilities (high - low) invariably refer to the item's characteristic curve. The results of group invariance in this study are presented in table 11.

Table 11: Results of Group Invariance

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE MEASURE	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	ITEM
11		***	2	0#	-.39	.34		.00	ITEM11
	0	0	376	91	-.46	.04	1.1	-.24	
	1	1	37	9	.20	.16	1.0	.24	
18	0	0	345	83	-.51	.04	1.0	-.31	ITEM18
	1	1	70	17	.15	.12	1.0	.31	
8	0	0	338	81	-.49	.04	1.1	-.23	ITEM8
	1	1	77	19	-.01	.10	1.1	.23	
26	0	0	332	80	-.52	.04	1.0	-.31	ITEM26
	1	1	83	20	.09	.10	1.0	.31	
23		***	3	1#	-.41	.26		.00	ITEM23
	0	0	317	77	-.53	.04	1.0	-.30	
	1	1	95	23	.03	.09	1.0	.30	
33	0	0	316	76	-.53	.04	1.0	-.29	ITEM33
	1	1	99	24	.02	.10	1.1	.29	
1		***	1	0#	-.15			.02	ITEM1
	0	0	311	75	-.57	.04	.9	-.38	
	1	1	103	25	.12	.10	1.0	.38	
15	0	0	308	74	-.58	.03	.9	-.39	ITEM15
	1	1	107	6	.13	.10	1.0	.39	
13	0	0	297	72	-.54	.04	1.1	-.28	ITEM13
	1	1	118	28	-.05	.08	1.0	.28	

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE MEASURE	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	ITEM
14	0	0	287	69	-.59	.04	1.0	-.36	ITEM14
	1	1	128	31	.03	.08	1.0	.36	
6	0	0	278	67	-.59	.04	.9	-.35	ITEM6
	1	1	137	33	.00	.08	1.0	.35	
36		***	1	0#	.34			.05	ITEM36
	0	0	265	64	-.54	.04	1.1	-.24	
	1	1	149	36	-.15	.08	1.1	.24	
17		***	1	0#	-.33			.00	ITEM17
	0	0	264	64	-.57	.04	1.0	-.28	
	1	1	150	36	-.10	.08	1.1	.28	
4		***	1	0#	-.16			.01	ITEM4
	0	0	262	63	-.59	.04	1.1	-.31	
	1	1	152	37	-.08	.07	1.0	.31	
40		***	2	0#	-.42	.34		.00	ITEM40
	0	0	260	63	-.61	.04	1.0	-.35	
	1	1	153	37	-.04	.07	1.0	.35	
24	0	0	255	61	-.56	.04	1.1	-.25	ITEM24
	1	1	160	39	-.15	.07	1.1	.25	
38		***	4	1#	.07	.47		.06	ITEM38
	0	0	251	61	-.58	.04	1.0	-.27	
	1	1	160	39	-.13	.07	1.1	.27	
29		***	1	0#	-.15			.02	ITEM29
	0	0	250	60	-.60	0.4	1.0	-.30	
	1	1	164	40	-.10	0.7	1.0	.30	
29		***	2	0#	-1.00	.32		-.05	ITEM34
	0	0	247	60	-.63	.04	1.0	-.36	
	1	1	166	40	-.05	.07	1.0	.36	
37		***	2	0#	-.37	.39		.00	ITEM37
	0	0	235	57	-.66	.04	.9	-.38	
	1	1	178	43	-.05	.07	1.0	.38	
3		***	1	0#	-1.08			-.04	ITEM3
	0	0	234	57	-.67	.04	.9	-.39	
	1	1	180	43	-.04	.06	.9	.39	
19		***	1	0#	.31			.04	ITEM19
	0	0	221	53	-.68	.04	.9	-.38	
	1	1	193	47	.08	.06	1.0	.38	
24	0	0	218	53	-.67	.04	.9	-.35	ITEM16
	1	1	197	47	-.10	.06	1.0	.35	
12		***	2	0#	-.97	.10		-.05	ITEM12
	0	0	211	51	-.70	.04	.9	-.40	
	1	1	202	49	-.08	.06	.9	.40	
30		***	6	1#	.01	.34		.06	ITEM30
	0	0	183	45	-.66	.04	.9	-.29	

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE MEASURE	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	ITEM
	1	1	226	55	-.20	.06	1.1	.29	
5	0	0	180	43	-.75	.04	.9	-.38	ITEM5
	1	1	235	57	-.13	.06	1.0	.38	
32		***	1	0#	-1.04			-.04	ITEM32
	0	0	152	37	-.73	.05	.9	-.32	
	1	1	262	63	-.21	.05	1.0	.32	
7		***	2	0#	-.07	.08		.03	ITEM7
	0	0	127	31	-.79	.05	.9	-.33	
	1	1	286	69	-.23	.05	1.0	.33	
20	0	0	61	15	-.90	.07	.8	-.26	ITEM20
	1	1	354	18	-.31	.04	1.0	.26	
10		***	2	0#	-.48	.09		-.01	ITEM10
	0	0	8	2	-1.64	.12	.4	-.22	
	1	1	405	98	-.37	.04	1.0	.22	

Based on table 11, it can be explained that the average measure for each item always increases, meaning that it meets the group invariance. So it can be concluded that the characteristics of things are still fixed and do not depend on the ability of test-takers and qualify to be analyzed using the Rasch Model.

After all, assumptions are met; the next step is to analyze quantitative literacy items' characteristics using the Rasch Model approach. The aspects of the things to be interpreted are: 1) the validity of the internal structure of the items, 2) the reliability of objects and instruments, and 3) DIF

The validity of Internal Structures

The internal structure's validity is the accuracy of the items in measuring the defined constructs through theory. The results of the fact of the internal system are presented in table 12.

Table 12: Results of Internal Structure Validity

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT OBS %	MATCH EXP %	ITEM	G
36	149	414	.22	.11	1.08	1.8	1.10	1.7	A .24	.34	67.4	69.0	ITEM36	0
24	160	415	.10	.11	1.07	1.8	1.09	1.7	B .25	.34	64.1	67.4	ITEM24	0
8	77	415	1.25	.13	1.09	1.0	1.08	.7	C .23	.32	80.2	82.6	ITEM8	0
17	150	414	.21	.11	1.04	1.0	1.07	1.2	D .28	.34	67.9	68.8	ITEM17	0
38	160	411	.08	.11	1.05	1.3	1.06	1.2	E .27	.34	63.7	67.1	ITEM38	0
33	99	415	.89	.12	1.03	.5	1.06	.7	F	.33	77.3	78.1	ITE	0

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PT-MEASURE		EXACT OBS%	MATCH EXP%	ITEM	G
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
									.29				M33	
13	118	415	.62	.12	1.05	1.0	1.05	.7	G .28	.34	71.6	74.5	ITEM13	0
11	37	413	2.18	.18	1.05	.4	1.01	.1	H .24	.28	90.8	91.3	ITEM11	0
23	95	412	.94	.12	1.04	.6	1.01	.2	I .30	.33	76.7	78.8	ITEM23	0
26	83	415	1.14	.13	1.03	.4	.98	-.1	J .31	.33	80.2	81.3	ITEM26	0
4	152	414	.19	.11	1.02	.5	1.03	.6	K .31	.34	67.9	68.5	ITEM4	0
29	164	414	.05	.11	1.03	.8	1.02	.4	L .30	.34	63.8	66.8	ITEM29	0
30	226	409	-.67	.11	1.03	.9	1.00	.0	M .29	.31	58.9	63.2	ITEM30	0
18	70	415	1.38	.14	1.01	.2	1.00	.0	N .31	.32	83.6	84.0	ITEM18	0
40	153	413	.17	.11	.99	-.2	1.00	.0	O .35	.34	69.0	68.3	ITEM40	0
6	137	415	.37	.11	.99	-.2	.99	-.1	o .35	.34	72.3	71.0	ITEM6	0
32	262	414	-1.03	.11	.99	-.2	.95	-.8	n .32	.30	65.9	66.2	ITEM32	0
34	166	413	.02	.11	.98	-.6	.99	-.2	m .36	.34	67.1	66.5	ITEM34	0
14	128	415	.49	.11	.99	-.2	.97	-.5	l .36	.34	73.7	72.6	ITEM14	0
16	197	415	-.31	.10	.98	-.5	.96	-.9	k .35	.33	64.6	63.6	ITEM16	0
20	354	415	-2.34	.14	.97	-.3	.87	-.9	j .26	.21	85.5	85.3	ITEM20	0
7	286	413	-1.33	.11	.97	-.7	.92	-1.0	i .33	.28	69.2	70.3	ITEM7	0
37	178	413	-.11	.11	.96	-1.0	.95	-1.1	h .38	.33	68.5	65.1	ITEM37	0
1	103	414	.82	.12	.96	-.6	.95	-.5	g .38	.34	77.8	77.3	ITEM1	0
19	193	414	-.28	.10	.96	-1.3	.94	-1.3	f .38	.33	65.0	63.8	ITEM19	0
3	180	414	-.13	.11	.96	-1.3	.93	-1.5	e .39	.33	67.9	64.9	ITEM3	0

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PT-MEASURE		EXACT OBS %	MATCH EXP %	ITEM	G
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
15	107	415	.77	.12	.94	-.9	.95	-.6	d .39	.34	79.3	76.5	ITEM15	0
10	405	413	-4.57	.36	.95	.0	.40	-1.7	c .22	.08	98.1	98.1	ITEM10	0
5	235	415	-.72	.10	.95	-1.7	.90	-1.8	b .38	.31	64.1	63.4	ITEM5	0
12	202	413	-.37	.10	.95	-1.8	.92	-1.7	a .40	.33	65.9	63.3	ITEM12	0
MEAN	167.5	413.8	.00	.12	1.00	.0	.97	-.2			72.3	72.6		
S.D.	79.8	1.4	1.21	.05	.04	1.0	.12	1.0			8.9	8.8		

Based on table 12, it can be explained that 30 items have a ZSTD Outfit value between -2.0 to +2.0. So it can be concluded that the 30 items have valid criteria.

Item and Instrument Reliability

Reliability is the level of consistency/stability of data from the measurement. The results of item and instrument reliability are presented in table 13.

Table 13: Item and Instrument Reliability Results

SUMMARY OF 415 MEASURED PERSON

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	12.1	29.9	-.40	.43	1.00	.0	.97	.0
S.D.	4.3	.4	.79	.05	.15	.8	.24	.6
MAX.	29.0	30.0	3.82	1.03	1.69	2.6	1.95	1.9
MIN.	3.0	27.0	-2.71	.40	.70	-2.5	.20	-1.3

REAL RMSE .45 TRUE SD .65 SEPARATION 1.44 PERSON RELIABILITY .67
MODEL RMSE .44 TRUE SD .66 SEPARATION 1.51 PERSON RELIABILITY .69
S.E. OF PERSON MEAN = .04

DELETED: 65 PERSON
VALID RESPONSES: 99.7%
PERSON RAW SCORE-TO-MEASURE CORRELATION = .99 (approximate due to missing data)
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .70 (approximate due to missing data)

SUMMARY OF 30 MEASURED ITEM

	RAW	COUNT	MEASURE	MODEL	INFIT	OUTFIT

	SCORE			ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	167.5	413.8	.00	.12	1.00	.0	.97	-.2
S.D.	79.8	1.4	1.21	.05	.04	1.0	.12	1.0
MAX.	405.0	415.0	2.18	.36	1.09	1.8	1.10	1.7
MIN.	37.0	409.0	-4.57	.10	.94	-1.8	.40	-1.8

REAL RMSE .13 TRUE SD 1.20 SEPARATION 8.98 ITEM RELIABILITY .99
MODEL RMSE .13 TRUE SD 1.20 SEPARATION 9.06 ITEM RELIABILITY .99
S.E. OF ITEM MEAN = .22

DELETED: 10 ITEM

Based on table 13, it can be explained that the Cronbach Alpha (KR-20) value is 0.70, so the reliability of the instrument has good criteria, which means the interaction between 415 people and 30 items as a whole has good criteria. The value of the item reliability is 0.99 with special criteria, which means consistency of answers on 30 items with special criteria. The grouping of items can be determined through formulas H = $\frac{(4\,x\,separation)+1}{3}$ by looking at the separation value of 8.98. Here are the results and the calculation process: H = $\frac{(4\,x\,8{,}98)+1}{3} = \frac{(35{,}92)+1}{3} = \frac{36{,}92}{3}$ = 12,3 or 12 groups of item.

Differential Item Function [DIF]

DIF is the similarity of opportunities for correct answers on items by respondents from different groups such as race, ethnicity, and Gender. The DIF in this study was analyzed by Gender, with the results presented in table 14.

Table 14: DIF Results by Gender

PERSON CLASS	DIF MEASURE	DIF S.E	PERSON CLASS	DIF MEASURE	DIF S.E	DIF CONTRAST	JOINT S.E	Welch t	Welch d.f.	Welch Prob.	MantelHanzl Prob.	MantelHanzl Size	ITEM Number	Name
L	.77	.17	P	.87	.17	-.11	.24	-.44	410	.65087	.83 27	.06	1	ITEM1
P	.87	.17	L	.77	.17	.11	.24	.44	410	.65087	.83 27	-.06	1	ITEM1
L	-.19	.15	P	-.07	.15	-.12	.21	-.58	410	.56034	.26 63	-.17	3	ITEM3
P	-.07	.15	L	-.19	.15	.12	.21	.58	410	.56034	.26 63	.17	3	ITEM3
L	.02	.15	P	.34	.15	-.32	.22	-1.46	411	.14150	.09 04	-.40	4	ITEM4
P	.34	.15	L	.02	.15	.32	.22	1.46	411	.14150	.09 04	.40	4	ITEM4
L	-.64	.15	P	-.81	.15	.17	.21	.82	412	.41251	.26 22	.17	5	ITEM5
P	-.81	.15	L	-.64	.15	-.17	.21	-.82	412	.41251	.26 22	-.17	5	ITEM5

PERSON CLASS	DIF MEASURE	DIF S.E	PERSON CLASS	DIF MEASURE	DIF S.E	DIF CONTRAST	JOINT S.E	Welch t	Welch d.f.	Welch Prob.	MantelHanzl Prob.	MantelHanzl Size	ITEM Number	Name
L	.29	.16	P	.44	.15	-.15	.22	-.67	411	.50 26	.38 37	-.10	6	ITEM6
P	.44	.15	L	.29	.16	.15	.22	.67	411	.50 26	.38 37	.10	6	ITEM6
L	-1.17	.16	P	-1.50	.16	.33	.22	1.47	410	.14 14	.15 56	.39	7	ITEM7
P	-1.50	.16	L	-1.17	.16	-.33	.22	-1.47	410	.14 14	.15 56	-.39	7	ITEM7
L	1.25	.20	P	1.25	.18	.00	.27	.00	410	1.0 00	.61 50	-.08	8	ITEM8
P	1.25	.18	L	1.25	.20	.00	.27	.00	410	1.0 00	.61 50	.08	8	ITEM8
L	-4.03	.39	P	-5.94	1.00	1.91	1.08	1.78	354	.07 67	.08 06	.00	10	ITEM10
P	-5.94	1.00	L	-4.03	.39	-1.91	1.08	-1.78	354	.07 67	.08 06	.00	10	ITEM10
L	2.32	.28	P	2.08	.24	.23	.37	.63	404	.52 94	.78 01	.05	11	ITEM11
P	2.08	.24	L	2.32	.28	-.23	.37	-.63	404	.52 94	.78 01	-.05	11	ITEM11
L	-.35	.15	P	-.40	.15	.06	.21	.27	410	.78 70	.98 62	.13	12	ITEM12
P	-.40	.15	L	-.35	.15	-.06	.21	-.27	410	.78 70	.98 62	-.13	12	ITEM12
L	.71	.17	P	.54	.16	.17	.23	.71	410	.47 89	.53 22	.12	13	ITEM13
P	.54	.16	L	.71	.17	-.17	.23	-.71	410	.47 89	.53 22	-.12	13	ITEM13
L	.35	.16	P	.62	.16	-.27	.23	-1.19	412	.23 48	.15 62	-.26	14	ITEM14
P	.62	.16	L	.35	.16	.27	.23	1.19	412	.23 48	.15 62	.26	14	ITEM14
L	.68	.17	P	.85	.17	-.18	.24	-.74	411	.46 06	.24 46	-.09	15	ITEM15
P	.85	.17	L	.68	.17	.18	.24	.74	411	.46 06	.24 46	.09	15	ITEM15
L	-.48	.15	P	-.16	.15	-.32	.21	-1.5	411	.12 11	.29 95	-.31	16	ITEM16

PERSON CLASS	DIF MEASURE	DIF S.E	PERSON CLASS	DIF MEASURE	DIF S.E	DIF CONTRAST	JOINT S.E	Welch t	Welch d.f.	Welch Prob.	MantelHanzl Prob.	MantelHanzl Size	ITEM Number	Name
								5						
P	-.16	.15	L	-.48	.15	.32	.21	1.55	411	.1211	.2995	.31	16	ITEM16
L	.09	.16	P	.33	.15	-.24	.22	-1.10	410	2705	.5768	-.02	17	ITEM17
P	.33	.15	L	.09	.16	.24	.22	1.10	410	2705	.5768	.02	17	ITEM17
L	1.41	.21	P	1.35	.19	.06	.28	.22	409	.8260	.4299	.19	18	ITEM18
P	1.35	.19	L	1.41	.21	-.06	.28	-.22	409	.8260	.4299	-.19	18	ITEM18
L	-.15	.15	P	-.39	.14	.23	.21	1.11	410	.2690	.3555	.23	19	ITEM19
P	-.39	.14	L	-.15	.15	-.23	.21	-1.11	410	.2690	.3555	-.23	19	ITEM19
L	-2.24	.19	P	-2.46	.21	.22	.29	.78	412	.4366	.6148	.11	20	ITEM20
P	-2.46	.21	L	-2.24	.19	-.22	.29	-.78	412	.4366	.6148	-.11	20	ITEM20
L	.69	.17	P	1.18	.18	-.49	.25	-1.96	409	.0508	.0675	-.68	23	ITEM23
P	1.18	.18	L	.69	.17	.49	.25	1.96	409	.0508	.0675	.68	23	ITEM23
L	.14	.16	P	.06	.15	.09	.22	.40	411	.6880	.4539	.17	24	ITEM24
P	.06	.15	L	.14	.16	-.09	.22	-.40	411	.6880	.4539	-.17	24	ITEM24
L	1.14	.19	P	1.14	.18	.00	.26	.00	410	1.000	9272	-.02	26	ITEM26
P	1.14	.18	L	1.14	.19	.00	.26	.00	410	1.000	9272	.02	26	ITEM26
L	.12	.16	P	-.02	.15	.14	.21	.64	410	.5232	.6947	.11	29	ITEM29
P	-.02	.15	L	.12	.16	-.14	.21	-.64	410	.5232	.6947	-.11	29	ITEM29
L	-.76	.15	P	-.59	.15	-.17	.21	-.82	405	.4123	.4103	-.25	30	ITEM30
P	-.59	.15	L	-.76	.15	.17	.21	.82	40	.41	.41	.25	30	ITE

PERSON CLASS	DIF MEASURE	DIF S.E	PERSON CLASS	DIF MEASURE	DIF S.E	DIF CONTRAST	JOINT S.E	Welch t	Welch d.f.	Welch Prob.	MantelHanzl Prob.	MantelHanzl Size	ITEM Number	Name
									5	23	03			M30
L	-.87	.15	P	-1.19	.15	.31	.22	1.46	41	.14	.30	.36	32	ITE
									1	56	14			M32
P	-1.19	.15	L	-.87	.15	-.31	.22	-1.46	41	.14	.30	-.36	32	ITE
									1	56	14			M32
L	.74	.17	P	1.02	.17	-.29	.24	-1.18	41	.23	.05	-.07	33	ITE
									2	84	25			M33
P	1.02	.17	L	.74	.17	.29	.24	1.18	41	.23	.05	.07	33	ITE
									2	84	25			M33
L	-.07	.15	P	.11	.15	-.18	.21	-.86	41	.39	.35	-.22	34	ITE
									0	12	26			M34
P	.11	.15	L	-.07	.15	.18	.21	.86	41	.39	.35	.22	34	ITE
									0	12	26			M34
L	.36	.16	P	.10	.15	.26	.22	1.17	40	.24	.09	.34	36	ITE
									9	17	11			M36
P	.10	.15	L	.36	.16	-.26	.22	-1.17	40	.24	.09	-.34	36	ITE
									9	17	11			M36
L	.08	.16	P	-.28	.14	.36	.21	1.69	40	.09	.17	.39	37	ITE
									8	24	69			M37
P	-.28	.14	L	.08	.16	-.36	.21	-1.69	40	.09	.17	-.39	37	ITE
									8	24	69			M37
L	.15	.16	P	.01	.15	.14	.22	.67	40	.50	.34	.17	38	ITE
									7	60	09			M38
P	.01	.15	L	.15	.16	-.14	.22	-.67	40	.50	.34	-.17	38	ITE
									7	60	09			M38
L	.17	.16	P	.17	.15	.00	.22	.00	40	1.00	.87	.03	40	ITE
									9	00	24			M40
P	.17	.15	L	.17	.16	.00	.22	.00	40	1.00	.87	-.03	40	ITE
									9	00	24			M40

Size of Mantel-Haenszel slice: MHSLICE = .010 logits

Based on table 14, it can be explained that all items have a Manthaenszel probability value exceeding 0.05, so all of the above items are not detected by DIF. So it can be concluded that the function of things in the male and female gender groups has the same position so that the opportunity to answer correctly on the person with the same ability as the male / female Gender is the same.

Discussions

Several Rasch Model assumptions must be met before analyzing item fit and person fit, unidimensional, local independence, and group invariance. Three assumptions in the Item Response Theory: unidimensional, local independence, and parameter invariance (Hambleton et al., 1991). Then Sumintono and Widhiarso (2015) also explained that the instrument development procedure using the Rasch Model approach consisted of several steps, namely: 1) verification of assumptions such as local unidimension and independence, 2) testing of item fit and person fit.

The item fit and person fit test results showed that there were 30 out of 40 articles and 415 out of 480 people with ZSTD Outfit values ranging from - 2.0 to 2.0, so 30 items and 415 people were under the Rasch Model. The criteria for determining qualified items and fit persons refer to the opinion of Boone et al. (2014), describes one of the requirements for checking right things and suitable persons is guided by the Outfit Z Standard (ZSTD) value with fit criteria if -2.0 <ZSTD <+2,0.

The unidimensional test showed the value of Raw Variance Explained by Empirical Measures was 24.6%, exceeding 20%, meaning that the unidimensional assumptions were fulfilled. So it can be concluded that the items only measure quantitative literacy abilities. According to Reckase (1979), the criteria for unidimensional testing explain that items have unidimensional standards if the test variant's value is ≥ 20%.

Furthermore, the results of local independence testing show the value of residual correlations is less than 0.20, so it meets local independence. This means that the answers between items and between people do not affect each other. The local independence testing criteria refer to Christensen et al. (2016), who explains that the object or person is detected locally if the residual correlation value is> 0.2.

The group invariance test shows that each item's average measure value always increases, meaning that the item meets the group invariance criteria. So it can be concluded that all groups of person abilities (high - low) invariably refer to the same grain's characteristic curve. Testing group invariance can be guided by an increase in accurate scores along with ability levels (Kang, et al, 2018).

The validity of Internal Structures

The validity of this study was determined through evidence of internal structure. Goodwin and Leech (2003) state that truth based on evidence of internal structure is the accuracy of the internal components of a test with a construct of measure. Then completed by Mardapi (2008), which states the construct validity refers to the extent to which the test measures the concept of a theory as the basis for preparing the test as evidenced by the internal structure of the test. The statistical analysis used to determine this validity is the item accuracy index defined through the infit and Outfit coefficients in Rasch modeling (Sumintono and Widhiarso, 2015).

Based on the study results, 30 items have valid internal structural criteria from 40 items developed. These items are spread over each quantitative literacy dimension, which is presented in table 15.

Table 15: Distribution of Valid Items on Each Dimension

No	Dimension	Item Indicator	Item Number	Outfit ZSTD	Criteria
1	Interpretation	Interpreting graphical information in solving problems	1	-0,5	Valid
			3	-1,5	Valid

No	Dimension	Item Indicator	Item Number	Outfit ZSTD	Criteria
		Interpreting tabular information in solving problems	4	0,6	Valid
			5	-1,8	Valid
			6	-0,1	Valid
2	Representation	Represent information to geometric patterns	7	-1,0	Valid
			8	0,7	Valid
		Represent information to mathematical models	10	-1,7	Valid
			11	0,1	Valid
		Represent information to diagrams, graphs, or tables	12	-1,7	Valid
			13	0,7	Valid
3	Calculation	Use the calculation operations of addition / subtraction in solving mathematical problems	14	-0,5	Valid
			15	-0,6	Valid
			16	-0,9	Valid
		Using multiplication / division operations in solving mathematical problems	17	1,2	Valid
			18	0	Valid
			19	-1,3	Valid
		Using mixed count operations in solving mathematical problems	20	-0,9	Valid
			23	0,2	Valid
4	Analysis	Analyzing information about story items in problem-solving,	24	1,7	Valid
			26	-0,1	Valid
			29	0,4	Valid
			30	0	Valid
5	Assumption	Assumes results from story item information	32	-0,8	Valid
			33	0,7	Valid
			34	-0,2	Valid
			36	1,7	Valid
6	Communication	Conceptual problem solving the mathematical model	37	-1,1	Valid
			38	1,2	Valid
			40	0	Valid
Item total			30	30	

Table 15 shows that 30 items have a ZSTD Outfit value between - 2 to + 2, which means that all things have valid structural criteria. This is under Sumintono and Widhiarso (2015) opinion that the statistical analysis used to determine the validity of internal structures is the accuracy of the model item accuracy, selected through the infit and Outfit coefficients Rasch modeling. Then the criteria used to check its suitability are based on the Outfit Z Standard (ZSTD) with a fit value of -2.0 <ZSTD <+2.0 (Boone et al., 2014). Furthermore, the item also spreads, representing each dimension with the number of questions ≥ 3 per size. According to Neill (2011) opinion, the situation states that a domain of measurement contains at least three items. So it can be concluded that quantitative literacy instruments are feasible to use in terms of the internal structure's validity.

Instrument and Item Reliability

The instrument reliability in this study was determined through the internal consistency estimation method. The method is a way to assess the consistency of items by measuring once and requiring only one instrument, then calculating the

reliability coefficient as a determinant of the reliability criteria (Urbina, 2004). The reliability coefficient used is the KR 20 reliability coefficient. The determination of the KR-20 reliability coefficient is based because quantitative literacy instruments produce dichotomous data. Therefore KR-20 is suitable for use. This is consistent with Naga's (2013) opinion, which explains that the reliability coefficient of Kuder Richardson was put forward in 1937, which applies to the dichotomous data.

Based on the results of the study, it can be seen that the KR-20 reliability coefficient value is 0.70, so the reliability of the instrument has excellent and acceptable criteria, which means that the interaction between 415 person and 30 items as a whole has excellent and proper standards. According to Naga (2013), instrument reliability criteria can be accepted if 0.7 ≤ KR-20 reliability coefficient < 0.8. This was also reinforced by Sumintono and Widhiarso (2015), who explained that reliability was an excellent criterion if the reliability coefficient value was 0.7 to 0.8

The value of item reliability is 0.99, so item reliability has particular criteria, which means consistency of answers on 30 items has specific criteria. This is consistent with Fisher (2007) if the value of item reliability> 0.94 has a particular item consistency. While based on the separation value of 1.44, we obtained 12 groups of items. According to Fisher (2007), grouping items of more than five groups, indicating the quality of items with particular criteria.

Differential Item Function [DIF]

According to Dorans and Holland (1992), DIF refers to differences in the performance of items between two groups of comparable examinees. Based on the study results, DIF was determined through two different groups: male and female Gender. The results explain that 30 items have a Mantelhaenszel probability value of more than 0.05, so all the things above do not contain DIF. So it can be concluded that the performance of items on the same capable person from two different groups (male and female sexes) is the same.

Conclusion

Based on the results and discussion, it can be concluded that: 1) 30 items have valid criteria with Outfit ZSTD values between -2.0 to +2.0 on each item of the 40 items developed, 2) Instrument reliability has the right measures with a Cronbach Alpha (KR-20) value of 0.70, 3) Reliability 30 items have particular criteria with the value of item reliability of 0.99 with 12 groups of objects, 4) 30 articles are free from DIF, with the Mantelhaenszel probability value on each item more than 0.05.

Limitations and Future Studies

This research is still limited to the analysis of quantitative literacy items such as validity, reliability, and DIF analysis so that further research will be carried out to analyze quantitative literacy skills using these items.

Acknowledgment

Thanks are given to 1) elementary school students in the City of Semarang and Demak Regency, Indonesia as test subjects of quantitative literacy instruments for elementary school students; 2) Prof. Dr. Ani Rusilowati, M.Pd.; Prof. Sukestyarno, M.S., Ph.D.; Dr. Masrukan, M.Pd.; and Tri Suryaningsih, M.Pd. as an expert validator of quantitative literacy instruments for elementary school students.

References

[1] Bookman, J., Ganter, S. L., & Morgan, R. (2008). Developing assessment methodologies for quantitative literacy: A formative study. *The American Mathematical Monthly*, *115*(10), 911-929. https://doi.org/10.1080/00029890.2008.11920609

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

[3] Burdette, A. M., & McLoughlin, K. (2010). Using census data in the classroom to increase quantitative literacy and promote critical sociological thinking. *Teaching Sociology*, *38*(3), 247-257. https://doi.org/10.1177/0092055X10370114

[4] Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. *Applied psychological measurement*, *41*(3), 178-194. https://doi.org/10.1177/0146621616677520

[5] Cohen & Swerdlik. (2010). *Psychological Testing and Assessment: An Introduction to Tests and Measurement 7th Edition*. USA: McGraw-Hill Primis.

[6] Crocker, L., & Algina, J., 2008. *Introduction to Classical and Modern Test Theory*. USA: Cengage Learning

[7] Dorans, N. J., & Holland, P. W. (1992). DIF Detection and Description: Mantel-Haenszel and Standardization 1, 2. *ETS Research Report Series*, *1992*(1), i-40. https://doi.org/10.1002/j.2333-8504.1992.tb01440.x

[8] Field, A. P. (2005). *Discovering Statistics Using SPSS*. Sage Publications Inc.

[9] Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and evaluation in Counseling and Development*, *36*(3), 181-191. https://doi.org/10.1080/07481756.2003.11909741

[10] Hallett, D. H. (2003). The role of mathematics courses in the development of quantitative literacy. *Quantitative Literacy: Why Numeracy Matters for School and Colleges*, 91-98.

[11] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.

[12] Irawan, H. (2019). Masalah Kesetaraan Dalam Pendidikan Matematika. *PAPATUNG: Jurnal Ilmu Administrasi Publik, Pemerintahan dan Politik*, *2*(1), 57-68.

[13] Irianto, P. O., & Febrianti, L. Y. (2017, June). Pentingnya penguasaan literasi bagi generasi muda dalam menghadapi MEA. In *Proceedings Education and Language International Conference* (Vol. 1, No. 1).

[14] Kang, H. A., Su, Y. H., & Chang, H. H. (2018). A note on monotonicity of item response functions for ordered polytomous item response theory models. *British Journal of Mathematical and Statistical Psychology*, *71*(3), 523-535. https://doi.org/10.1111/bmsp.12131

[15] De Lange, J. (2006). Mathematical literacy for living from OECD-PISA perspective.

[16] Mardapi, D. (2008). Teknik penyusunan instrumen tes dan nontes. *Yogyakarta: Mitra Cendekia*, *127*, 88.

[17] Martony, O. (2019). The Influence of Balanced Nutrition Extension through Poster Media on the Improvement of Knowledge and Change Attitude. *International Journal of Science and Society*, *1*(1), 98-106.

[18] Mayes, R. L., Forrester, J., Christus, J. S., Peterson, F., & Walker, R. (2014). Quantitative reasoning learning progression: The matrix. *Numeracy: Advancing Education in Quantitative Literacy*, *7*(2), 1-20.

[19] *McArthur D. L., 1987. Alternative Approaches to the Assessment of Achievement. Kluwer Academic Publishers*

[20] Naga, D. S. (2012). Teori sekor pada pengukuran mental. *Jakarta: PT Nagarani Citrayasa*

[21] Neill, J., 2011. "Exploratory Factor Analysis Survey Research and Design Psikologi." Lecture Material. Centre or Applied Psychology University of Canberra.

[22] OECD. (2018). *Pisa 2015 Result In Focus*. PISA: OECD Publishing

[23] Ontario Education. (2014). *Leading Math Success: Mathematical Literacy, Grades 7-12*. The Final Report of the Expert Panel on Student Success in Ontario

[24] *Peraturan Menteri Pendidikan Nasional Nomor 22 Tahun 2006 Tentang Standar Isi Pendidikan Dasar dan Menengah*. 2006. Jakarta: Lembaran Negara RI Nomor 4496.

[25] Rafianti, I., Setiani, Y., & Novaliyosi, N. (2018). Profil kemampuan literasi kuantitatif calon guru matematika. *JPPM (Jurnal Penelitian dan Pembelajaran Matematika)*, *11*(1). http://dx.doi.org/10.30870/jppm.v11i1.2985

[26] Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of educational statistics*, *4*(3), 207-230. https://doi.org/10.3102/10769986004003207

[27] Schuhmann, P. W., McGoldrick, K., & Burrus, R. T. (2005). Student quantitative literacy: Importance, measurement, and correlation with economic literacy. *The American Economist*, *49*(1), 49-65. https://doi.org/10.1177/056943450504900104

[28] Steen, L. A. (2001). *Mathematics And Democracy: The Case For Quantitative Literacy*. USA: Woodrow Wilson National Fellowship Foundation

[29] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.

[30] Suwarsono, St. (1987). *Trait-Treatment Interaction dalam Pendidikan Matematika.* Yogyakarta: IKIP Sanata Dharma Yogyakarta

[31] Sweet, S., & Strand, K. (2006). Cultivating quantitative literacy: The role of sociology. *Teaching Sociology*, *34*(1), 1-4. https://doi.org/10.1177/0092055X0603400101

[32] Urbina, S. (2014). *Essentials of psychological testing*. John Wiley & Sons.

[33] Wijanto, S.H. 2008. *Structural Equation Modeling dengan Lisrel 8.8*. Yogyakarta: Graha Ilmu

[34] Wilkins, J. L. (2010). Modeling quantitative literacy. *Educational and Psychological Measurement*, *70*(2), 267-290. https://doi.org/10.1177/0013164409344506