The Seybold Report

ISSN 1533-9211

DOI: 10.5281/zenodo.11108908

INSTRUMENTS DEVELOPMENT FOR THE ABILITY TO SOLVING MATHEMATICS STORY QUESTIONS FOR PRIMARY SCHOOL STUDENTS

SEPNI YANTI [1], AWALUDDIN TJALLA [2]* and ANAN SUTISNA [3]

[1,2,3] Educational Research and Evaluation Doctoral Program, Universitas Negeri Jakarta.
Email: [1]sepniyanti@gmail.com, [2]awaluddin.tjalla@unj.ac.id (*Corresponding Author), [3]asutisna@unj.ac.id

Abstract

This research aims to develop an instrument for elementary school students to solve mathematics story problems. Students' ability to solve math story problems is often a challenge, and previous research shows that these difficulties can be overcome through appropriate instruments. In this context, the ADDIE Model is used as a framework for developing educational instruments. Considering the low ranking of Indonesian students' mathematics abilities in international assessments, this research is necessary. The question instrument was created using the Rasch Model to obtain fit question items and analyze the Winsteps software output results. This measurement aims to provide information that can be used in developing better item instruments related to the ability to solve math story problems. This research shows the importance of developing appropriate and valid instruments to measure students' abilities in solving mathematics story problems. Using the ADDIE Model, instrument development can be carried out in a structured, practical and focused manner. The results of this research can improve the mathematical problem-solving abilities of elementary school students in Indonesia.

Keywords: Development of Educational Instruments, Mathematical Problem Solving Ability, Mathematical Story Question Instruments, Elementary School.

INTRODUCTION

Mathematics is a knowledge that everyone needs in everyday life. Mathematics is widely used in everyday life. Good skills in using mathematics are necessary for solving problems in everyday life. Mathematics learning is carried out from elementary to tertiary level, so Indonesian students will likely have good skills. The better Indonesian students' skills in Mathematics, the better the quality of education in Indonesia will be. Evaluation is needed to determine the quality of education in Indonesia. For this purpose, Indonesian students have taken part in several international scale assessments. Several international scale assessments include PISA (The Program International Student Assessment) and TIMSS (The Trends in International Mathematics and Science Study).

Indonesian students' mathematics abilities can be proven to be at the bottom of the PISA (The Program for International Student Assessment) activities for children aged 15 years in Mathematics, Science and Literacy. Since Indonesia participated in the first PISA from 2000 until 2018, Indonesian students have been in the lowest position among participating countries (Budi, 2018; OECD, 2015). Indonesia's ranking experienced a significant increase in score in 2015 by 11 points, but the ranking is still below OECD (Organization for Economic Co-operation Development) countries. The results of the latest release in 2018, Indonesia's ranking fell again. Indonesia is ranked 7th from the bottom (Permana, 2019; Program, Assessment,

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

2018). Since Indonesia participated in the first PISA from 2000 until 2018, Indonesian students were in the lowest position of the participating countries (Budi, 2018; OECD, 2015b). Indonesia's ranking experienced a significant increase in score in 2015 by 11 points, but the ranking is still below OECD (Organization for Economic Co-operation Development) countries. The results of the latest release in 2018, Indonesia's ranking fell again. Indonesia is ranked seventh from the bottom (Permana, 2019; Program, Assessment, 2018).

The PISA ranking of Indonesian students in Mathematics in 2000 had an average score of 367 points, which was ranked 39th, then in 2003, it had an average score of 360 points, which was ranked 38th, then in 2006, it had an average score of 391 points, which was ranked 50th. In 2009, it had an average score of 371 points, which was ranked 61st; in 2012, it had an average score of 375 points, which was ranked 64th. Then, in 2015, it had an average score of 386 points, which was ranked 62nd, and finally, in 2018, it had an average of 379 scores with a ranking of 67 (Permana, 2019; Program, Assessment, 2018). Based on the PISA report, Indonesian students' abilities in Mathematics are in the low category and need to receive optimal attention.

Large-scale assessments have been enhanced to measure student ability in the reading, mathematics, and science content domains. The Trend in International Mathematics and Science Study (TIMSS) is a large-scale international and prestigious assessment in Mathematics and Science. TIMSS was first held in 1995 and is carried out every 4 years from 1999, 2003, 2007, and so on by the International Association for the Evaluation of Educational Achievement (IEA).

The main goal of TIMSS is to act as an educational policy maker and educator through reliable and timely data. As a reliable basis for improving their educational performance. TIMSS cognitive domain (Prastyo, 2020): knowing, applying and reasoning. TIMSS ability levels, namely: level 1 advanced ability, level 2 high-level ability, level 3 intermediate-level ability, and level 4 low-level ability. TIMSS Indonesia results from 2003 to 2015 including in 2003, it was ranked 35th with an average score of 411 points, then in 2007 it was ranked 36th with an average score of 397 points, then in 2011 it was ranked 38th with an average score was 386 points and in 2015 it was ranked 44th with a total average score of 397 points (Hadi & Novaliyosi, 2019; Nizam, 2016).

The TIMSS score categories are a score of 400 in the low category, a score of 475 in the medium category, a score of 550 in the high category, and a score of 625 in the very high category (Hadi & Novaliyosi, 2019; Nizam, 2016). Based on the information above, Indonesia's position from 2003 to 2015 was in the low category.

The factors that cause students to experience learning difficulties consist of internal factors and external factors. Internal factors include students' attitudes towards learning, students' learning motivation, intellectual abilities and students' physical health. Meanwhile, external factors include more variety in teachers' teaching methods, inadequate use of media, school facilities and infrastructure, and a less supportive family environment (Unaenah et al., 2022; Cahyono, 2019).

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

Students' mathematical problem-solving abilities can be interpreted as students' ability to understand problems, plan problem-solving strategies, implement the chosen solution strategy, and re-examine the solution to the problem so that the solution can then be taken systematically and inseparably from the correct representation of the problem (Polya, Batubara, 2017 ; Siagan et al., 2019). Story problems are an essential part of the mathematics curriculum in elementary schools (Kurshumlia and Vula, 2019). Story questions have an essential role that is usually used to determine students' problem-solving abilities. Story problems are questions considered to have a higher level of difficulty than mathematics questions that display mathematical models directly. In story problems, students are expected to be able to find problems that must be solved in the problem (Dwidarti et al., 2019). The problem that students often need helping in learning mathematics is solving word problems.

Other facts also show that students' mathematical problem-solving abilities still need to improve. This is proven by research conducted by Sol et al., 2020; Pertiwi et al., 2021; Simamora et al., 2019; Surya et al., 2017; Meryansumayeka et al., 2021; Hasbi et al., 2019; Wasiah et al., 2020; Utami et al., 2018; students experience students' inability to solve mathematical problems.

Mathematical ability is a person's ability to memorize, understand, explain and apply mathematical concepts to solve problems in everyday life. One form of application for solving mathematical problems in everyday life is using story problems (Utami and Endaryanto, 2018). The ability to complete math word problems is an essential skill in students' mathematical development at the elementary school level based on research conducted by Patel et al., 2021; Cobbe et al., 2021; Tambunan, 2019; Kenedi et al., 2019; Math word problems allow students to apply mathematical concepts in real-life situations, helping them develop a deeper understanding of mathematics and problem-solving skills.

We can see students' difficulties in solving math problems in the form of story problems from their ability to read, understand, process, transport, solve process skills and write answers (Surya, 2018) based on Newman's theory of five important activities in learning to bring out students' abilities in solving story problems. Includes the following five stages: (1) reading, (2) understanding, (3) transformation, (4) process skills, and (5) writing answers/encoding (Sesanti and Bere 2020). Problems experienced by elementary school students who experience errors in solving story problems based on the Polya procedure include the following four stages: (1) understanding the problem, (2) making a plan, (3) implementing the plan, (4) checking again (Utami et al., 2019; Yuwono et al., 2018; Vilianti et al., 2018). Solving math word problems is more challenging than solving multiple-choice questions or short descriptions. Story problems require students to understand the problem in the story, determine strategies, carry out strategies, and conclude answers (Octavia, 2017). They need a good understanding of mathematical concepts, analyzing problems, and applying their knowledge. As students move to higher levels, the difficulty level of math word problems tends to increase. An instrument is needed to evaluate students' abilities in solving mathematical story problems.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

Measuring the ability to solve math story problems can be challenging to do. The instruments used to measure this ability must consider the diversity of types of story questions, difficulty level, and cognitive abilities being tested. Developing appropriate and valid instruments is essential to understand student abilities accurately. The ADDIE model is an instructional design development model with a learner-centered learning approach rather than a traditional teacher-centered approach so that effective learning can occur (Harmelin, 2022). Each learning component is regulated by learning outcomes, which have been determined after thoroughly analyzing student needs. These phases sometimes overlap and can be interrelated, but they provide dynamic and flexible guidelines for developing effective and efficient teaching.

The ADDIE model is a framework for the development of educational instruments. A theoretical study of this model will help researchers understand the steps to develop educational instruments, including analysis, design, development, implementation, and evaluation. The advantages of the ADDIE model for developing instruments to solve story problems include being structured, practical and focused (Harmelin, 2022). Compared to using the previous model, it requires a relatively long time because the procedure is relatively complex and requires quite significant financial resources, as in the Borg and Gill development model (Maydiantoro, 2021). Applying the ADDIE Model to develop an instrument for solving math story problems for fourth-grade elementary school students has never been done before. The development of story problem-solving instruments using the ADDIE Model has been carried out by several researchers, with research subjects being middle-class and upper-class students, as evidenced by research conducted by (Rochsun & Agustin, 2020; Prabowo et al., 2020; Riyati & Suparman, 2019; Azzatia, 2019).

Based on learning difficulties, especially in the ability to solve math story problems experienced by fourth-grade elementary school students, the researchers wanted to develop an instrument for the ability to solve math story problems for fourth-grade elementary school students in Depok City. Questions were discussed in the Number Domain in the Odd Semester of the 2023-2024 Academic Year. Question items in the Number Domain include numbers and their place values, the value of the rupiah currency, the most significant common factor and the smallest common multiple. Student ability measurements are carried out after the Mid-Odd Semester learning activities. This measurement can be used as a reference for developing the items related to the ability to solve story problems (word problems).

Statement of the Problem

The author focuses on two instrument development processes, including measuring elementary school student's ability to solve mathematics story problems. The steps to achieve this goal will include needs analysis, instrument design, development, implementation, and evaluation. Then, the research will explore the characteristics of the developed test items, with particular emphasis on each question item's difficulty level. Thus, this research will provide in-depth insight into developing an effective evaluation instrument and understanding the critical aspects that must be considered in developing mathematics story questions for elementary school students.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

Theoretical Framework

The theoretical framework for this research includes several main concepts relevant to the development of an evaluation instrument for students' abilities in solving student-centered mathematics story problems by applying the ADDIE Model (Analysis, Design, Development, Implementation, Evaluation), which will be used as a framework for the development of evaluation instruments. The ADDIE model emphasizes the importance of analyzing students' needs, designing appropriate instruments, developing them, implementing them, and evaluating their effectiveness.

Understanding the characteristics of mathematics story problems and the difficulty level of the questions will be an essential basis for developing the instrument. This concept includes an understanding of reading, understanding, transformation, process skills, and writing answers in solving story problems. Thus, this theoretical framework will help develop appropriate and effective evaluation instruments, considering an understanding of mathematical problem-solving, student-centered learning models, and the characteristics of mathematical story questions and their difficulty level.

METHODOLOGY

This research uses research and development methods. The method in this research is a combination of instrument development and research and development models. This research aims to evaluate elementary school students' understanding and skills in solving mathematical story problems, focusing on measuring their achievement in the mathematical concepts contained in the word problems. This research assesses the potential positive impact of using the developed instrument on improving elementary school students' learning and understanding in solving mathematics story problems, focusing on the practical application of research results in an educational context.

The research and development model is used to produce specific products and test the product's effectiveness (Donovan, 2013; Zulyadaini, 2017). In this research, the product is an instrument for solving story problems in elementary school mathematics. The development research model used is the ADDIE model. The research and development model of the ADDIE Model (Branch, 2009; Branch, 2013; Sharifah & Faaizah, 2015; Muruganantham, 2015; Hess & Greer, 2016; Zulyadaini, 2017) includes: 1) Analyze, 2) Design, 3) Development, 4) Implementation, and 5) Evaluation.

The target population in this study was limited to grade IV elementary school students in Depok City. Meanwhile, the affordable population is represented by class IV students at SDN with An accreditation and the Merdeka Curriculum. Phase I of the trial is a small-scale sample, followed by data collection, and phase II is the actual sample. The samples used were five elementary schools in Depok City, with a sample size of 500 respondents.

The instrument was tested for validity by 4 experts to see the validity of the content of the development of the school mathematics story question instrument. The analysis obtained valid evidence, showing that the Aiken' V content validity score was 0.81. This proves that the

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

mathematical content related to the development of the Story Question Instrument is by the measurement objectives or measurement accuracy (Heale & Twycross, 2015). The test-retest reliability of the mathematics story problem instrument was ensured. The scores obtained from the two responses were subjected to statistical analysis involving Pearson Product Moment Correlation at a significance level of 0.05. This instrument produces a reliability coefficient of 0.918, which is considered high enough for this research.

RESULTS

This research is intended to seek information and data that can be used to empirically describe the quality of test instruments based on elements of validity and reliability, which are processed with the help of the Rasch model and Winsteps software. Following are the results of phase 1 of the trial

Table 1: Data on the distribution of misfit or unfit questions using the Rasch model for the first test

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
17	43	200	1.28	.18	1.06	.7	1.19	1.4	A .13	.24	78.0	78.6	S17
15	31	200	1.70	.20	1.03	.3	1.16	.9	B .14	.22	85.0	84.5	S15
14	93	200	.02	.15	1.11	2.5	1.16	2.9	C .13	.30	60.5	63.0	S14
10	53	200	.98	.17	1.07	.9	1.15	1.4	D .14	.26	73.0	73.7	S10
16	98	200	-.09	.15	1.11	2.4	1.13	2.5	E .15	.31	54.5	62.9	S16
8	129	200	-.79	.16	1.10	1.6	1.11	1.6	F .16	.30	65.5	67.8	S8
22	85	200	.20	.15	1.08	1.8	1.11	1.8	G .18	.30	56.0	63.6	S22
9	44	200	1.24	.18	1.05	.6	1.10	.8	H .16	.25	78.5	78.1	S9
21	70	200	.55	.15	1.03	.5	1.09	1.2	I .23	.29	67.0	66.7	S21
24	83	200	.24	.15	1.07	1.4	1.08	1.3	J .20	.30	58.0	63.8	S24
23	63	200	.72	.16	1.06	.9	1.07	.8	K .19	.28	67.0	69.3	S23
6	138	200	-1.02	.16	.99	-.1	1.06	.7	L .29	.29	74.0	70.7	S6
12	94	200	.00	.15	1.01	.3	1.01	.2	M .29	.30	63.0	63.0	S12
26	93	200	.02	.15	1.01	.2	1.01	.1	N .30	.30	64.5	63.0	S26
3	148	200	-1.28	.17	1.00	.0	1.00	.0	O .28	.28	74.5	74.5	S3
5	156	200	-1.52	.18	.98	-.2	1.00	.0	n .28	.27	81.0	78.1	S5
7	118	200	-.54	.15	1.00	.0	.99	-.2	m .31	.31	65.5	65.2	S7
4	165	200	-1.83	.19	.98	-.1	.98	-.1	l .27	.25	82.5	82.5	S4
11	53	200	.98	.17	.98	-.3	.94	-.5	k .31	.26	75.0	73.7	S11
20	113	200	-.42	.15	.97	-.5	.96	-.7	j .35	.31	66.0	64.3	S20
27	74	200	.45	.15	.97	-.5	.95	-.7	i .34	.29	65.0	65.6	S27
13	133	200	-.89	.16	.94	-.9	.97	-.4	h .37	.30	76.5	69.0	S13
19	114	200	-.45	.15	.96	-.9	.95	-.9	g .37	.31	68.5	64.5	S19
25	104	200	-.22	.15	.95	-1.2	.95	-1.1	f .38	.31	69.5	63.1	S25
2	94	200	.00	.15	.94	-1.5	.91	-1.8	e .40	.30	63.0	63.0	S2
18	90	200	.09	.15	.92	-1.9	.90	-1.9	d .42	.30	67.5	63.1	S18
28	81	200	.29	.15	.89	-2.3	.88	-2.1	c .45	.30	72.5	64.2	S28
30	103	200	-.20	.15	.85	-3.6	.85	-3.3	b .52	.31	72.0	63.1	S30
29	72	200	.50	.15	.83	-3.3	.78	-3.2	a .54	.29	73.0	66.2	S29
MEAN	97.8	200.0	-.23	.21	1.00	-.1	1.01	.0			69.5	68.6	
S.D.	38.3	.0	1.48	.30	.07	1.5	.10	1.5			7.6	6.4	

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

Based on the results of the analysis of instrument trials to solve Mathematics story questions from 200 respondents with 30 questions. The obtained information for each evaluation item shows the level of suitability of the item to the measurement model used. Some items, such as Numbers 10, 15, and 23, show an excellent fit to the model, whereas others, such as Numbers 6 and 14, tend to be poor fit. Despite this, most items are still acceptable with quite good fit scores, although some have infit and outfit scores that are slightly above average. Further analysis may be needed to determine whether inappropriate items require revision or adjustment to improve their quality in the context of the measurement model used.

Furthermore, the second testing stage used much larger respondent data. In the second stage, the following results were obtained:

Table 2: Data on the distribution of misfit or unfit questions using the Rasch model for the second test

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
12	94	300	1.66	.13	1.30	4.8	1.38	4.2	A .00	.35	60.7	72.4	SB12	0
17	54	300	2.49	.16	1.17	1.7	1.35	2.3	B .10	.31	80.7	82.8	SB17	0
18	120	300	1.23	.13	1.14	3.0	1.26	4.0	C .17	.35	60.7	66.9	SB18	0
24	128	300	1.10	.12	1.18	4.1	1.24	3.9	D .14	.35	60.0	65.5	SB24	0
23	99	300	1.57	.13	1.14	2.4	1.21	2.6	E .18	.35	65.0	71.3	SB23	0
21	142	300	.88	.12	1.16	3.9	1.20	3.4	F .17	.35	54.3	64.0	SB21	0
16	156	299	.67	.12	1.04	1.0	1.16	2.8	G .28	.35	66.2	64.0	SB16	0
25	129	300	1.08	.12	1.10	2.4	1.10	1.8	H .24	.35	55.0	65.4	SB25	0
3	299	300	-5.25	1.00	1.01	.3	1.09	.4	I .02	.04	99.7	99.7	SB3	0
19	189	300	.16	.13	1.02	.5	1.00	.0	J .31	.33	66.7	67.5	SB19	0
26	190	300	.14	.13	1.00	.1	.96	-.5	K .34	.33	69.0	67.7	SB26	0
10	134	300	1.01	.12	.95	-1.2	1.00	.0	L .39	.35	70.0	64.8	SB10	0
20	176	300	.36	.12	.99	-.2	.95	-.8	M .36	.34	65.3	65.2	SB20	0
14	134	300	1.01	.12	.98	-.4	.99	-.2	N .37	.35	64.0	64.8	SB14	0
4	292	300	-3.13	.36	.92	-.1	.97	.1	O .24	.12	97.3	97.3	SB4	0
13	193	300	.09	.13	.97	-.7	.91	-1.2	n .38	.33	70.7	68.3	SB13	0
7	258	300	-1.26	.17	.96	-.3	.93	-.3	m .29	.25	85.7	86.1	SB7	0
8	270	300	-1.67	.20	.95	-.3	.82	-.8	l .30	.22	90.0	90.0	SB8	0
30	210	300	-.20	.13	.94	-1.0	.92	-.9	k .38	.32	72.7	72.3	SB30	0
9	156	300	.67	.12	.94	-1.6	.93	-1.2	j .41	.35	71.0	63.9	SB9	0
5	288	300	-2.70	.30	.91	-.3	.51	-1.5	i .30	.15	96.0	96.0	SB5	0
29	177	300	.35	.13	.91	-2.4	.85	-2.6	h .46	.34	67.0	65.3	SB29	0
2	287	300	-2.61	.29	.90	-.3	.47	-1.7	g .32	.15	95.7	95.7	SB2	0
22	122	300	1.19	.13	.90	-2.3	.85	-2.6	f .47	.35	70.0	66.5	SB22	0
15	112	300	1.35	.13	.89	-2.3	.89	-1.7	e .47	.35	76.0	68.4	SB15	0
11	139	300	.93	.12	.88	-3.1	.86	-2.7	d .49	.35	73.7	64.2	SB11	0
6	283	300	-2.32	.25	.88	-.5	.55	-1.6	c .35	.17	94.3	94.3	SB6	0
28	153	300	.72	.12	.87	-3.5	.83	-3.4	b .50	.35	70.7	63.8	SB28	0
27	170	300	.46	.12	.84	-4.3	.79	-3.7	a .53	.34	76.0	64.7	SB27	0
MEAN	177.7	300.0	.00	.19	1.00	.0	.96	-.1			73.9	73.8		
S.D.	67.4	.3	1.72	.17	.11	2.2	.22	2.2			12.9	12.3		

Based on the analysis of instrument trials for solving Mathematics story questions from 300 respondents with 30 questions. The obtained information for each evaluation item shows the level of suitability of the item to the measurement model used. In this table, several items, such as Numbers 12, 17, and 18, show poor fit with the measurement model as measured by the

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

value INFIT and quite a high OUTFIT. Despite this, most items were still well accepted, with some showing good fit values to the model. However, items Numbers 3 and 4 stand out as items with low goodness-of-fit values, which may require special attention to be corrected in the context of the measurement model. Further analysis can be performed to determine the cause of the nonconformity and appropriate corrective steps.

DISCUSSION

The analysis results show significant differences between the first trial and instrument development using the ADDIE method. The increase in item reliability scores from 0.96 to 0.98 indicates an increase in consistency in measuring students' abilities in solving mathematics story problems. Although person reliability increased from 0.58 to 0.67, this value still needs to be improved to ensure greater consistency in student performance evaluation. In addition, the Cronbach's alpha value remains stable at 0.918, indicating good internal consistency of the instrument.

Meanwhile, the increase in the separation score from 5.03 to 6.74 shows an increase in accuracy in measuring variations in student abilities. This indicates that the instrument developed can differentiate between students with different abilities more accurately.

Although the Outfit Mean Square Statistics (Outfit MNSQ) value tends to be stable at around 1.01 in both the person and item columns, a slight decrease from 1.01 to 0.96 in instrument development indicates an increase in the suitability between student responses and the Rasch measurement model. This indicates that the instrument developed can measure students' abilities better than the previous version.

In addition, the Outfit Z Standard (Outfit ZSTD) value, which is close to 0 in the person and item tables, indicates that student responses are consistent with the Rasch measurement model. However, it should be noted that an Outfit ZSTD value close to 0 does not guarantee that the instrument fully complies with the Rasch model, so further evaluation must be carried out to ensure its overall validity and reliability.

CONCLUSION

Based on the analysis results and comparison between the first trial and the development of the instrument using the ADDIE method, there is a significant increase in the quality of the instrument for the ability to solve mathematics story problems. The development of instruments using the ADDIE method and the application of the Rasch Model increased item reliability, personal reliability, and the instrument's ability to differentiate between students with different abilities. A stable Cronbach's alpha value indicates good internal consistency of the instrument. In contrast, an increase in the separation value indicates an increase in the accuracy of measuring variations in student abilities. Although some improvement is still needed in personal reliability, the instrument developed can measure student abilities better and more consistently than the previous version. These results indicate that developing an instrument for solving mathematics story problems using the ADDIE method and the Rasch Model is a

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

practical approach to improving the quality of evaluation in this field. The next step is to carry out further validation to ensure the suitability of the instrument shows an item score (item reliability) of 0.98, person reliability of 0.67, and a Cronbach's alpha value of 0.918, a separation value of 6.74 while an Outfit Mean Square Statistics (Outfit MNSQ) value) of 0.96 in the person and item columns. The Outfit Z Standard (Outfit ZSTD) value is 0.01 in the person table and -0.01 for the item table.

References

1) Abdullah, A. H., Nurarfah, S., & Rahman, S. A. (2017). *Metacognitive Skills of Malaysian Students in Non-Routine Mathematical Problem Solving Habilidades Metacognitivas dos Estudantes Malaios na Resolução de Problemas Não Rotineiros de Matemática*. 310–322.

2) Ali, R. (2018). *Rasch Model Validation Of Instrument To Measure Students Readiness To Embedded Systems Design Course*. *February*.

3) Anderson. (2001). *Taxonomy for Learning, Teaching, and Assessing:*

4) Andrich, D. (1989). Distinctions Between Assumptions and Requirements in Measurement in the Social Sciences. *Mathematical and Theoretical Systems*, 7–16.

5) Anees, S. (2017). Analysis of Assessment Levels of Students' Learning According t Further after analysis, the frequency and percentage on each level according to cognitive domain of Bloom's Taxonomy was found to measure the results. Cognitive Domain of Bloom's Taxonomy. *Online Submission*, 1–14.

6) Baghaei, P., Yanagida, T., & Heene, M. (2017). Development Of A Descriptive Fit Statistic for the Rasch Model. *North American Journal of Psychology*, *19*(1), 155–168.

7) Bahar, M., & Aksüt, P. (2020). Investigation on the effects of activity-based science teaching practices in the acquisition of problem solving skills for 5-6 year old pre-school children. *Journal of Turkish Science Education*, *17*(1), 22–39. https://doi.org/10.36681/tused.2020.11

8) Bernard, H. (2011). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*.

9) Bichi, A. A., Talib, R., Atan, N. A., & Ibrahim, H. (2019). Validation of a Developed University Placement Test Using Classical Test Theory and Rasch Measurement Approach. *International Journal of ADVANCED AND APPLIED SCIENCES*, *6*(6), 22–29. https://doi.org/10.21833/ijaas.2019.06.004

10) Bloom, B. S. (1956). *Taxonomy Of Education Objectives*.

11) Boman, J., Currie, G., MacDonald, R., Miller-Young, J., Yeo, M., & Zettel, S. (2017). Overview of Decoding Across the Disciplines. *New Directions for Teaching and Learning*, *2017*(150), 13–18. https://doi.org/10.1002/tl.20234

12) Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates Publishers.

13) Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, *15*(4), 1–7. https://doi.org/10.1187/cbe.16-04-0148

14) Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in Developmental Science : Situations , Shortcomings , Solutions , and Standards Q. *Developmental Review*, *33*(4), 357–370. https://doi.org/10.1016/j.dr.2013.08.003

15) Branch, R. M. (2009). Instructional Design: The ADDIE Approach. In *Springer*. # Springer ScienceþBusiness Media, LLC 2009. https://doi.org/10.1016/b978-0-12-411355-8.50014-0

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

16) Branch, R. M. (2013). *Robert Maribe Branch*. 24–27. https://doi.org/10.13811/j.cnki.eer.2013.08.005

17) Budi, K. (2018, April). *Indonesia Kirim Guru Ke Korea Selatan untuk Pelajari HOTS*.

18) Cappelleri, J. C., Jason Lundy, J., & Hays, R. D. (2014). Overview of Classical Test Theory and Item Response Theory for the Quantitative Assessment of Items in Developing Patient-Reported Outcomes Measures. *Clinical Therapeutics*, *36*(5), 648–662. https://doi.org/10.1016/j.clinthera.2014.04.006

19) Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics. *Procedia - Social and Behavioral Sciences*, *129*, 133–139. https://doi.org/10.1016/j.sbspro.2014.03.658

20) Chan, S. W., Looi, C. K., & Sumintono, B. (2020). Assessing computational Thinking Abilities Among Singapore Secondary Students: a Rasch Model Measurement Analysis. *Journal of Computers in Education*, *0123456789*. https://doi.org/10.1007/s40692-020-00177-2

21) Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, *41*(3), 178–194. https://doi.org/10.1177/0146621616677520

22) Conrad, K. J., Conrad, K. M., Mazza, J., Riley, B. B., Funk, R., Stein, M. A., & Dennis, M. L. (2017). Dimensionality, Hierarchical Structure, Age Generalizability, and Criterion Validity of the GAIN's Behavioral Complexity Scale. *Physiology & Behavior*, *176*(1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

23) CoPo, A. R. I. (2016). Students Initial Knowledge State And Test Design: Towards A Valid And Reliable Test Instrument. *Journal of College Teaching & Learning (TLC)*, *12*(3), 189. https://doi.org/10.19030/tlc.v12i3.9310

24) Crane, P. K., Gibbons, L. E., Jolley, L., & Van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care*, *44*(11 SUPPL. 3), 115–123. https://doi.org/10.1097/01.mlr.0000245183.28384.ed

25) Creswell, J., & Clark, V. (2011). Designing and Conducting Mixed Method Research. In *Analysing and interpreting data in mixed methods research*.

26) Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's Taxonomy to Enhance Student Learning in Biology. *CBE—Life Sciences Education*, *7*, 368–381.

27) Dagostino, L., Carifio, J., Bauer, J. D. C., Zhao, Q., & Hashim, N. H. (2015). Using Bloom's Revised Taxonomy to Analyze A Reading Comprehension Instrument. *Current Issues in Education*, *18*(2).

28) Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch Measurement Model to Explore the Relationship Between Content Knowledge and Topic-Specific Pedagogical Content Knowledge for Organic Chemistry. *International Journal of Science Education*, *38*(9), 1483–1503. https://doi.org/10.1080/09500693.2016.1196843

29) Devon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A Psychometric Toolbox For Testing Validity and Reliability. *Journal of Nursing Scholarship*, *39*(2), 155–164. https://doi.org/10.1111/j.1547-5069.2007.00161.x

30) Dikko, M. (2016). Establishing construct validity and reliability: Pilot testing of a qualitative interview for research in takaful (Islamic insurance). *Qualitative Report*, *21*(3), 521–528.

31) Donovan, M. S., & Donovan, M. S. (2013). *in Education Systems*. *317*(April), 317–320. https://doi.org/10.1126/science.1236180

32) Drost, E. A. (2004). *Validity and Reliability in Social Science Research*. *38*(1), 105–125.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

33) Etikan, I., Musa, S. A., & Alkassim, R. S. (2017). *Comparison of Convenience Sampling and Purposive Sampling Comparison of Convenience Sampling and Purposive Sampling*. *January 2016*. https://doi.org/10.11648/j.ajtas.20160501.11

34) Facione, P. A. (1990). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations. *Eric*.

35) Faryadi, Q. (2012). Effective Teaching and Effective Learning : Instructional Design Perspective. *International Journal of Engineering Research and Applications*, *2*(1), 222–228.

36) Fokides, E. (2018). Digital educational games and mathematics. Results of a case study in primary school settings. *Education and Information Technologies*, *23*(2), 851–867. https://doi.org/10.1007/s10639-017-9639-5

37) Forza, C. (2002). Survey Research in Operations Management: A process-Based Perspective. *International Journal of Operations and Production Management*, *22*(2), 152–194. https://doi.org/10.1108/01443570210414310

38) Freitas, S., Prieto, G., Simões, M. R., & Santana, I. (2014). *Psychometric properties of the Montreal Cognitive Assessment (MoCA): An analysis using the Rasch model*. 1–37.

39) Gall, M. D. (1942). *Educational Research : An Introduction / Meredith D. Gall, Joyce P. Gall, Walter R. Borg.* (Eighth Edi). Boston : Pearson/Allyn & Bacon, [2007].

40) Gooch, D. L. (2012). *Research, Development, and Validation of a School Leader's Resource Guide for the Facilitation of Social Media Use by School Staff*.

41) Greer, B. (1997). *MODELLING REALITY IN MATHEMATICS CLASSROOMS : THE CASE OF WORD PROBLEMS*. *7*(4).

42) Guilleux, A., Blanchin, M., & Hardouin, J. (2014). *Power and Sample Size Determination in the Rasch Model : Evaluation of the Robustness of a Numerical Method to Non-Normality of the Latent Trait*. *9*(1), 1–7. https://doi.org/10.1371/journal.pone.0083652

43) Güner, P., & Erbay, H. N. (2021). Metacognitive Skills and Problem-Solving. *International Journal of Research in Education and Science*, 715–734. https://doi.org/10.46328/ijres.1594

44) Hadi, S., & Novaliyosi. (2019). TIMSS INDONESIA (Trends in International Mathematics and Science Study). *Prosiding Seminar Nasional & Call For Papers*, 562–569. https://doi.org/10.4135/9781412957403.n438

45) Hambleton, R. K., Swaminathan, H., Rogers, D. J., & Hambleton, R. K. (1991). *Fundamentals of Item Response Theory Library of Congress Cataloging-in-Publication Data*.

46) Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

47) Helmi, S. A., Mohd-Yusof, K., & Phang, F. A. (2016). Enhancement of team-based problem solving skills in engineering students through cooperative problem-based learning. *International Journal of Engineering Education*, *32*(6), 2401–2414.

48) Hess, A. K. N., & Greer, K. (2016). Designing For Engagement: Using The ADDIE Model To Integrate High-Impact Practices Into An Online Information Literacy Course. *Communications in Information Literacy*, *10*(2), 264–282.

49) Ho, T. M. P. (2020). Measuring Conceptual Understanding, Procedural Fluency and Integrating Procedural and Conceptual Knowledge in Mathematical Problem Solving. *International Journal of Scientific Research and Management*, *8*(05), 1334–1350. https://doi.org/10.18535/ijsrm/v8i05.el02

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

50) Holmefur, M., Sundberg, K., Wettergren, L., & Langius-Eklöf, A. (2015). Measurement Properties of The 13-Item Sense of Coherence Scale Using Rasch Analysis. *Quality of Life Research*, *24*(6), 1455–1463. https://doi.org/10.1007/s11136-014-0866-6

51) Huitt, W. (2011). Bloom et al.'s taxonomy of the cognitive domain. *Educational Psychology Interactive*.

52) Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of Two- and Three-Parameter Logistic Item Characteristic Curves: A Monte Carlo Study. *Applied Psychological Measurement*, *6*(3), 249–260. https://doi.org/10.1177/014662168200600301

53) Irvine, J. (2017). A Comparison of Revised Bloom and Marzano' s New Taxonomy of Learning. *Research in Higher Education Journal*, *33*(08), 16.

54) Istiyono, E., Sunu, W., Dwandaru, B., & Megawati, I. (2018). Application of Bloomian and Marzanoian Higher Order Thinking Skills in the Physics Learning Assessment : an Inevitability. *International Conference on Learning Innovation*, *164*(Icli 2017), 136–142.

55) Janssen, K. C., Phillipson, S., O'Connor, J., & Johns, M. W. (2017). Validation of the Epworth Sleepiness Scale for Children and Adolescents using Rasch analysis. *Sleep Medicine*, *33*, 30–35. https://doi.org/10.1016/j.sleep.2017.01.014

56) 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

57) Kaspersen, E., Pepin, B., & Sikko, S. A. (2017). Measuring Student Teachers' Practices and Beliefs About Teaching Mathematics Using The Rasch Model. *International Journal of Research and Method in Education*, *40*(4), 421–442. https://doi.org/10.1080/1743727X.2016.1152468

58) Kazerooni, E. A. (2001). *Fundamentals of Clinical Research for Radiologists*. *November*, 993–999.

59) Kenedi, A. K., Helsa, Y., Ariani, Y., Zainil, M., & Hendri, S. (2019). Mathematical Connection Of Elementary School Students To Solve Mathematical Problems. *Journal on Mathematics Education*, *10*(1), 69–80.

60) Koller, I., Maier, M. J., & Hatzinger, R. (2015). *An Empirical Power Analysis of Quasi-Exact Tests for the Rasch Model Measurement Invariance in Small Samples*. https://doi.org/10.1027/1614-2241/a000090

61) Kumar, R. (2012). Research Methodology a step-by-step guide for beginners. In *עלון הנוטע* (Vol. 66).

62) Lambert, M. C., Garcia, A. G., January, S. A. A., & Epstein, M. H. (2018). The impact of English language learner status on screening for emotional and behavioral disorders: A differential item functioning (DIF) study. *Psychology in the Schools*, *55*(3), 229–239. https://doi.org/10.1002/pits.22103

63) Lawhon, D. (1976). Instructional development for training teachers of exceptional children: A sourcebook. *Journal of School Psychology*, *14*(1), 75. https://doi.org/10.1016/0022-4405(76)90066-2

64) *Learning Disabilities: A Multidisciplinary Journal, 24 (2), 1–14.* (2019).

65) Linacre, J. M. (1994). Sample Size and Item Calibration or Person Measure Stability. *Rasch Measurement Transactions*, *7*(4), 328.

66) Lub, V. (2015). Validity in Qualitative Evaluation. *International Journal of Qualitative Methods*, *14*(5), 160940691562140. https://doi.org/10.1177/1609406915621406

67) Md, M. R. (2019). 21st Century Skill "Problem Solving": Defining the Concept. *Asian Journal of Interdisciplinary Research*, *2*(1), 64–74. https://doi.org/10.34256/ajir1917

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

68) Ministry of Trade & Industry Republic Singapore. (2016). *The Impact of Research and Development on Productivity* (pp. 40–49).

69) Mohamad, M. M., Sulaiman, N. L., Sern, L. C., Mohd, K., & Salleh. (2015). *Measuring the Validity and Reliability of Research Instruments*.

70) Moradmand, N. (2014). *A Computer-Assisted Framework Based On The Bloom-Anderson Taxonomy For Teaching*. *April*.

71) Muruganantham, G. (2015). *Developing of E-Content Package By Using ADDIE Model*. *1*, 52–54.

72) Mutlu, Y. (2019). *Math Anxiety in Students With and Without Math Learning Difficulties*. *11*(5), 471–475. https://doi.org/10.26822/iejee.2019553343

73) Nizam. (2016). Ringkasan Hasil-hasil Asesmen Belajar dari Hasil UN, PISA, TIMSS, INAP/AKSI. In *Puspendik* (p. 94).

74) Noble, H., & Smith, J. (2015). Issues Of Validity and Reliability In Qualitative Research. *Evidence-Based Nursing*, *18*(2), 34–35. https://doi.org/10.1136/eb-2015-102054

75) OECD. (2015a). Concepts and Definitions for Identifying R&D. *The Measurement of Scientific, Technological and Innovation Activities*, 43–79. https://doi.org/10.1787/9789264239012-4-en

76) OECD. (2015b). *PISA 2015 Result in Focus*. https://doi.org/10.1596/28293

77) Osman, S., Nurul, C., Che, A., Abu, M. S., & Ismail, N. (2018). *Enhancing Students 'Mathematical Problem-Solving Skills through Bar Model Visualisation Technique*. *13*(3), 273–279.

78) Peng, A., Cao, L., & Yu, B. (2020). *Reciprocal Learning in Mathematics Problem Posing and Problem Solving : An Interactive Study between Canadian and Chinese Elementary School Students*. *16*(12).

79) Permana, R. H. (2019). *Survei Kualitas Pendidikan PISA 2018: RI Sepuluh Besar dari Bawah*. 1.

80) Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM - Mathematics Education*, *52*(1), 33–44. https://doi.org/10.1007/s11858-019-01118-9

81) Prastyo, H. (2020). Kemampuan Matematika Siswa Indonesia Berdasarkan TIMSS. *Jurnal Padegogik*, *3*(2), 111–117. https://doi.org/10.35974/jpd.v3i2.2367

82) Programme, T., Assessment, I. S., & Tables, I. (2018). *Indonesia What 15-year-old Students in Indonesia Know and Can Do Figure 1. Snapshot of Performance in Reading , Mathematics and Science*. 1–10.

83) R.Lynn, M. (1986). Determination and Quantification of Content Validity. *Nursing Research*, *35*, 382–386. https://doi.org/10.1097/00006199-198611000-00017

84) Radmehr, F., & Drake, M. (2019). Revised Bloom's Taxonomy and Major Theories and Frameworks That Influence the Teaching, Learning, and Assessment of Mathematics: A Comparison. *International Journal of Mathematical Education in Science and Technology*, *50*(6), 895–920. https://doi.org/10.1080/0020739X.2018.1549336

85) Rasch, G. (1960). *Probabilistic Models For Some Intelligence And Attainment Tests.*

86) 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

87) Retnawati, H. (2017). *Validitas Reliabilitas dan Karakteristik Butir* (2nd ed.). Parama Publishing.

88) Richey, R. C., & Klein, J. D. (2008). *Research on Design and Development* (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

89) Richey, R. C., & Nelson, W. (1996). *Developmental Research*. Simon & Schuster.

90) Richey, R. C., Klein, J. D., & Nelson, W. (2004). *Developmental Research: Studies of Instructional Design and Development* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

91) Richey, R. C. (2005). *Validating Instructional Design Models* (I. J. M. S. & D. A. Wiley (Ed.)). Mahwah, NJ; Lawrence Erlbaum Associates.

92) Richey, R. C. J. D. K. (2007). *Design and Developtnent Research* (A. Messina (Ed.)). Lawrence Erlbaum Associates, Inc. Routledge Taylor & Francis Group.

93) Rohmah, M., & Sutiarso, S. (2018). Analysis problem solving in mathematical using theory Newman. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(2), 671–681. https://doi.org/10.12973/ejmste/80630

94) Russo, J., & Minas, M. (2020). *Student Attitudes Towards Learning Mathematics Through Challenging , Problem Solving Tasks : " It 's so Hard – in a Good Way "*. *13*(2), 215–225.

95) Said, R. F. M. (2016). Application of Rasch Measurement Model in Evaluating Student Performance for Foundation of Computing II. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (Issue InCULT). https://doi.org/10.1007/978-981-287-664-5_58

96) Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. *International Electronic Journal of Mathematics Education*, *14*(3), 535–546. https://doi.org/10.29333/iejme/5755

97) Salihu, L., & Räsänen, P. (2018). *Mathematics Skills of Kosovar Primary School Children : A Special View on Children with Mathematical Learning Difficulties* *. *10*(4), 421–430. https://doi.org/10.26822/iejee.2018438132

98) Santun Naga, D. (2012). *Teori Sekor Pada Pengukuran Mental* (2nd ed.). PT Nagarani Citrayasa.

99) Sharifah, R., & Faaizah, S. (2015). The Development of Online Project Based Collaborative Learning using ADDIE Model. *Procedia - Social and Behavioral Sciences*, *195*, 1803–1812. https://doi.org/10.1016/j.sbspro.2015.06.392

100) Stenner, A. J., Fisher, W. P., Stone, M. H., & Burdick, D. S. (2013). Causal Rasch models. *Frontiers in Psychology*, *4*(AUG), 1–14. https://doi.org/10.3389/fpsyg.2013.00536

101) Sumintoro, Bambang, Widhiarso, W. (2013). *Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial* (B. Sumintoro (Ed.); II).

102) Susetyo, B. (2015). *Prosedur Penyusunan dan Analisis Tes*. PT Refika Aditama.

103) Tabatabaee-Yazdi, M. et. al. (2018). Development and Validation of a Teacher Success Questionnaire Using the Rasch Model. *International Journal of Instruction*, *11*(2), 129–144. https://doi.org/10.12973/iji.2018.11210a

104) Tachie, S. A. (2019). *Meta-cognitive Skills and Strategies Application : How this Helps Learners in Mathematics Problem-solving*. *15*(5).

105) Taherdoost, H. (2018). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*, *January 2016*. https://doi.org/10.2139/ssrn.3205040

106) Tambunan, H. (2019). The Effectiveness of the Problem Solving Strategy and the Scientific Approach to Students ' Mathematical Capabilities in High Order Thinking Skills. *International Electronic Journal of Mathematics Education*, *14*(2), 293–302.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

107) Tarim, K., & Bilgisi, M. (2010). *Problem Solving Levels Of Elementary School Students On Mathematical Word Problems And The Distribution of These Problems in Textbooks İlkokul Öğrencilerinin Matematiksel Sözel Problemleri Çözme Düzeyleri ve Bu Problemlerin Ders Kitaplarındaki Dağılımları*. *1997*, 639–648. https://doi.org/10.14812/cuefd.306025

108) Taylor, L. (2004). Educational Theories and Instructional Design Models . Their Place in Simulation . *Health*. https://doi.org/10.1177/026327695012004006

109) Tösten, R., Han, B., & Anik, S. (2017). The Impact of Parental Attitudes on Problem Solving Skills in High School Students. *Universal Journal of Educational Research*, *5*(1), 170–174. https://doi.org/10.13189/ujer.2017.050121

110) Uli Sihombing, Rosita; S. Naga, Dali; Rahayu, W. (2019). *A Rasch Model Measurement Analysis On Science Literacy Test Of Indonesian Student : Smart Way To Improve Learning Assessment*. *6*(1), 44–55.

111) Urbina, A. A. dan S. (2007). *Tes Psikologi*. PT Indeks.

112) Van Zile-Tamsen, C. (2017). Using Rasch Analysis to Inform Rating Scale Development. *Research in Higher Education*, *58*(8), 922–933. https://doi.org/10.1007/s11162-017-9448-0

113) VanderStoep, Scott W; Johnston, D. D. (n.d.). *Research Methods For Everyaday Life*.

114) Verhelst, N. D., & Verstralen, H. H. F. M. (2008). Some Considerations on The Partial Credit Model. *Psicologica*, *29*(2), 229–254.

115) Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM - Mathematics Education*, *52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4

116) Wainer, Howard; I.Braun, H. (2009). *Test Validity* (II). 365 Broadway Hillsdale, New Jersey.

117) Waltz, C. F., & Bausell, B. R. (1981). *Nursing Research: Design Statistics and Computer Analysis*. Davis FA.

118) Wei, S., Liu, X., & Jia, Y. (2014). Using Rasch Measurement to Validate the Instrument of Students' Understanding of Models in Science (SUMS). *International Journal of Science and Mathematics Education*, *12*(5), 1067–1082. https://doi.org/10.1007/s10763-013-9459-z

119) Wright, B. D., & Stone, M. H. (1979). *Best Test Design*.

120) Yang, Y., He, P., & Liu, X. (2016). Validation of an Instrument for Measuring Students' Understanding of Interdisciplinary Science in Grades 4-8 over Multiple Semesters: A Rasch Measurement Study. *NARST*, 1–20.

121) Zulyadaini, D. (2017). A Development of Students' Worksheet Based on Contextual Teaching and Learning. *IOSR Journal of Mathematics*, *13*(01), 30–38. https://doi.org/10.9790/5728-1301033038

122) Weible, Christopher M (2017). Instrument Constituencies and the Advocacy Coalition Framework: an essay on the comparisons, opportunities, and intersections. Policy and Society, 37(1), 59-73, ISSN 1449-4035, Oxford University Press (OUP), https://doi.org/10.1080/14494035.2018.1417705

123) Pomalato et al. (2021). Instrument Test Development Of Mathematics Skill On Elementary School. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(6), 2447-2455, ISSN 1309-4653, Auricle Technologies, Pvt., Ltd., https://doi.org/10.17762/turcomat.v12i6.5688

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.11108908

124) Jameela, Ingkha, & Destania, Yuriska (2020). Developing Essay Questions on Prism and Pyramid for the Ability to Understand Mathematical Concept. Hipotenusa : Journal of Mathematical Society, 2(2), 83-97, ISSN 2716-3156, IAIN Salatiga, https://doi.org/10.18326/hipotenusa.v2i2.83-97

125) Wafa, Mohammad Nasim (2019). Assessing School Students' Mathematic Ability Using DINA and DINO Models. International Journal of Mathematics Trends and Technology, 65(12), 153-165, ISSN 2231-5373, Seventh Sense Research Group Journals, https://doi.org/10.14445/22315373/ijmtt-v65i12p517

126) Waluyo,S., Armanto, D., Mansyur, A. (2021). Analysis of Mathematic Reasoning Ability and Self-Confidence Students Using Models PBL Reviewed from Intelligence Intrapersonal in SMP N 5 Youth Vocationality. Journal of Education and Practice, International Institute for Science, Technology and Education, https://doi.org/10.7176/jep/12-16-08

127) Hasbi, Muhammad, Lukito, Agung, & Sulaiman, Raden (2019). The Realistic of Mathematic Educational Approach to Enhancing Ability Mathematical Connections. International Journal of Trends in Mathematics Education Research, 2(4), 179-183, ISSN 2621-8488, SAINTIS Publishing, https://doi.org/10.33122/ijtmer.v2i4.82

128) Sinuraya, Motla, & Mihardi (2021). Need Analysis of TPACK Oriented of ICARE Based Student Worksheet for Improving Students' Ability to Complete HOTS Questions. Proceedings of the 3rd International Conference on Innovation in Education, Science and Culture, ICIESC 2021, 31 August 2021, Medan, North Sumatera Province, Indonesia, EAI, https://doi.org/10.4108/eai.31-8-2021.2313804

129) Siagan, MV, Saragih, S, & Sinaga, B (2019). Development of Learning Materials Oriented on Problem-Based Learning Model to Improve Students' Mathematical Problem Solving Ability and Metacognition Ability.. International electronic journal of …, ERIC, https://eric.ed.gov/?id=EJ1227196

130) Can, Derya (2020). The mediator effect of reading comprehension in the relationship between logical reasoning and word problem solving. Participatory Educational Research, dergipark.org.tr, https://dergipark.org.tr/en/pub/per/article/726691

131) Spencer, M., Fuchs, L. S. & Fuchs, D. (2020). Language-related longitudinal predictors of arithmetic word problem solving: A structural equation modeling approach. Contemporary Educational Psychology, 60, 1-16.

132) Ulu, M (2017). Errors made by elementary fourth grade students when modelling word problems and the elimination of those errors through scaffolding. International Electronic Journal of Elementary …, iejee.com, https://www.iejee.com/index.php/iejee/article/view/176

133) Unaenah, Een & Zamroni, Moh. (2022). Analisis Kesulitan Belajar Matematika dalam Menyelessaykan Soal Cerita Siswa Kelas V SDN Karawaci 1. TSAQOFAH. 2. 529-540. 10.58578/tsaqofah.v2i5.536.