International Journal of Trends in Mathematics Education Research
Vol. 6, No. 3 (2023), pp. 269~283
ISSN 2621-8488 (online)
DOI: https://doi.org/10.33122/ijtmer.v6i3.240

SAINTIS publishing

Research Article

Development of an instrument to measure student's self-efficacy level in mathematics learning

Wildan Nugraha[1] & Khoerul Umam[2]*
[1] Universitas Muhammadiyah Prof. Dr. Hamka Jakarta, Indonesia, 60231
[2] Postgraduate of the Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia, 60231
*Corresponding Author: khoerul.umam@uhamka.ac.id

Received: 28 July 2023 *Revised: 29 August 2023* *Accepted: 20 September 2023* *Available online: 30 September 2023*

ABSTRACT

The purpose of this research is to measure the level of students' self-efficacy by developing a mathematical self-efficacy instrument and to determine the quality of the developed instrument. This research is categorized as Research and Development (R&D). The resulting product of the research is a questionnaire consisting of 28 statements. The development of the mathematical self-efficacy instrument was carried out using a formative research development model. The study was conducted in 11th-grade classes at SMA 90 in South Jakarta. The findings of this research indicate that the developed instrument is valid, reliable, and of reasonably good quality, as it passed the instrument testing stages. Additionally, this research also demonstrates the criteria for students' self-efficacy level using the developed instrument.

Keywords: Development Research; Instruments; Self-Efficacy; Mathematics Teaching; Mathematics Learning

1. INTRODUCTION

According to (Dewi, 2018), education can be defined as a process of preparing individuals to adapt and survive in their environment (life skills). It serves as a crucial guide in life that every individual must possess and master in order to enhance their quality and abilities. The quality of human resources, driven by self-empowerment, becomes a benchmark for a country's educational quality and helps in developing individuals' capabilities through education (Maskar & Dewi, 2021). In addition, education also helps individuals to keep up with all the existing developments of the era (Puspaningtyas, 2019). Based on research conducted by (Anderha & Maskar, 2020), it is stated that education is an essential aspect that every individual must have in their life because it guides each individual to act and think according to their life goals. This is also reflected in Law No. 20 of 2003 concerning the National Education System, where self-efficacy is crucial for the nation and the country (Sistem Pendidikan Nasional, 2003).

The indicator of success in education for an individual can be seen from their academic achievement (Afniola et al., 2020). According to (Slameto, 2012), an individual's academic achievement can be influenced by psychological, physical, and environmental factors. One of the psychological factors that affect an individual's academic achievement is self-efficacy.. According to (Bandura, 1997), self-efficacy can be defined as an individual's level of confidence in their abilities to achieve the goals they desire in life. Self-efficacy has a significant influence on an individual's academic achievement, as it instills in them a high level of curiosity, perseverance, resilience, and self-confidence in problem solving. Self-efficacy can be developed in high school students or adolescents aged 16 to 18 years, who are becoming more independent and seeking their own identity during this period. According John M. Ortiz in (Tanjung & Amelia, 2017), self-confidence is believing in one's own abilities and being able to rely on oneself. Students need a sense of self-efficacy to meet the demands of an increasingly sophisticated era. According to Ormrod in (Hairida, 2017), if someone believes they can perform certain actions successfully, it means they have a high level of self-efficacy.

Based on the above, self-efficacy in students needs to be attended to by educators so that their potential can be developed optimally. In the world of education, students with high self-efficacy tend to have high learning motivation and a strong enthusiasm for acquiring knowledge. They are more capable of adapting to difficulties in completing tasks assigned by teachers. This is in line with the opinion of (Schunk, 1990), who stated that students with high self-efficacy can find solutions or a way out when facing challenges or obstacles in the learning process, making learning more effective. When discussing the subject of mathematics, self-efficacy is crucial in the learning process. In the study conducted by (Husna, 2016), the development of self-efficacy in mathematics is essential to be nurtured in order to instill an appreciation for the vital role of mathematics in life, in accordance with the National Education Minister Regulation No. 54 of the Year. Self-efficacy in mathematics learning needs to be given attention because it can influence the continuity of the mathematics learning process (Shadiq, 2007). Additionally, self-efficacy is also important in reducing students' stress levels towards mathematics, which is often seen as a challenging subject for many students (Supardi US, 2010).

In the field of education, measuring the level of self-efficacy is an important activity to assess someone's potential or abilities based on their learning outcomes. This measurement is conducted using measurement tools or instruments. Research instruments serve as measurement tools that aid in data collection and have specific criteria or qualifications (Djaali & Mulyono, 2000). In educational research, these instruments are used to measure students' learning achievements and the factors and indicators influencing the learning process. Research instruments can be divided into two techniques: test instruments and non-test instruments.

The previous research conducted by (Ahriana et al., 2016) regarding self-efficacy in mathematics learning indicated that the level of self-efficacy among students was still low. This finding was supported by the results of interviews conducted with the students. When faced with mathematics problems of medium to high difficulty levels, students tended to give up on solving them and preferred to ask their friends who could solve the problems. The desire to be able to solve the problems independently was still very low. According to (Hendriana & Kadarisma, 2019), self-efficacy has a positive influence on students' communication abilities. This means that the higher the self-efficacy of the students, the higher their communication skills will be. Furthermore, the correlation coefficient value falls into the category of very strong, indicating a robust relationship between self-efficacy and students' communication abilities.

Based on the expert's opinion above, it is crucial for researchers to develop a self-efficacy instrument that is proven to be valid, reliable, and practical. This will provide a clear understanding of students' self-efficacy levels in mathematics learning through the instrument administered to the students. Based on the issues outlined, the objectives of this research are as follows; 1) To develop a self-efficacy instrument for mathematics learning; 2) To assess the quality of the developed mathematical self-efficacy instrument; 3) To measure the level of students' self-efficacy in mathematics learning using the developed instruments.

2. RESEARCH METHOD

This research is classified as Research and Development (R&D). According to (Sugiyono, 2017), the R&D method is used to create, produce, and test the effectiveness of a product. The development model utilized in this research is the formative research model (Tessmer, 1993). There are four stages involved in the formative research development model: the preliminary stage, self-evaluation stage, prototyping stage, and field test stage. The following is the flow of formative research steps design.

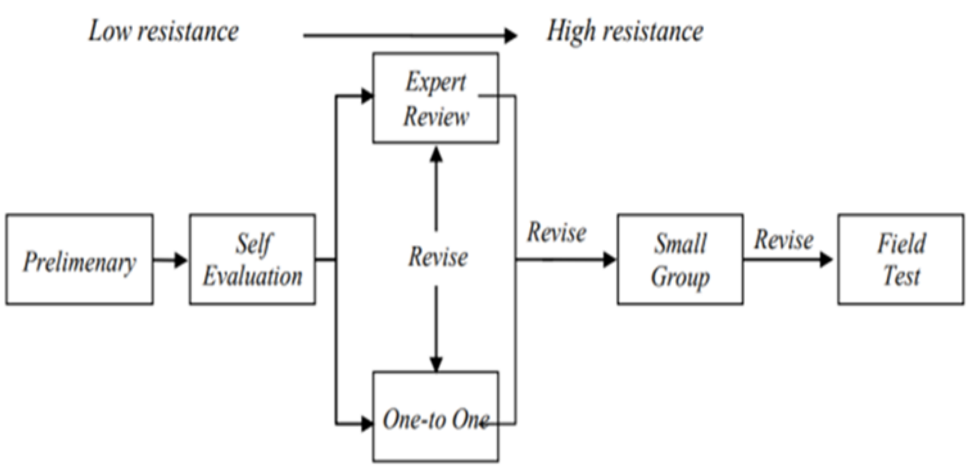

Figure 1. Flow of Formative Research Steps Design

The researchers conducted the development of an instrument to measure the level of self-efficacy among respondents, who are 11th-grade students at SMA 90 in South Jakarta. The sampling technique used in the study was Cluster Random Sampling, resulting in the selection of class XI IPA-4 with a total of 36 student respondents.

The developed instrument is a non-test research instrument in the form of a questionnaire. The questionnaire consists of 28 statements regarding the review of mathematics learning and is based on self-efficacy indicators. A Likert scale with a range of 1 to 5 is used to fill in the statements in the questionnaire, which include both positive and negative statements. The Likert scale criteria for positive statements in the questionnaire are as follows: 1) Strongly Disagree; 2) Disagree; 3) Agree; 4) Strongly Agree; 5) Very Agree (Muslimah, 2013). He same applies to negative statements in the questionnaire. The questionnaire or survey used in this research is a closed-ended questionnaire, where respondents are required to select only one correct answer. This questionnaire is used to obtain data on students' self-efficacy, allowing the researcher to analyze the data in accordance with the research objectives.

The data analysis technique used in this research is the Rasch model (quantitative research) to describe the designed learning instrument and assess the instrument's appropriateness after implementation in the final product. The Rasch model is also employed to test the validity of the items, reliability, difficulty level, and test discrimination. In determining the instrument's quality, the analysis process takes into account the difficulty level and test discrimination (Wahyuningsih, 2015). Data is analyzed using Microsoft Excel and the Rasch model with Winstep software. To test the validity of an item, you can refer to the output from the Winstep software, specifically the table item fit order. The table item fit order is used to determine whether the questionnaire items used are appropriate or not. If the instrument does not meet one of the criteria in **Table 1**, the instrument can still be considered valid (Sumintono & Widhiarso, 2015).

Table 1. Criteria of Validity Test

Values	Information
Outfit Mean Square (MNSQ)	0,5 < MNSQ < 1,5
Outfit Z-Standart (ZSTD)	-2,0 < ZSTD < +2,0
Point Measure Correlation (Pt Mean Corr)	0,4 < Pt Mean Corr < 0,85

Furthermore, to test the reliability of an item, you can refer to the output from the Winstep software, particularly the table summary statistics. In the table summary statistics, there is also Cronbach's alpha to assess the reliability of the items and persons (Sumintono & Widhiarso, 2015). In the reliability test, you can use the following formula to calculate Alpha: [Alpha formula].

$$r_{11} = \left[\frac{n}{n-1}\right]\left[1 - \frac{\sum_{\sigma_i} 2}{{\sigma_i}^2}\right]$$

Information:

r_{11} : Instrument Reliability
$\sum_{\sigma_i} 2$: The sum of varians score test
${\sigma_i}^2$: Total of variances
n : The sum of item

Next, the following table presents the criteria for Cronbach's Alpha (KR-20), Item and Person Reliability, and Item and Person Separation as follows:

Table 2. Reliability in Rasch Analysis

Statistics	Fit Indices	Interpretation
Cronbach's Alpha (KR-20)	< 0.5	Very bad
	0.5 – 0.6	Bad
	0.6 – 0.7	Enough
	0.7 – 0.8	Good
	> 0.8	Very Good
Item and Person Reliability	< 0.67	Weak
	0.67 – 0.80	Enough
	0.81 – 0.90	Good
	0.91 – 0.94	Very Good
	> 0.94	Special
Item and Person Separation		The higher the separation value, it can be said that the quality of the instrument is better. Besides that, groups of item and person can be identified.

Furthermore, the level of difficulty can also be observed in the output of the Winstep software, specifically in the table item measure. In the table item measure, a higher measure value indicates a higher level of difficulty for the item (Sumintono & Widhiarso, 2015). The difficulty level of questionnaire items can be seen through the Wright Person Item Map. The criteria for the difficulty level of questionnaire items using the Rasch model can be seen in **Table 3.**

Table 3. Criteria of Hardness Level Item

Measured Value (logit)	Interpretation
$\boldsymbol{Measure\ logit < -SD\ logit}$	Easy
$\boldsymbol{-SD\ logit \leq Measure\ logit \leq 0}$	Medium
$\boldsymbol{0 \leq Measure\ logit \leq SD\ logit}$	Hard
$\boldsymbol{Measure\ logit > SD\ logit}$	Very Hard

(Handayani & Iba, 2020) stated that the item discrimination for each item can be observed through the item's bias value in the questionnaire. In determining item discrimination, the separation value is needed, which functions to classify persons and items into certain categories.

$$H = \frac{[(4\ \times Separation) + 1]}{3}$$

The higher the item separation value, the higher the quality of the instrument, and the quality of the questionnaire items or statements also improves. This is because a higher item separation value indicates that the instrument can effectively distinguish among respondents as a whole (Sumintono & Widhiarso, 2015). After going through several stages of development and testing the instrument, it is essential to assess the quality of the non-test instrument, which is the questionnaire, to determine the extent of its development process. The questionnaire items that have been developed in accordance with the criteria of the instrument testing are then arranged as follows.

1. If the questionnaire items have the minimum validity level, then the validity criteria are considered fulfilled;
2. If the questionnaire items have the minimum reliability level, then the reliability criteria are considered met;
3. The difficulty level criteria of questionnaire items are considered good if the items can determine the level of difficulty among the group of items;
4. The discrimination criteria are considered good if the instrument can classify students.

Based on the research objectives, the main goal of this study is to measure the level of self-efficacy among students in mathematics learning using the developed instrument. After obtaining the self-efficacy data, a descriptive analysis will be conducted by creating a frequency distribution table to depict the frequency of each variable and categorize them into very high, high, moderate, low, and very low categories. The assessment criteria for each data will refer to the limits proposed by (Husaini Usman, 2017).

Table 4. Criteria of Mathematics Self-Efficacy Level of Students

Score	Criteria
$X \geq (M + 1{,}5\ Sdi)$	Very High
$(M + 0{,}5\ Sdi) < X < (M + 1{,}5\ Sdi)$	High
$(M - 0{,}5\ Sdi) < X < (M + 0{,}5\ Sdi)$	Medium
$(M - 1{,}5\ Sdi) < X < (M - 0{,}5\ Sdi)$	Low
$X \leq (M - 1{,}5\ Sdi)$	Very Low

3. RESULTS AND DISCUSSION

Description of Instrument Development Stage

The instrument development is carried out using the formative research development model, which consists of four stages: preliminary stage, self-evaluation stage, prototyping stage, and field test stage (Tessmer, 1993). Here is a detailed flow of the formative research stages:

1. Description of Preliminary Stage

In the preliminary stage, a thorough review of research literature is conducted. Once all the literature is gathered, the location and research subjects are determined by coordinating with the targeted school and specifically focusing on mathematics teachers since this research is about students' self-efficacy in mathematics learning. Based on observation and analysis, this research is conducted at SMA Negeri 90 Jakarta with the research subjects being 11th-grade students. The 11th-grade class is selected using the cluster random sampling technique, resulting in the selection of class XI IPA-4 with a total of 36 students.

2. Description of Self-Evaluation Stage

The self-evaluation stage aims to develop an instrument that can measure the level of mathematical self-efficacy. In this stage, analysis and design are conducted regarding the non-test research instrument, including:

a) Curriculum analysis

In the curriculum analysis stage, it is carried out to align the objectives in the development of the non-test instrument, so that it can measure the level of students' mathematical self-efficacy. The curriculum under review in this study is the 2013 curriculum.

b) Students Analysis

The student analysis stage focuses on 11th-grade students who have undergone comprehensive mathematics learning. Based on the problem analysis, it is found that the 11th-grade students at SMA 90 in South Jakarta have varying levels of self-efficacy. Some students have low self-efficacy, some have moderate self-efficacy, and some have high self-efficacy. This variation in self-efficacy levels is possible due to the individual differences in students' interest in mathematics learning.

c) Design

In the design stage, the objective is to develop a non-test instrument capable of measuring students' mathematical self-efficacy. The initial step involves designing mathematical self-efficacy questionnaire items, with a total of 28 items having been designed. These questionnaire items cover various indicators of mathematical self-efficacy. This design is referred to as the initial prototype.

3. Description of Prototype Stage

The Prototyping stage aims to evaluate the designed instrument. The evaluation process in the Prototyping stage consists of three processes: expert review, one-to-one, and small group evaluations. The purpose of these processes is to identify and rectify errors and shortcomings in the initial prototype. The revisions made during the evaluation process are referred to as Prototype I.

a) Expert Review

Expert Review is a form of validation performed by experts on the created instrument. The results of this review serve as the basis for revising and improving the instrument to be developed. Validators assess the designed instrument from two aspects related to filling out the Likert scale using the provided validation sheet. The Likert scale rating criteria for the validators are from 1 to 4, where a score of 1 indicates "strongly disagree," a score of 2 indicates "disagree," a score of 3 indicates "agree," and a score of 4 indicates "strongly agree." In this study, the validators consist of one Mathematics Faculty of Education lecturer (Validator 1) and one mathematics subject teacher from SMA Negeri 90 in South Jakarta (Validator 2). Based on the expert assessment of the created instrument, quantitative calculations can be performed using the following formula (Sari et al., 2020):

$$P = \frac{\sum x_i}{\sum x} \times 100\%$$

Information:

P : Percentage of assessment
$\sum x_i$: The total score obtained
$\sum x$: The total score altogether

Table 5. Percentage Criteria of Validator's Assessment

Percentage	Criteria
76% - 100%	Valid
56% - 75%	Quite Valid
40% - 55%	Less Valid
0% - 39%	Invalid

The self-efficacy instrument meets validity if the validators rate the average of all questionnaire items above the minimum percentage threshold of 56%, which qualifies as sufficiently valid or valid. If it does not meet these criteria, it needs to be revised again to obtain an instrument that meets the criteria of sufficiently valid or valid (Prototype I).

b. One-to-one

In addition to expert review, the instrument is also validated through the one-to-one process. In the one-to-one process, three non-subject students with various levels of self-efficacy are needed to provide feedback on the questionnaire items until completion. The comments given by the students are used to revise the developed instrument (Prototype II).

c. Small Group

The validation from expert review and one-to-one processes is used as the basis for revising the design to conduct validity through the small group process. The small group consists of nine non-subject students with different characteristics: three high-ability students, three moderate-ability students, and three low-ability students. Students are asked to fill out a questionnaire consisting of 28 statements related to mathematics learning reviews and referring to self-efficacy indicators using the Likert scale.

After obtaining the data from the small group stage, data analysis is conducted using the Rasch model with the assistance of Microsoft Excel and Winstep software to test the validity and reliability of the items. The output from the Winstep software used to test the validity and reliability of the items is the table item fit order. The instrument design analyzed in this stage is then revised and developed according to the criteria for testing the validity and reliability of the items (Prototype III) to be implemented in the field test stage. The criteria for testing the validity and reliability of the items can be seen in **Table 1** and **Table 2**.

4. Description of Field Test Stage

The revised and validated prototype is then tested on the research subjects, which are 36 students from class XI IPA-4 of SMA Negeri 90 in South Jakarta, as the field test stage. At the beginning of the field test phase, the researcher distributes a Google Form link as a medium for the students to fill out the questionnaire that has been developed in Prototype III using the Likert scale.

The results obtained from the students' responses in the field test stage aim to determine the quality of the non-test instrument by conducting validity tests, reliability tests, difficulty levels, item discrimination, and criteria for non-test instrument quality using Microsoft Excel and the Rasch model with Winstep software. Based on the research objectives, the results obtained in this stage are also used to measure the students' self-efficacy level in mathematics learning using the developed instrument.

Description of the Test Results

After going through several stages of design and development of the non-test instrument, the results from the instrument evaluation process and the field test stage are as follows:

1. Expert Review

In the expert review process, the validator is asked to assess the designed instrument by providing opinions on questionnaire items that need revision. The overall assessment by the validator can be seen in **Table 6**.

Table 6. Validator's Assessment

Validators	Assessments
Validator 1	The instrument is already quite good but needs to be improved to be suitable for use.
Validator 2	The instrument is already good enough and suitable for use.

After going through several revisions, the validator assessed the designed instrument from two aspects related to filling out the Likert scale through the provided validation sheet. The results of the validator's assessment can be seen in **Table 7**.

Table 7. The Result of Validator's Assessment

Validators	Aspects	Criteria	Assessment Scale			
			1	2	3	4
Validator 1	Contents	Non-test measurement tools or questionnaires can be used to measure the level of self-efficacy.				√
		In accordance with the definition of indicators.				√
	Languages	The sentence does not cause any ambiguity.				√
		Using language that conforms to the rules of good and proper grammar.				√
		The use of words appropriate for the students' school level.			√	
Validator 2	Contents	Non-test measurement tools or questionnaires can be used to measure the level of self-efficacy.				√
		In accordance with the definition of indicators.				√
	Languages	The sentence does not cause any ambiguity.				√
		Using language that conforms to the rules of good and proper grammar.				√
		The use of words appropriate for the students' school level.				√

Based on the validator's assessment, the percentage result of the validator's evaluation is obtained using the quantity calculation formula.

Table 8. Percentage of Validator Assessment's Result

Validators	Percentage of Assessments	Criteria
Validator 1	95%	Valid
Validator 2	100%	Valid

Based on the results, the percentage has exceeded the minimum percentage threshold of 56%, with scores of 95% and 100%. The validator's assessment for all aspects falls within the valid criteria, indicating that the instrument has achieved a valid level.

2. One-to-one

One-to-one was conducted by providing the non-test instrument in the form of a questionnaire to 3 students from class XI (non-subject). Below are the comments provided by the students regarding the given instrument.

Table 9. The Result of One-to-one Process

Students	Comments
1	The questionnaire has diverse statements.
2	The language is easy to understand, and sentence construction is clear enough.
3	The questionnaire is easily accessible as it uses a digital medium in the form of a Google Form.

The comments provided by the students are used to revise the prepared instrument, and then the instrument can proceed to the next process.

3. Small Group

In this process, the researcher analyzes the data using the Rasch model to test the validity and reliability of the questionnaire items with the help of Microsoft Excel and Winstep software. To test the validity of an item, it can be observed in the output of the Winstep software, which is presented in the table item fit order. The results of the validity test are analyzed according to the criteria, and the obtained results are as **Figure 2**.

In the above figure, items I19, I23, and I25 do not meet the validity test criteria in the table item fit order, as their Outfit Mean Square (MNSQ), Point Measure Correlation (Pt Mean Corr), and Outfit Z-Standard (ZSTD) values do not meet the criteria. However, items I9, I11, I8, I15, I14, I16, I5, I27, and I28 are considered valid because they only fail to meet one of the criteria. Therefore, items I19, I23, and I25 should be removed or further developed to meet the validity criteria.

After the items were revised by eliminating those that did not meet the validity test criteria, a second validity test was conducted with 25 remaining items. The results are as **Figure 3**.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
9	22	9	2.24	.47	1.89	1.9	2.06	2.0	A .58	.58	44.4	49.0	I9
11	37	9	-1.65	.53	1.88	1.8	1.65	1.3	B .50	.55	55.6	56.0	I11
19	20	9	2.68	.47	1.64	1.5	1.66	1.4	C .30	.60	44.4	45.4	I19
8	24	9	1.80	.48	1.61	1.3	1.59	1.2	D .55	.56	33.3	52.4	I8
10	38	9	-1.93	.54	1.60	1.3	1.37	.8	E .52	.52	33.3	57.2	I10
18	31	9	.00	.53	1.47	1.0	1.47	1.0	F .62	.60	66.7	59.2	I18
15	29	9	.55	.52	1.36	.8	1.36	.8	G .11	.58	33.3	62.6	I15
2	30	9	.28	.53	1.33	.8	1.35	.8	H .66	.59	44.4	61.8	I2
26	34	9	-.83	.52	1.29	.7	1.26	.7	I .42	.59	33.3	54.3	I26
13	30	9	.28	.53	1.17	.5	1.25	.6	J .56	.59	44.4	61.8	I13
17	34	9	-.83	.52	1.24	.6	1.24	.6	K .47	.59	77.8	54.3	I17
24	37	9	-1.65	.53	1.22	.6	1.09	.3	L .44	.55	66.7	56.0	I24
21	32	9	-.28	.53	1.11	.4	1.10	.4	M .63	.60	44.4	55.2	I21
1	31	9	.00	.53	1.06	.3	1.03	.2	N .62	.60	55.6	59.2	I1
12	34	9	-.83	.52	.95	.0	.89	-.1	n .62	.59	77.8	54.3	I12
14	32	9	-.28	.53	.83	-.2	.84	-.2	m .89	.60	44.4	55.2	I14
6	36	9	-1.37	.52	.80	-.3	.78	-.4	l .54	.57	55.6	56.7	I6
16	30	9	.28	.53	.74	-.4	.68	-.5	k .92	.59	66.7	61.8	I16
7	35	9	-1.10	.52	.72	-.5	.69	-.6	j .69	.58	66.7	55.8	I7
20	24	9	1.80	.48	.68	-.7	.61	-.8	i .49	.56	66.7	52.4	I20
4	33	9	-.56	.52	.63	-.7	.67	-.6	h .63	.60	55.6	55.2	I4
3	22	9	2.24	.47	.62	-.9	.58	-1.0	g .57	.58	55.6	49.0	I3
22	32	9	-.28	.53	.52	-1.0	.52	-1.0	f .57	.60	66.7	55.2	I22
5	30	9	.28	.53	.39	-1.4	.42	-1.2	e .62	.59	88.9	61.8	I5
27	32	9	-.28	.53	.37	-1.5	.36	-1.6	d .84	.60	66.7	55.2	I27
28	28	9	.82	.52	.30	-1.7	.30	-1.7	c .79	.57	88.9	61.8	I28
23	37	9	-1.65	.53	.21	-2.7	.22	-2.3	b .87	.55	88.9	56.0	I23
25	30	9	.28	.53	.20	-2.2	.20	-2.1	a .81	.59	88.9	61.8	I25
MEAN	30.9	9.0	.00	.52	.99	.0	.97	-.1			59.1	56.3	
S.D.	4.7	.0	1.24	.02	.49	1.2	.48	1.1			17.6	4.3	

Figure 2. **First Stage of Small Group's Fit Order Item**

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
9	22	9	2.24	.46	1.94	1.9	2.10	2.1	A .55	.58	33.3	45.7	I9
11	37	9	-1.55	.52	1.84	1.7	1.64	1.3	B .48	.55	55.6	55.7	I11
8	24	9	1.81	.47	1.60	1.3	1.62	1.3	C .53	.57	33.3	52.1	I8
10	38	9	-1.82	.53	1.57	1.3	1.34	.8	D .51	.53	33.3	55.4	I10
15	29	9	.59	.52	1.39	.8	1.39	.8	E .09	.59	33.3	62.2	I15
2	30	9	.32	.52	1.34	.8	1.36	.8	F .63	.60	44.4	61.4	I2
18	31	9	.05	.52	1.33	.8	1.30	.7	G .66	.60	55.6	57.6	I18
13	30	9	.32	.52	1.18	.5	1.28	.7	H .53	.60	44.4	61.4	I13
23	34	9	-.76	.51	1.22	.6	1.19	.5	I .44	.60	33.3	53.7	I23
17	34	9	-.76	.51	1.18	.5	1.20	.6	J .49	.60	77.8	53.7	I17
22	37	9	-1.55	.52	1.16	.5	1.01	.2	K .46	.55	55.6	55.7	I22
20	32	9	-.23	.52	1.11	.4	1.10	.4	L .61	.61	55.6	56.1	I20
1	31	9	.05	.52	1.02	.2	.99	.2	M .62	.60	66.7	57.6	I1
12	34	9	-.76	.51	.87	-.1	.81	-.3	l .65	.60	77.8	53.7	I12
14	32	9	-.23	.52	.76	-.4	.78	-.3	k .90	.61	55.6	56.1	I14
6	36	9	-1.28	.51	.73	-.6	.71	-.5	j .58	.57	55.6	56.2	I6
16	30	9	.32	.52	.66	-.6	.59	-.7	i .93	.60	66.7	61.4	I16
7	35	9	-1.02	.51	.66	-.7	.62	-.8	h .72	.59	66.7	55.3	I7
4	33	9	-.49	.52	.62	-.8	.65	-.7	g .64	.60	55.6	54.8	I4
3	22	9	2.24	.46	.62	-.9	.58	-1.0	f .57	.58	44.4	45.7	I3
19	24	9	1.81	.47	.59	-.9	.53	-1.1	e .54	.57	66.7	52.1	I19
21	32	9	-.23	.52	.48	-1.2	.49	-1.1	d .60	.61	55.6	56.1	I21
5	30	9	.32	.52	.39	-1.4	.42	-1.2	c .62	.60	88.9	61.4	I5
24	32	9	-.23	.52	.38	-1.5	.36	-1.5	b .83	.61	77.8	56.1	I24
25	28	9	.85	.51	.29	-1.8	.29	-1.7	a .79	.58	88.9	61.5	I25
MEAN	31.1	9.0	.00	.51	1.00	.0	.97	.0			56.9	56.0	
S.D.	4.4	.0	1.11	.02	.46	1.0	.45	1.0			16.7	4.2	

Figure 3. **Second Stage of Small Group's Fit Order Item**

In the **Figure 3**, item I9 does not meet the validity test criteria in the table item fit order, as the values of Outfit Mean Square (MNSQ), Point Measure Correlation (Pt Mean Corr), and Outfit Z-Standard (ZSTD) do not meet the criteria. However, items I11, I8, I15, I14, I16, I21, I5, I24, and I25 are considered valid as only one criterion does not meet the requirements. Therefore, item I9 needs to be removed or further developed to achieve validity criteria.

After revising the items by eliminating those that did not meet the validity test criteria, a second validity test was conducted with 24 items. The results are as **Figure 4**.

The image above shows the results of the item fit order for the third stage, where in the first and second stages, I23, I25, I19, and I9 were eliminated. Thus, there are 24 remaining items. In this third stage, items I8, I10, I14, I13, I15, I5, I23, and I24 are still considered valid because only one value does not meet the criteria. The validity test criteria in the table item fit order can be seen in Table 1. Therefore, all items in this third stage are considered valid.

After that, to test the reliability, it can be observed from the output of the Winstep software, which includes the value of the Cronbach's Alpha statistic (KR-20), Item and Person Reliability, and Item and Person Separation. The results obtained are as **Figure 5**. In the **Figure 5,** the results of the analysis are presented in the following table, showing the obtained values along with the criteria based on those values.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
8	24	9	2.03	.49	1.85	1.7	1.92	1.7	A .48	.57	22.2	55.3	I8
10	37	9	-1.53	.53	1.85	1.7	1.67	1.3	B .51	.57	55.6	56.7	I10
9	38	9	-1.81	.54	1.64	1.4	1.41	.9	C .51	.55	33.3	55.9	I9
2	30	9	.43	.54	1.45	.9	1.48	1.0	D .60	.60	44.4	62.7	I2
14	29	9	.71	.53	1.41	.9	1.40	.8	E .13	.59	33.3	63.7	I14
12	30	9	.43	.54	1.26	.6	1.36	.8	F .53	.60	44.4	62.7	I12
17	31	9	.14	.54	1.30	.7	1.26	.6	G .71	.61	66.7	60.1	I17
21	37	9	-1.53	.53	1.28	.8	1.12	.4	H .42	.57	55.6	56.7	I21
16	34	9	-.70	.53	1.22	.6	1.27	.7	I .49	.61	77.8	57.0	I16
19	32	9	-.15	.53	1.19	.5	1.19	.5	J .60	.61	44.4	54.9	I19
22	34	9	-.70	.53	1.19	.5	1.13	.4	K .50	.61	44.4	57.0	I22
1	31	9	.14	.54	1.05	.3	1.01	.2	L .63	.61	55.6	60.1	I1
11	34	9	-.70	.53	.86	-.2	.80	-.3	l .68	.61	66.7	57.0	I11
13	32	9	-.15	.53	.77	-.3	.81	-.2	k .91	.61	66.7	54.9	I13
6	36	9	-1.25	.52	.78	-.4	.79	-.3	j .56	.59	55.6	57.3	I6
7	35	9	-.98	.52	.71	-.6	.67	-.7	i .70	.60	55.6	57.8	I7
15	30	9	.43	.54	.71	-.5	.63	-.6	h .93	.60	66.7	62.7	I15
4	33	9	-.43	.53	.65	-.7	.67	-.6	g .64	.61	44.4	54.0	I4
3	22	9	2.50	.48	.66	-.8	.62	-.8	f .57	.59	33.3	49.7	I3
18	24	9	2.03	.49	.58	-.9	.52	-1.1	e .57	.57	77.8	55.3	I18
20	32	9	-.15	.53	.50	-1.1	.51	-1.0	d .60	.61	88.9	54.9	I20
5	30	9	.43	.54	.42	-1.3	.47	-1.1	c .59	.60	88.9	62.7	I5
23	32	9	-.15	.53	.42	-1.4	.39	-1.4	b .81	.61	66.7	54.9	I23
24	28	9	.99	.53	.30	-1.7	.29	-1.7	a .80	.58	88.9	62.9	I24
MEAN	31.5	9.0	.00	.53	1.00	.0	.97	.0			57.4	57.8	
S.D.	4.0	.0	1.10	.01	.44	1.0	.43	.9			18.3	3.6	

Figure 4. Third Stage of Small Group's Fit Order Item

SUMMARY OF 9 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	83.9	24.0	1.13	.32	.99	-.1	.97	-.2
S.D.	11.3	.0	1.14	.01	.38	1.4	.39	1.4
MAX.	104.0	24.0	3.17	.34	1.76	2.3	1.82	2.4
MIN.	69.0	24.0	-.39	.31	.42	-2.6	.42	-2.6

REAL RMSE .34 TRUE SD 1.09 SEPARATION 3.15 Person RELIABILITY .91
MODEL RMSE .32 TRUE SD 1.09 SEPARATION 3.40 Person RELIABILITY .92
S.E. OF Person MEAN = .40

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .92

SUMMARY OF 24 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	31.5	9.0	.00	.53	1.00	.0	.97	.0
S.D.	4.0	.0	1.10	.01	.44	1.0	.43	.9
MAX.	38.0	9.0	2.50	.54	1.85	1.7	1.92	1.7
MIN.	22.0	9.0	-1.81	.48	.30	-1.7	.29	-1.7

REAL RMSE .58 TRUE SD .94 SEPARATION 1.63 Item RELIABILITY .73
MODEL RMSE .53 TRUE SD .96 SEPARATION 1.83 Item RELIABILITY .77
S.E. OF Item MEAN = .23

Figure 5. The Results of Reliability Test in Small Group Stage

In the provided image, the results of the analysis are presented in the following table, showing the obtained values along with the criteria based on those values.

Table 10. The Results of Reliability Analyze in Small Group Stage

Statistics	Fit Indices	Criteria
Cronbach's Alpha (KR-20)	0.92	Very Good
Item Reliability	0.73	Enough
Person Reliability	0.91	Very Good
Item Separation	1.63	The higher the separation value, it can be said that the quality of the instrument is better.
Person Separation	3.15	Besides that, groups of item and person can be identified.

From the analysis results above, the value of Cronbach's Alpha (KR-20) is higher than the minimum threshold of 0.92, indicating that the instrument can be considered reliable

4. Field Test Stage

During the field test phase, the researcher analyzes the data from the research subjects to assess the quality of the instrument by conducting tests for validity, reliability, item difficulty, item discrimination, and the overall quality of the instrument. Microsoft Excel and the Rasch Model with Winstep software are used to determine the instrument's quality. To test the validity of an item during the field test phase, it can be observed in the output of the Winstep software, specifically in the table item fit order. The results of the validity test are analyzed according to the established criteria, and the following results are obtained:

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
1	121	36	.16	.22	1.42	1.7	1.49	1.8	A .55	.63	41.7	49.9	I1
14	118	36	.30	.22	1.41	1.6	1.36	1.4	B .47	.63	52.8	49.6	I14
5	117	36	.35	.22	1.38	1.5	1.34	1.4	C .44	.64	66.7	50.7	I5
8	126	36	-.09	.22	1.29	1.2	1.37	1.5	D .58	.63	38.9	49.0	I8
10	141	36	-.85	.23	1.36	1.5	1.34	1.3	E .64	.60	41.7	48.9	I10
3	124	36	.01	.22	1.25	1.1	1.25	1.0	F .56	.63	52.8	50.5	I3
15	126	36	-.09	.22	1.23	1.0	1.19	.8	G .59	.63	38.9	49.0	I15
2	125	36	-.04	.22	1.07	.4	1.03	.2	H .53	.63	52.8	50.1	I2
18	101	36	1.13	.22	1.02	.2	1.03	.2	I .49	.64	52.8	49.4	I18
19	114	36	.50	.22	1.01	.1	1.00	.1	J .71	.64	47.2	51.2	I19
17	123	36	.06	.22	1.01	.1	.97	.0	K .61	.63	55.6	49.9	I17
16	128	36	-.19	.22	.98	.0	.94	-.2	L .66	.62	61.1	48.7	I16
22	130	36	-.29	.22	.93	-.2	.90	-.3	l .63	.62	38.9	48.0	I22
21	138	36	-.69	.23	.93	-.3	.87	-.5	k .67	.60	50.0	49.1	I21
24	113	36	.55	.22	.87	-.5	.85	-.6	j .51	.64	72.2	51.2	I24
6	137	36	-.64	.23	.84	-.7	.79	-.8	i .76	.61	52.8	49.5	I6
20	115	36	.45	.22	.84	-.6	.84	-.6	h .68	.64	61.1	51.2	I20
7	132	36	-.39	.22	.82	-.8	.78	-.9	g .70	.62	47.2	47.9	I7
12	123	36	.06	.22	.81	-.8	.77	-.9	f .71	.63	52.8	49.9	I12
13	127	36	-.14	.22	.68	-1.5	.78	-.9	e .72	.62	55.6	49.0	I13
9	131	36	-.34	.22	.77	-1.0	.76	-1.0	d .79	.62	61.1	48.0	I9
4	121	36	.16	.22	.68	-1.5	.73	-1.1	c .72	.63	55.6	49.9	I4
23	116	36	.40	.22	.66	-1.5	.66	-1.5	b .62	.64	58.3	50.8	I23
11	132	36	-.39	.22	.59	-2.1	.58	-2.0	a .75	.62	66.7	47.9	I11
MEAN	124.1	36.0	.00	.22	.99	.0	.98	-.1			53.1	49.6	
S.D.	8.9	.0	.44	.00	.25	1.1	.25	1.0			8.9	1.0	

Figure 6. Fit Order Item of Field Test Stage

The picture above shows the results of the item fit order in the field test phase. All items meet the criteria for Outfit Mean Square (MNSQ), Point Measure Correlation (Pt Mean Corr), and Outfit Z-Standard (ZSTD) values. Therefore, all items in the field test phase are considered valid. To test the reliability during the field test phase, it can be observed in the output of the Winstep software, specifically in the statistics for Cronbach's Alpha (KR-20), Item and Person Reliability, and Item and Person Separation. The following results are obtained.

In the **Figure 7**, the analysis results are presented in the following table, including the obtained values and the criteria based on those values.

```
     SUMMARY OF 36 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      82.8       24.0        .77     .28        .99    -.2    .98    -.2 |
| S.D.      14.6         .0       1.11     .04        .59    1.9    .59    1.9 |
| MAX.     116.0       24.0       3.93     .51       2.92    5.5   2.86    5.4 |
| MIN.      49.0       24.0      -1.62     .26        .27   -3.2    .27   -3.2 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .31 TRUE SD    1.06  SEPARATION  3.48  Person RELIABILITY  .92 |
|MODEL RMSE    .28 TRUE SD    1.07  SEPARATION  3.82  Person RELIABILITY  .94 |
| S.E. OF Person MEAN = .19                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .93

     SUMMARY OF 24 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     124.1       36.0        .00     .22        .99     .0    .98    -.1 |
| S.D.       8.9         .0        .44     .00        .25    1.1    .25    1.0 |
| MAX.     141.0       36.0       1.13     .23       1.42    1.7   1.49    1.8 |
| MIN.     101.0       36.0       -.85     .22        .59   -2.1    .58   -2.0 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .23 TRUE SD     .37  SEPARATION  1.59  Item   RELIABILITY  .72 |
|MODEL RMSE    .22 TRUE SD     .38  SEPARATION  1.70  Item   RELIABILITY  .74 |
| S.E. OF Item MEAN = .09                                                     |
-------------------------------------------------------------------------------
```

Figure 7. Results of Reliability Test in Field Test Stage

Table 11. The Results of Reliability Analyze in Field Test Stage.

Statistics	Fit Indices	Criteria
Cronbach's Alpha (KR-20)	0.93	Very Good
Item Reliability	0.72	Enough
Person Reliability	0.92	Very Good
Item Separation	1.59	The higher the separation value, it can be said that the quality of the instrument is better. Besides that, groups of item and person can be identified.
Person Separation	3.48	

Based on the **Table 11,** analysis the results show that the Cronbach's Alpha (KR-20) value is above the minimum threshold of 0.93, indicating that the instrument can be considered reliable. To measure the level of difficulty, it can also be seen in the output from the Winstep software, which is the item measure table. Here are the results of the item measure:

Based on the **Figure 8**, it shows that the obtained item Separation Reliability (SD logit) value is 0.44, while the Measure logit value is obtained based on each item's value. The analysis of the difficulty level for each item in the questionnaire is presented in the **Table 12**.

The instrument is considered good because it has various levels of item difficulty, namely very difficult, difficult, moderate, and easy. Furthermore, to determine the item's discriminative power, the value of person separation is needed to classify the respondents into specific groups. The value of person separation can be seen in **Figure 7**, which is obtained as 3.48.

$$H = \frac{[(4 \times Separation) + 1]}{3}$$
$$H = \frac{[(4 \times 3.48) + 1]}{3}$$
$$H = 4.97$$
$$H = 5$$

Based on the results, it means that the instrument can classify respondents into 5 groups, namely very high, high, moderate, low, and very low. After obtaining all the analysis results through several stages of instrument testing that have been conducted, the quality of the non-test instrument in the form of a questionnaire is considered good enough because it meets all the criteria in the instrument testing.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
18	101	36	1.13	.22	1.02	.2	1.03	.2	.49	.64	52.8	49.4	I18
24	113	36	.55	.22	.87	-.5	.85	-.6	.51	.64	72.2	51.2	I24
19	114	36	.50	.22	1.01	.1	1.00	.1	.71	.64	47.2	51.2	I19
20	115	36	.45	.22	.84	-.6	.84	-.6	.68	.64	61.1	51.2	I20
23	116	36	.40	.22	.66	-1.5	.66	-1.5	.62	.64	58.3	50.8	I23
5	117	36	.35	.22	1.38	1.5	1.34	1.4	.44	.64	66.7	50.7	I5
14	118	36	.30	.22	1.41	1.6	1.36	1.4	.47	.63	52.8	49.6	I14
1	121	36	.16	.22	1.42	1.7	1.49	1.8	.55	.63	41.7	49.9	I1
4	121	36	.16	.22	.68	-1.5	.73	-1.1	.72	.63	55.6	49.9	I4
12	123	36	.06	.22	.81	-.8	.77	-.9	.71	.63	52.8	49.9	I12
17	123	36	.06	.22	1.01	.1	.97	.0	.61	.63	55.6	49.9	I17
3	124	36	.01	.22	1.25	1.1	1.25	1.0	.56	.63	52.8	50.5	I3
2	125	36	-.04	.22	1.07	.4	1.03	.2	.53	.63	52.8	50.1	I2
8	126	36	-.09	.22	1.29	1.2	1.37	1.5	.58	.63	38.9	49.0	I8
15	126	36	-.09	.22	1.23	1.0	1.19	.8	.59	.63	38.9	49.0	I15
13	127	36	-.14	.22	.68	-1.5	.78	-.9	.72	.62	55.6	49.0	I13
16	128	36	-.19	.22	.98	.0	.94	-.2	.66	.62	61.1	48.7	I16
22	130	36	-.29	.22	.93	-.2	.90	-.3	.63	.62	38.9	48.0	I22
9	131	36	-.34	.22	.77	-1.0	.76	-1.0	.79	.62	61.1	48.0	I9
7	132	36	-.39	.22	.82	-.8	.78	-.9	.70	.62	47.2	47.9	I7
11	132	36	-.39	.22	.59	-2.1	.58	-2.0	.75	.62	66.7	47.9	I11
6	137	36	-.64	.23	.84	-.7	.79	-.8	.76	.61	52.8	49.5	I6
21	138	36	-.69	.23	.93	-.3	.87	-.5	.67	.60	50.0	49.1	I21
10	141	36	-.85	.23	1.36	1.5	1.34	1.3	.64	.60	41.7	48.9	I10
MEAN	124.1	36.0	.00	.22	.99	.0	.98	-.1			53.1	49.6	
S.D.	8.9	.0	.44	.00	.25	1.1	.25	1.0			8.9	1.0	

Figure 8. The Results of Measure Item in Field Test Stage

Table 12. Analysis of Hardness Level in Field Test Stage

Item	Item measure	Criteria
18	1.13	Very Hard
24	0.55	Very Hard
19	0.50	Very Hard
20	0.45	Very Hard
23	0.40	Hard
5	0.35	Hard
14	0.30	Hard
1	0.16	Hard
4	0.16	Hard
12	0.06	Hard
17	0.06	Hard
3	0.01	Hard
2	-0.04	Medium
8	-0.09	Medium
15	-0.09	Medium
13	-0.14	Medium
16	-0.19	Medium
22	-0.29	Medium
9	-0.34	Medium
7	-0.39	Medium

Analysis of Student's Mathematical Self-Efficacy Level

In measuring students' self-efficacy level, analysis based on the criteria of mathematical self-efficacy at students' ability level is necessary. Students' ability level is used to identify their self-efficacy level in answering the questionnaire. The students' ability levels have been sorted from very high to very low based on the logit values for each person. Based on the **Table 4,** the criteria for grouping students' abilities are obtained based on the Standard Deviation Logit (SD Logit)..

Table 13. Criteria for Classifying Students' Level of Self-Efficacy

Score	Criteria
$X \geq 0.66$	Very High
$0.22 \leq X < 0.66$	High
$-0.22 < X < 0.22$	Medium
$-0.66 < X \leq -0.22$	Low
$X \leq -0.66$	Very Low

High logit values indicate high self-efficacy in answering the questionnaire. This corresponds to the total score column. The following are the results of person measures on the test instrument.

Person STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
36	116	24	3.93	.51	.94	.1	.95	.1	.02	.18	83.3	84.4	36
22	108	24	2.68	.32	.78	-.7	.72	-.9	.55	.28	54.2	54.2	22
1	103	24	2.22	.29	.88	-.4	.83	-.6	.65	.30	54.2	43.4	01
6	102	24	2.13	.28	1.03	.2	1.03	.2	.36	.31	37.5	43.8	06
20	102	24	2.13	.28	1.49	1.8	1.51	1.8	.30	.31	50.0	43.8	20
21	97	24	1.75	.27	1.09	.4	1.06	.3	.52	.32	37.5	42.4	21
34	97	24	1.75	.27	1.11	.5	1.08	.4	-.08	.32	54.2	42.4	34
15	95	24	1.61	.27	2.92	5.5	2.86	5.4	.63	.32	.0	42.7	15
18	93	24	1.47	.27	.43	-2.9	.43	-3.0	.54	.32	66.7	42.4	18
23	90	24	1.26	.26	.50	-2.4	.50	-2.4	.15	.32	75.0	43.3	23
8	89	24	1.19	.26	1.42	1.6	1.41	1.5	.49	.32	33.3	44.2	08
28	89	24	1.19	.26	1.06	.3	1.07	.4	.61	.32	37.5	44.2	28
29	89	24	1.19	.26	.55	-2.0	.54	-2.1	.55	.32	70.8	44.2	29
3	88	24	1.12	.26	.40	-2.9	.40	-2.9	.46	.32	70.8	44.6	03
26	87	24	1.05	.26	1.58	2.0	1.61	2.0	.62	.32	45.8	44.6	26
17	84	24	.84	.27	2.34	3.6	2.36	3.6	.18	.32	29.2	47.0	17
7	83	24	.77	.27	.48	-2.2	.48	-2.2	.14	.32	50.0	47.7	07
9	83	24	.77	.27	.82	-.6	.80	-.6	.24	.32	50.0	47.7	09
33	83	24	.77	.27	.57	-1.7	.57	-1.7	.20	.32	58.3	47.7	33
14	82	24	.70	.27	.82	-.6	.81	-.6	.23	.32	41.7	49.6	14
10	81	24	.62	.27	.45	-2.3	.45	-2.2	.42	.32	66.7	50.5	10
16	81	24	.62	.27	2.56	3.9	2.54	3.8	.25	.32	33.3	50.5	16
11	78	24	.41	.27	.91	-.2	.90	-.2	.72	.31	45.8	53.8	11
19	77	24	.33	.27	.77	-.7	.78	-.7	.14	.31	54.2	54.3	19
2	75	24	.18	.27	.81	-.5	.79	-.6	.36	.31	62.5	54.8	02
32	74	24	.11	.27	.27	-3.2	.27	-3.2	-.11	.31	83.3	55.6	32
5	73	24	.04	.27	.60	-1.4	.60	-1.3	.33	.31	54.2	55.9	05
4	72	24	-.04	.27	.41	-2.3	.40	-2.3	.34	.31	83.3	56.5	04
12	70	24	-.19	.27	.79	-.6	.80	-.6	.51	.32	58.3	56.5	12
24	70	24	-.19	.27	1.48	1.4	1.50	1.5	-.09	.32	41.7	56.5	24
25	70	24	-.19	.27	.78	-.7	.77	-.7	.10	.32	66.7	56.5	25
31	70	24	-.19	.27	.56	-1.5	.56	-1.5	-.05	.32	83.3	56.5	31
30	64	24	-.62	.27	.78	-.7	.79	-.7	.20	.32	58.3	50.5	30
13	63	24	-.69	.26	1.12	.5	1.13	.5	.30	.32	41.7	50.1	13
35	52	24	-1.42	.26	1.20	.9	1.20	.9	-.09	.33	33.3	40.5	35
27	49	24	-1.62	.26	.92	-.3	.91	-.3	.10	.32	45.8	40.6	27
MEAN	82.8	24.0	.77	.28	.99	-.2	.98	-.2			53.1	49.6	
S.D.	14.6	.0	1.11	.04	.59	1.9	.59	1.9			17.8	8.0	

Figure 9. The Results of Measure Person in Field Test Stage

Based on the **Figure 9**, the grouping of students' self-efficacy levels can be seen in the **Table 14.**

Table 14. The Results of Classifying Student's Self-Efficacy Level

Respondents	Total Score	Total Counts	Measure	Level Self-Efficacy
36	116	24	3.93	Very high
22	108	24	2.68	Very high
1	103	24	2.22	Very high
6, 20	102	24	2.13	Very high
21, 34	97	24	1.75	Very high
15	95	24	1.61	Very high
18	93	24	1.47	Very high
23	90	24	1.26	Very high
8, 28, 29	89	24	1.19	Very high
3	88	24	1.12	Very high
26	87	24	1.05	Very high
17	84	24	0.84	Very high
7, 9, 33	83	24	0.77	Very high
14	82	24	0.70	Very high
10, 16	81	24	0.62	High
11	78	24	0.41	High
19	77	24	0.33	High
2	75	24	0.18	Medium
32	74	24	0.11	Medium
5	73	24	0.04	Medium
4	72	24	-0.04	Medium
12, 24, 25, 31	70	24	-0.19	Medium
30	64	24	-0.62	Low
13	63	24	-0.69	Very low
35	52	24	-1.42	Very low
27	49	24	-1.62	Very low

The measurement results indicate that 89% of the respondents have a sufficiently good level of self-efficacy, while only 11% of the respondents have a low level of self-efficacy.

4. CONCLUSION

Based on the results of the description and data analysis in this research, several conclusions can be drawn as follows: 1) The development of self-efficacy instrument in mathematics learning meets the validity criteria. 2) The quality of the developed mathematical self-efficacy instrument is considered quite good as it meets all the criteria in the instrument testing. 3) The measurement results of students' self-efficacy level in mathematics learning using the developed instrument indicate that 89% of the respondents have a sufficiently good level of self-efficacy, while only 11% of the respondents have a low level of self-efficacy.

CONFLICT OF INTEREST

There are no conflicts of interest declared by the authors.

REFERENCES

Afniola, S., Ruslana, R., & Artika, W. (2020). Intelegensi Dan Bakat Pada Prestasi Siswa. *Al-Din: Jurnal Dakwah Dan Sosial Keagamaan*, *6*(1). https://doi.org/10.35673/ajdsk.v6i1.844

Ahriana, A., Yani, A., & Maruf, M. (2016). Studi Analisis Hubungan Antara Self Efficacy Dengan Hasil Belajar Fisika Siswa Kelas XI MIA SMA Negeri 1 Takalar. *Jurnal Pendidikan Fisika Unismuh*, *4*(2).

Anderha, R. R., & Maskar, S. (2020). Analisis Kemampuan Komunikasi Matematis Siswa Pada Pembelajaran Daring Materi Eksponensial. *Jurnal Ilmiah Matematika Realistik*, *1*(2). https://doi.org/10.33365/ji-mr.v1i2.438

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W.H. Freeman and Company. *American Psychological Association*, *23*.

Dewi, P. S. (2018). Efektivitas Pendekatan Open Ended Ditinjau dari Kemampuan Pemecahan Masalah Matematis. *Prisma*, *7*(1).

Djaali, & Mulyono, P. (2000). Pengukuran Dalam Bidang Pendidikan. In *Grasindo Publisher*.

Hairida, H. (2017). Pengembangan Instrumen Untuk Mengukur Self Efficacy Siswa Dalam Pembelajaran Kimia. *Edusains*, *9*(1). https://doi.org/10.15408/es.v9i1.4000

Handayani, S. L., & Iba, K. (2020). Karakteristik Tes Keterampilan Proses Sains: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal. *Publikasi Pendidikan*, *10*(2). https://doi.org/10.26858/publikan.v10i2.13051

Hendriana, H., & Kadarisma, G. (2019). Self-Efficacy dan Kemampuan Komunikasi Matematis Siswa SMP. *JNPM (Jurnal Nasional Pendidikan Matematika)*, *3*(1). https://doi.org/10.33603/jnpm.v3i1.2033

Husaini Usman. (2017). Pengantar Statistika (Edisi Ketiga): Cara Mudah Memahami Statistika. In *Hukum Perumahan*.

Husna, B. N. (2016). Implementasi Permendikbud Kurikulum 2013 Dalam Pelajaran Bahasa Indonesia Smk Kelas X. *Jurnal Ilmiah Profesi Pendidikan*, *1*(2). https://doi.org/10.29303/jipp.v1i2.13

Maskar, S., & Dewi, P. S. (2021). Peningkatan Kompetensi Guru Ma Darur Ridho Al-Irsyad Al Islamiyyah Pada Pembelajaran Daring Melalui Moodle. *Journal of Social Sciences and Technology for Community Service (JSSTCS)*, *2*(1). https://doi.org/10.33365/jsstcs.v2i1.875

Muslimah. (2013). Persepsi Mahasiswa Akuntansi terhadap Praktik-Praktik Kecurangan Akademik (Academic Fraud). In *Jurnal Ilmiah Mahasiswa FEB: Vol. Vol 2, No* (Issue No 1).

Puspaningtyas, N. D. (2019). Proses Berpikir Lateral Siswa SD dalam Menyelesaikan Masalah Matematika Open-Ended Ditinjau dari Perbedaan Gaya Belajar. *MAJAMATH: Jurnal Matematika Dan Pendidikan Matematika*, *2*(2). https://doi.org/10.36815/majamath.v2i2.373

Sari, A. D., Hastuti, S., & Asmiati, A. (2020). Pengembangan Model Creative Problem Solving (CPS) Untuk Meningkatkan Kemampuan Berpikir Reflektif Siswa. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *4*(2). https://doi.org/10.31004/cendekia.v4i2.318

Schunk, D. H. (1990). Goal Setting and Self-Efficacy During Self-Regulated Learning. *Educational Psychologist*, *25*(1). https://doi.org/10.1207/s15326985ep2501_6

Shadiq, F. (2007). Apa Dan Mengapa Matematika Begitu Penting? *Departemen Pendidikan Nasional Direktorat Jenderal Peningkatan Mutu Pendidik Dan Tenaga Kependidikan Pusat Pengembangan Dan Pemberdayaan Pendidik Dan Tenaga Kependidikan (PPPPTK) Matematika*.

Sistem Pendidikan Nasional. (2003). *Undang-Undang No.20 Tahun 2003*. Departemen Pendidikan Nasional.

Slameto. (2012). Belajar Dan Faktor-faktor Yang Mempengaruhinya. Jakarta: Rineke Cipta. *Journal of Chemical Information and Modeling*, *53*(9).

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif Dan R&D*.

Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan RASCH Pada Assessment Pendidikan. *AplikAsi RascH PemodelAn Pada Assessment Pendidikan*, *October*, 1–142.

Supardi US, L. &. (2010). Pengaruh Konsep Diri, Sikap Siswa Pada Matematika, Dan Kecemasan Siswa Terhadap Hasil Belajar Matematika. *Jurnal Cakrawala Pendidikan*, *3*(3). https://doi.org/10.21831/cp.v3i3.362

Tanjung, Z., & Amelia, S. (2017). Menumbuhkan Kepercayaan Diri Siswa. *JRTI (Jurnal Riset Tindakan Indonesia)*, *2*(2). https://doi.org/10.29210/3003205000

Tessmer, M. (1993). Planning and conducting formative evaluations: Improving the quality of education and training. In *Planning and Conducting Formative Evaluations*.

Wahyuningsih, E. T. (2015). Analisis Butir Soal Tes Objektif Buatan Guru Ulangan Semester Ganjil Mata Pelajaran Ekonomi Kelas X Di Sma Negeri 1 Mlati Tahun Ajaran 2013/2014. *Skripsi*, *97*(12).