Assessment of Validity and Reliability of Mathematical Attitude Instruments of Elementary School Students

Trisnani, N.*[1], Retnawati, H.[2], and Wuryandani, W.[3]

[1]*Faculty of Education, IKIP PGRI Wates, Indonesia*
[2]*Research and Evaluation Education Department,*
Universitas Negeri Yogyakarta, DI Yogyakarta 55281, Indonesia
[3]*Educational and Psychology Department,*
Universitas Negeri Yogyakarta, DI Yogyakarta 54363, Indonesia

E-mail: novytrisnani@ipw.ac.id
*Corresponding author

Abstract

Previous research has used various instruments to measure mathematical attitudes, but validated and reliable tools for Indonesian elementary students are limited. This study aims to fill this gap by developing an instrument suited to Indonesian elementary students to measure their mathematical attitudes effectively. Data were collected from 187 students using a Likert scale. Two experts assessed content validity, examining construct, content, and language aspects. To test validity and reliability, Exploratory Factor Analysis (EFA) was performed with SPSS 25.0, followed by Confirmatory Factor Analysis (CFA) using LISREL 10.20. Criteria included a Standard Loading Factor (SLF) of at least 0.30, a t-value of 1.96 or higher, and construct reliability (CR) above 0.70. The analysis showed that 20 items in the instrument met the criteria for validity and reliability, confirming their suitability for measuring mathematical attitudes. The Goodness-of-Fit test indicated a strong fit between the theoretical model and the empirical data, supporting the robustness of the instrument. The developed instrument reliably measures elementary students' mathematical attitudes, meeting a key need in the field. It aligns well with Indonesian students' cultural and educational context, serving as a valuable tool for educators and researchers.

Keywords: elementary school; mathematical attitude; reliability; validity.

1   Introduction

Research on mathematical attitudes has garnered significant attention due to its crucial role in the learning process. It is important to distinguish between attitudes toward mathematics, which generally encompass feelings and perceptions about the subject, and the more specific concept of mathematical attitude. The latter refers to students' readiness to approach, think about, and engage with mathematical problems in a logical and systematic manner [21]. Mathematical attitude involves a combination of emotional responses, beliefs, and behaviors toward mathematics [30]. While general attitudes include affective and cognitive responses, mathematical attitude emphasizes the cognitive effort put into understanding mathematical concepts, solving problems, and communicating ideas clearly [3].

Mathematical attitudes play a crucial role in the learning process of mathematics [15]. Previous studies have highlighted the significant influence of mathematical attitude on academic achievement, motivation, and students' overall success in learning mathematics [19]. A positive mathematical attitude has been shown to increase student engagement, leading to better problem-solving skills and enhanced persistence in facing mathematical challenges [2]. The importance of mathematical attitude is also supported by several studies that indicate affective factors play a role in students' success in mathematics [14]. Katagiri [21] emphasizes that mathematical thinking is one aspect influenced by students' attitudes, which can be observed in how students attempt to solve mathematical problems logically and systematically.

Attitudes toward mathematics encompass a multidimensional construct, including cognitive, affective, and behavioral components [37]. This attitude can be interpreted as a form of liking, enjoyment, interest, and belief in one's mathematical abilities and usefulness [31]. Mathematical attitudes can be reflected through indications of enthusiasm and sincerity in learning mathematics, support for the mathematics learning process, adequate knowledge to deal with mathematical material, curiosity, desire to ask questions, and enthusiasm to develop skills and experience in mathematics [28]. Katagiri [21] revealed that students' mathematical attitudes can be observed through the effort they make when overcoming a mathematical challenge. These efforts are reflected in students' attitudes toward understanding problems, thinking logically, conveying information clearly, and seeking better solutions.

Many instruments have been developed to measure mathematical attitudes, such as the Fennema-Sherman Mathematics Attitudes Scales (FSMAS) and the Mathematics Attitude Scale (MAS). These tools have been widely used in various studies to assess emotional responses and perceptions toward mathematics [12]. However, these instruments tend to focus more on affective aspects, such as anxiety, confidence, and enjoyment, without fully capturing the cognitive dimensions of mathematical attitude [25]. Additionally, these instruments were primarily developed in Western contexts, and their validity and reliability have not been thoroughly tested in culturally and educationally diverse settings like Indonesia [33].

This lack of contextual adaptation limits the reliability and validity of these instruments across various educational environments in Indonesia. For instance, FSMAS is designed to assess students' general feelings toward mathematics, but it may not adequately capture students' cognitive efforts in logical reasoning, problem-solving, and systematic thinking, which are key elements of mathematical attitude. Furthermore, these instruments do not fully address the role of cultural and social factors that may influence students' attitudes and approaches to mathematics [29]. Research has emphasized the importance of establishing the validity and reliability of mathematical attitude instruments [25] and it is also crucial to highlight that the validity and reliability of instruments used to measure mathematical attitudes have not been extensively studied [16].

Given these limitations, there is an urgent need for a contextually valid and reliable instrument specifically designed to assess mathematical attitudes among elementary school students in Indonesia. This study seeks to address this gap by developing and validating a new instrument specifically tailored to measure mathematical attitudes among Indonesian elementary students. Unlike existing instruments, this tool will incorporate both cognitive and affective dimensions, making it more suitable for understanding how students approach mathematical problem-solving within their cultural and educational contexts. Furthermore, the need for this research is also driven by challenges encountered in factor analysis, which is often faced when measuring mathematical attitudes, particularly on a Likert scale. The high correlation between factors can complicate the creation of accurate statements about students' mathematical attitudes [24]. By focusing on validity and reliability, the instrument developed in this study aims to provide a robust tool for educators and researchers to gain deeper insights into students' mathematical attitudes and how these attitudes can be developed to support their learning success.

In conclusion, this research is important not only for enriching the literature on attitudes toward mathematics but also for providing a contextually valid and reliable instrument for educators and researchers to measure mathematical attitudes in elementary school students. By filling the gap in existing research and addressing the limitations of previous instruments, this study contributes to the development of a more accurate tool for assessing mathematical attitudes in Indonesia.

2 Methods

2.1 Research design

Non-experimental quantitative research with a cross-sectional survey approach was applied in this study. Cross-sectional surveys are useful for assessing the prevalence of various characteristics, conditions, or behaviors within a population, as they allow researchers to analyze relationships between different variables without the need for longitudinal data collection [22]. The study population consisted of fourth-grade elementary school students. The sampling approach used was convenience sampling, based on the proximity and accessibility of the researchers, and the survey was conducted online.

Previous studies [12] have highlighted that cultural and linguistic factors significantly affect the validity and reliability of attitude measurements, emphasizing the need for contextually adapted instruments for the target population. Therefore, the items in the questionnaire were distributed in Indonesian to ensure clarity and comprehension for the respondents, who were elementary school students in Indonesia. Before distributing the questionnaire to students, we ensured that information about the research purpose and voluntary participation was clearly communicated to the relevant institutions, parents, and students. All research procedures adhered to ethical guidelines for educational research, including confidentiality and parental consent principles.

2.2 Sampling

The sample size was calculated based on the recommendations of Gable and Wolf [9], who suggest $6 - 10$ respondents per item for standard instrument development. Using the lower limit of six respondents per item and accounting for an anticipated response rate of 75%, the sample

size was determined as follows:

$$32 \text{ items} \times 6 \times \frac{100}{75} = 256, \text{ rounded up to } 250.$$

The survey was distributed to 850 students, resulting in a response rate of 74.8% (187 questionnaires returned). In the first section of the collected data, we gathered biographical responses such as gender and school origin. In the second section, we examined variables relevant to the research questions. Ultimately, this study involved 84 male students (45%) and 103 female participants (55%). A total of 89 participants (47%) were from the western region of Indonesia, 66 participants (35%) were from the central region, and 32 participants (17%) were from the eastern region of Indonesia.

2.3 Instrument development

The development of the questionnaire items was an iterative process that included both theoretical and empirical validation stages [27]. To ensure alignment with the theoretical model, the initial items were developed through a comprehensive literature review of similar instruments, including the Fennema-Sherman Mathematics Attitudes Scales (FSMAS) and the Mathematics Attitude Scale (MAS), which have been widely used but primarily focus on emotional responses to mathematics [12]. Theoretical validation was conducted through an extensive review of relevant literature and consultation with two experts specializing in elementary education and assessment. This theoretical validation ensured that the instrument was age-appropriate and aligned with the cognitive abilities of elementary students.

The development of the mathematics attitude questionnaire items was primarily guided by the basic framework of mathematical attitudes as outlined by Katagiri [21]. There are four main dimensions: the effort to clearly understand problems, the effort to act logically, the effort to communicate information clearly and concisely, and the effort to seek better solutions. These dimensions were selected based on their relevance to students' cognitive engagement with mathematical tasks, as they reflect the deeper thinking processes required to systematically solve problems.

For content validation, the Content Validity Index (CVI) was employed, following the procedure recommended by Lynn (1986). Experts rated each item on a scale from 1 (not relevant) to 4 (very relevant). The results showed that the I-CVI for all items ranged from 0.83 to 1.00, indicating excellent content validity. Meanwhile, the S-CVI, which is the average of all I-CVIs, reached 0.90, exceeding the recommended threshold (0.80), indicating that this questionnaire has strong content validity.

2.4 Variable measurement

There are four main dimensions, including the effort to clearly understand problems (8 items; for example, the student's effort to ask questions), the effort to act logically (8 items; for example, the effort to think based on relevant data), the effort to communicate information clearly and concisely (8 items; for example, the effort to organize information in a logical sequence), and the effort to seek better solutions (8 items; for example, the effort to evaluate thinking results both objectively and subjectively). The indicators for each dimension of mathematical attitudes are outlined in Table 1.

Table 1: Mathematical attitude items.

Dimension	Items		$\sum$
	Positive $(+)$	**Negative** $(-)$	
The effort to clearly understand problems	1, 2, 3, 4, 5	6, 7, 8, 9, 10	8
The effort to act logically	11, 12, 13, 14, 15	16, 17, 18, 19, 20	8
The effort to communicate information clearly and concisely	21, 22, 23, 24, 25	26, 27, 28, 29, 30	8
The effort to seek better solutions	31, 32, 33, 34, 35	36, 37, 38, 39, 40	8
Total	**16**	**16**	**32**

2.5   Data analysis

To ensure the credibility of the research results, data analysis was conducted using both EFA and CFA models. EFA allows researchers to analyze and differentiate the number of common factors and factor loading patterns from a set of observed variables [36], which can then be tested and confirmed through CFA [10]. CFA is tools to test the unidimensionality, reliability and validity of the models [26]. The CFA models fit were examined based on Chi-square statistics, Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Comparative Fit Index (CFI), Incremental Fit Index (IFI), Normed Fit Index (NFI), Tucker Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA) indices (set at 0.05) [20]. The main components used in this study included the Kaiser-Meyer-Olkin (KMO) value, Bartlett's test, factor loadings, eigenvalues, scree plot, and varimax rotation. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett's Test of Sphericity were applied to determine whether the data were suitable for factor analysis. A KMO value above 0.90 is considered excellent, while values between 0.60 and 0.70 are deemed acceptable [5]. A $p-$value of less than 0.05 for Bartlett's test indicates that the data are not an identity matrix. Eigenvalues greater than one suggest the presence of distinct components. Additionally, the scree plot was analyzed to confirm the number of dimensions to be retained. According to Hair et al. (2010), factor loadings greater than 0.50 for each item are considered significant for validating the questionnaire's meaning.

Validity testing of the items is conducted utilizing Structural Equation Modeling (SEM) analysis and Confirmatory Factor Analysis (CFA) through the LISREL 10.20 Full Version software, which is employed to assess the mathematical attitude instrument model. Observed verification will be eliminated if the Standard Loading Factor (SLF) analysis value is less than the critical limit of 0.50. Apart from this critical limit [18], Wijayanto [38] suggests that items with loading factors ranging from 0.30 to less than 0.50 and a t-value of ≥ 1.96 will be retained in the model. Hair et al [13] assert that a loading factor of ≥ 0.30 is considered a valid criterion in CFA. The testing technique calculates reliability using the consistency level of the manifestation variable which can be determined by measuring latent constructs using Construct Reliability (CR). The standard for accepting reliability, CR > 0.70 [23].

The Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) are the two primary measures used to evaluate model fit. Additionally, indicators of model fit such as the Normed Fit Index (NFI), Comparative Fit Index (CFI), Goodness-of-Fit Index (GFI), and non-NFI or Tucker-Lewis Index (TLI) are also utilized [1]. Previous research suggested that RMSEA values were below 0.08 [1], and the Standardized Root Mean Square Residual (SRMR) should be 0.80 [41]. The amount of cut-off value for CFI and NFI values above 0.95 and 0.80, respectively [40]. CFI and NFI range from 0 to 1, with values closer

N. Trisnani *et al.*

Malaysian J. Math. Sci. 20(1): 183–198(2026) *183 - 198*

to 1 indicating a stronger relationship between variance and covariance [8]. GFI ranges from 0 to 1, with higher values indicating better fit. While there is no universally agreed cutoff value for GFI, it is commonly suggested that a GFI value of 0.90 or higher indicates a good model fit [7]. The reference states that in the context of the study, the Tucker-Lewis Index (TLI) limit value for appropriate suitability is TLI $\geq$ 0.90 [17]. TLI values greater than one are possible (i.e., a value of one does not signify a perfect fit, but instead any other fit), and they can be interpreted as an indication of an excellent model fit [11].

3 Results

The theoretical validity of the mathematical attitude instrument by experts indicates improvement in structure or composition, language usage, and content. Initial items with a total of 40 items, corrected and revised following expert judgment input. The final result of the revision produces 32 items of valid statements to measure mathematical attitudes according to expert judgment. Some points of note for improvement include correcting sentences in statements containing compound sentences, improving sentences in items to better fit the grid, the need to eliminating items because they do not fit the grid. For example, the sentence in the item "I feel proud when I can solve a math problem without help", the statement does not fit the grid because in this statement there is no indication of student effort, change the redaction of the statement item because it can make the data biased, and correct the word containing the answer option in the item.

| 3 | Saya suka belajar matematika dan berusaha memahami konsepnya dengan sungguh-sungguh. |

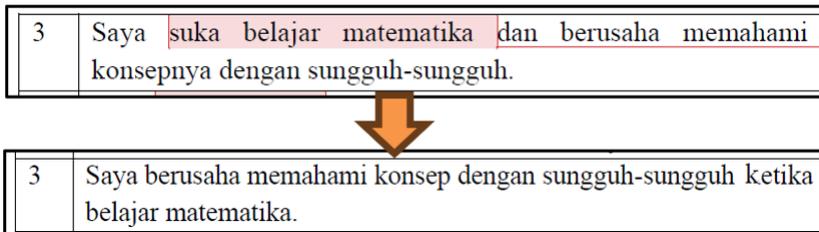| 3 | Saya berusaha memahami konsep dengan sungguh-sungguh ketika belajar matematika. |

Figure 1: Redaction revision of items.

The analysis to test the fit between the theoretical model and the empirical data was conducted using Exploratory Factor Analysis (EFA) with IBM SPSS Statistics Version 25.0 and Confirmatory Factor Analysis (CFA) with the Lisrel 10.20 Full Version program. Item validity testing was performed using Structural Equation Modeling (SEM).

Descriptive analysis for each dimension includes the calculation of the mean, standard deviation (SD), skewness, kurtosis, and intercorrelation. Based on Table 2, the skewness and kurtosis values for all dimensions and items fall within the range of -3 to $+3$ when using SEM.

Table 2: Descriptive analysis.

Dimension	Mean	SD	Skewness	Kurtosis	1	2	3	4
Understanding the problem	25.79	4.96	−0.67	−0.05	1	0.23**	0.139	0.002
Act logically	19.20	6.42	0.82	−0.06		1	−0.55**	−0.080
Convey information	27.59	4.97	−0.09	−0.72			1	0.67**
Understanding the problem	27.36	4.94	−0.25	−0.53				1

** Correlation is significant at the 0.01 level (2-tailed)

The overall items in the dimension have moderate to strong associations (ranging ($r = -0.080$ to $r = 0.67$, $p \leq 0.01$) all associations less than 0.90 indicate that there is no multicollinearity. The average score differed in each dimension, with $M = 25.79$ and $SD = 4,959$ for understanding the problem clearly, $M = 19.20$ and $SD = 6,420$ for efforts to act logically, $M = 27.59$ and $SD = 4,969$ for trying to convey information clearly and concisely, and $M = 27.36$ and $SD = 4,938$ for efforts to find better solutions.

Table 3: EFA results.

Dimension	Items	Communalities	Eigenvalue	% of Variance	Loading factor 1	2	3	4
Understanding the problem	F1I1	0.78	5.71	35.70	0.79			
	F1I2	0.80			0.84			
	F1I3	0.86			0.89			
	F1I4	0.88			0.91			
	F1I5	0.82			0.83			
	F1I6	0.83			0.83			
	F1I7	0.82			0.83			
	F1I8	0.46			0.67			
Act logically	F2I9	0.74	2.22	13.89	0.72			
	F2I10	0.93			0.93			
	F2I11	0.51			0.70			
	F2I12	0.89			0.91			
	F2I13	0.88			0.84			
	F2I14	0.75			0.76			
	F2I15	0.91			0.90			
	F2I16	0.76	1.66	10.38	0.74			
Understanding the problem	F4I25	0.80				0.82		
	F4I26	0.63				0.70		
	F4I27	0.24				0.41		
	F4I28	0.92				0.91		
	F4I29	0.82				0.80		
	F4I30	0.53				0.65		
	F4I31	0.92				0.92		
	F4I32	0.51				0.67		
Convey information	F3I17	0.64	1.17	7.32				0.70
	F3I18	0.65						0.70
	F3I19	0.71						0.74
	F3I20	0.81						0.83
	F3I21	0.68						0.72
	F3I22	0.74						0.74
	F3I23	0.73						0.76
	F3I24	0.83						0.84

The EFA output recommends that four factors explain 67.29% of the variance. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was 0.812, indicating that the sample was appropriate, and Bartlett's Test of Sphericity yielded a $p-$value of less than 0.001. Therefore, this study employed KMO, Bartlett's test, factor loadings, eigenvalues, scree plot, and varimax rotation. According to Table 3, the communalities for the 32 questions ranged from 0.49 to 0.79. The first factor, explaining 35.7% of the variance, was related to the effort to understand problems clearly. The second factor, accounting for 13.86% of the variance, was associated with logical reasoning. The third factor, explaining 10.38% of the variance, was linked to efforts to find better solutions, and the final factor, explaining 7.32% of the variance, was related to the effort to communicate information clearly and concisely. Overall, the loading value factor of all items was more than 0.50, so there was no cross loading in this study. The scree-plot test supports the option of maintaining four factors; therefore the study maintains four factors (Figure 2).
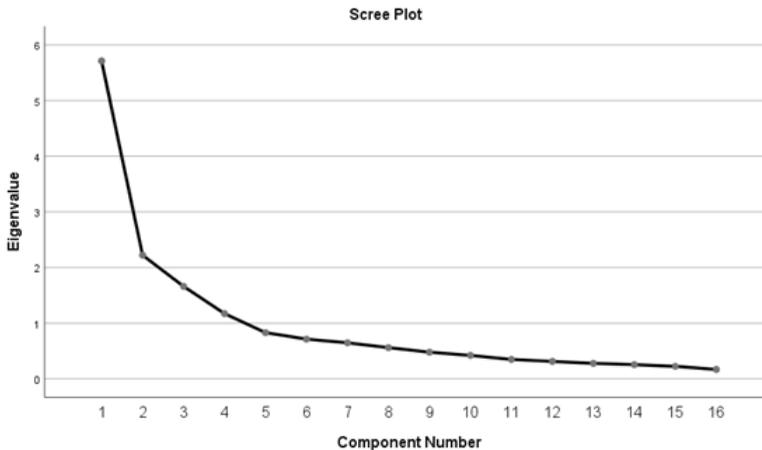

Figure 2: Scree plot 32 mathematical attitude items.

The mathematical attitude questionnaire instrument was created using Confirmatory Factor Analysis (CFA), which includes assessing the measurement model, evaluating the structural model, and testing model fit. The CFA analysis revealed that the instrument comprises 32 valid items, as shown in Table 4.

Table 4: CFA model loading factor mathematical attitude instrument (Part 1).

Indicator	Manifest	Loading factor		Error	Verification
		SLF	t-value		
F1	F1I1	0.64	8.37	0.59	Valid
	F1I2	0.66	8.78	0.56	Valid
	F1I3	0.57	7.41	0.67	Valid
	F1I4	0.64	8.39	0.59	Valid
	F1I5	0.58	7.52	0.66	Valid
	F1I6	-0.02	-0.25	1.00	Not valid
	F1I7	-0.02	-0.27	1.00	Not valid
	F1I8	0.03	0.30	1.00	Not valid
F2	F2I9	0.49	6.45	0.76	Valid
	F2I10	0.65	8.90	0.58	Valid
	F2I11	0.45	5.85	0.79	Valid

	F2I12	0.62	8.44	0.61	Valid
	F2I13	0.62	8.41	0.62	Valid
	F2I14	0.51	6.67	0.74	Valid
	F2I15	0.65	8.98	0.57	Valid
	F2I16	0.38	4.88	0.85	Valid
F3	F3I17	0.10	1.15	0.90	Not valid
	F3I18	0.13	1.57	0.98	Not valid
	F3I19	0.50	6.29	0.75	Valid
	F3I20	0.62	7.98	0.61	Valid
	F3I21	0.59	7.46	0.66	Valid
	F3I22	0.60	7.64	0.64	Valid
	F3I23	0.62	7.91	0.62	Valid
	F3I24	0.12	1.43	0.99	Not valid
	F4I25	0.41	4.83	0.83	Valid
F4	F4I26	0.36	4.17	0.87	Valid
	F4I27	−0.01	−0.16	1.00	Not valid
	F4I28	0.60	7.24	0.64	Valid
	F4I29	0.61	7.36	0.62	Valid
	F4I30	0.38	4.47	0.86	Valid
	F4I31	0.58	7.08	0.66	Valid
	F4I32	−0.12	−1.34	0.99	Not valid

The result of calculating Construct Reliability (CR) is 0.89, which shows that the accepted reliability standard, Construct Reliability (CR) > 0.70.

Figure 3 presents the first-order measurement of CFA with the t-value model, and shows that 8 items do not have significant weight, namely 3 items from indicator 1 (trying to understand the problem, and the purpose and substance of the problem clearly, by oneself) namely items number $6, 7$ and 8; 3 items from indicator 3 (showing efforts in conveying information clearly and concisely) namely items number $17, 18$ and 24 and 2 items from indicator 4 (indicating efforts in finding a better solution) namely items number 27 and 32.

The results in Figure 3 show that the model depicted is fit ($p-$value $= 0.02237 > \alpha$). Based on the results of LISREL analysis for the mathematical attitude questionnaire instrument in the figure above, Chi-Square $= 520.74$ with df $= 458$, $p-$value $= 0.02237$, and RMSEA $= 0.027$. In more detail, the results of the model fit test can be seen in Table 5.
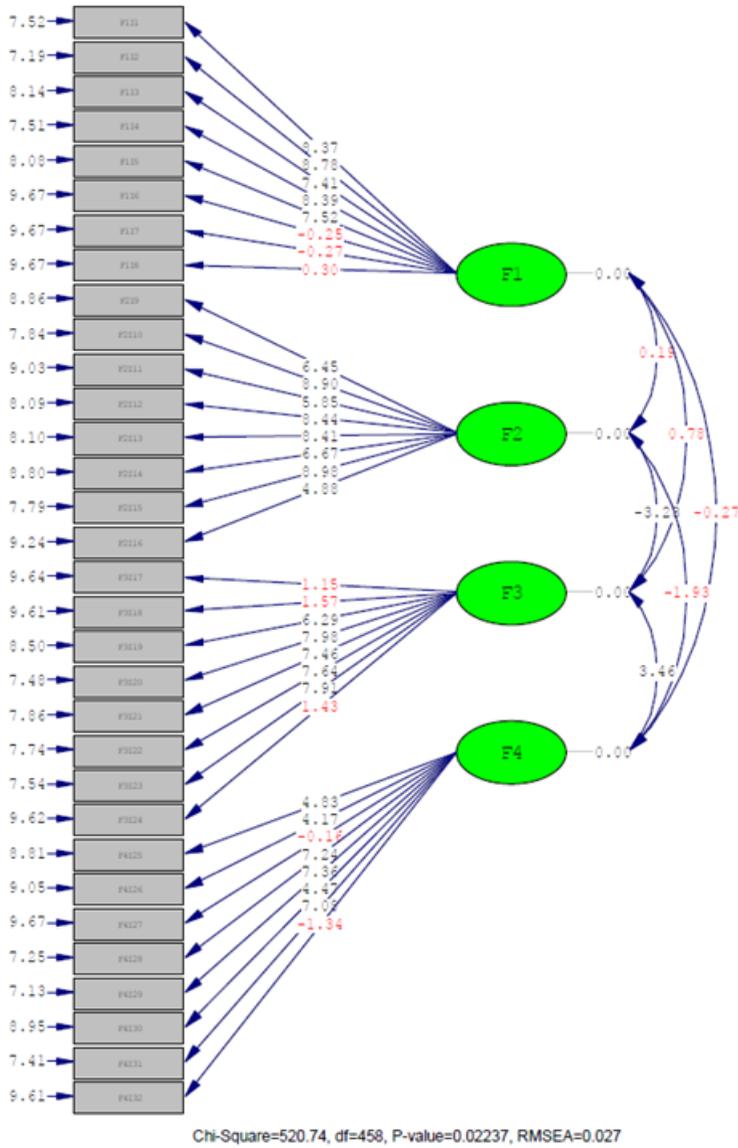
Figure 3: Path diagram for first-order CFA.

According to Table 5, there are 4 indices classified as poor fit, 3 indices categorized as marginal fit, and 5 indices falling under the good fit category. Based on these results, it can be concluded that the overall model fit is good fit.

Table 5: CFA model-goodness of fit test mathematical attitude instrument.

Index	Cut-off	GoF value	Verification
χ^2	$p-$value < 0.05	0.0000	Good fit
RMSEA	≤ 0.08	0.0511	Good fit
Root Mean Residual/RMR	≥ 0.80	0.0856	Good fit
Normed Fit Index/NFI	≥ 0.80	0.8930	Good fit
Incremental Fit Index/IFI	≥ 0.90	0.9590	Good fit
Comparative Fit Index/CFI	≥ 0.90	0.9580	Good fit
Goodness of Fit Index/GFI	≥ 0.90	0.8630	Marginal fit
Relative Fit Index/RFI	≥ 0.90	0.8560	Marginal fit
Non Normed Fit Index/NNFI	≥ 0.90	0.9430	Good fit
Parsimony Goodness of Fit Index/PGFI	≥ 0.60	0.6000	Good fit
Parsimonious Normed Fit Index/PNFI	≥ 0.90	0.6610	Poor fit
Adjusted Goodness of Fit Index/AGFI	≥ 0.90	0.8030	Marginal fit

The theoretical and empirical validation processes conducted on the mathematics attitude questionnaire instrument indicate that 8 items are invalid in the theoretical test and 8 items are invalid in the empirical test. Subsequently, the 24 valid items, the aspect of trying to understand the problem, and the purpose and substance of the problem clearly, by oneself is represented by 5 items, the aspect showing effort in taking logical action is represented by 8 items, the aspect shows effort in conveying information clearly and concisely represented by 5 items, and the aspect shows effort in finding a better solution is represented by 6 items.

Table 6: Mathematical attitude instrument after validation.

Dimension of mathematical attitude	Items		$\sum$
	$(+)$	$(-)$	
The effort to clearly understand problems	1, 2, 3, 4	5	5
The effort to act logically	6, 7, 8, 9	10, 11, 12, 13	8
The effort to communicate information clearly and concisely	14, 15	16, 17, 18	5
The effort to seek better solutions	19, 20, 21	22, 23, 24	6
Total	13	11	24

The measurement of mathematical attitudes carried out provides an overview of the condition of students' mathematical attitudes based on 24 questionnaire items. The aspect of trying to understand the problem, and the purpose and substance of the problem clearly, by oneself is represented by 5 items, the aspect showing effort in taking logical action is represented by 8 items, the aspect shows effort in conveying information clearly and concisely represented by 5 items, and the aspect shows effort in finding a better solution is represented by 6 items.

4 Discussion

Validity and reliability play a critical role in maintaining the effectiveness and credibility of measurement instruments [6]. Especially within educational contexts, the creation and validation of tools for gauging attitudes or competencies demand a thorough validation process to uphold the reliability of the findings [32]. This study aims to test the validity and reliability of a mathematics attitude questionnaire, a standard instrument used to measure the mathematics attitudes of elementary school students. EFA and CFA models were used to analyze and validate the measurement of mathematics attitudes. The findings of this study indicate that the items in the mathematics attitude questionnaire are generally suitable and applicable for assessing the mathematics attitudes of elementary school students, as reflected by the responses from participants in Indonesia. The mathematics attitude questionnaire aligns with our initial hypothesis, we found that the mathematics attitude questionnaire exhibits generally satisfactory EFA and CFA characteristics. The CFA model was used in this study to analyze and validate the mathematics attitude questionnaire, processed using LISREL 10.20 Full Version. The primary goal of conducting CFA is to evaluate the fit of the measurement model by confirming the relationships between observed variables and the underlying latent constructs [35]. LISREL is generally used for latent variable analysis and can be employed to confirm the validity of instrument scales and assess structural relationships among scales [39].

In addition to the EFA and CFA results, content validity was carefully evaluated by expert reviewers, ensuring the items' relevance to the intended construct. Initial revisions based on expert feedback led to 32 final items that were considered valid for measuring mathematical attitudes. Experts focused on refining the wording of items to improve clarity, alignment with the construct, and cultural appropriateness for Indonesian students. This process confirmed that the instrument reflects the desired aspects of mathematical attitudes, such as problem-solving effort, logical reasoning, and clarity in communication.

According to the EFA output, the data on the mathematics attitudes of elementary school students revealed four factors: efforts to understand problems clearly, efforts to take logical actions, efforts to convey information clearly and concisely, and efforts to seek better solutions. The CFA findings confirmed the EFA conclusions, as the mathematics attitude measure fit the data. The Confirmatory Factor Analysis (CFA) conducted using LISREL 10.20 on the mathematics attitude questionnaire indicated that the model overall exhibited a good fit with the data, with a Chi-Square value of 520.74 for 458 degrees of freedom and a ratio approaching 1.14, which aligns with the criteria for a good fit [23]. The $p-$value of this model is 0.02237, indicating that the model is statistically significant, and the RMSEA value of 0.027 confirms an excellent fit, consistent with the criteria set by Browne and Cudeck [1]. The quality of the items is assessed from the factor loadings represented by the path coefficients in the measurement model. Standard Loading Factor (SLF) is ≥ 0.30 and the calculated t-value is ≥ 1.96. Additionally, reliability is assessed through SEM analysis, which measures reliability via Construct Reliability, helping to determine the consistency of manifest variables in evaluating latent constructs. The standard for acceptable reliability is a construct reliability (CR) value > 0.70.

The findings confirm that the mathematical attitude instrument derived from assessing the CFA model's adequacy exhibits a commendable Goodness of Fit, as indicated by 11 GoF indices, encompassing RMSEA, NNFI, NFI, CFI, IFI, RFI, GFI, RMR, AGFI, and PGFI measures of suitability. The application of EFA and CFA to the mathematics attitude measure contributes to the confirmation of the validity and utility of the mathematics attitude questionnaire. The study has demonstrated that mathematics attitudes possess adequate reliability and validity, highlighting the importance of these factors in understanding students' attitudes towards mathematics [4].

The impact of mathematical attitudes on educational outcomes has been extensively documented, underscoring their significance across various academic and psychological domains, thereby emphasizing the need for valid and reliable measures of affective skills [34]. Furthermore, the reliability and validity of the four-factor mathematics attitude model are supported by strong alpha coefficients and factor loadings, highlighting the importance of these factors in understanding mathematics attitudes.

5 Conclusions, Limitations, and Future Work

This study successfully developed and validated an instrument for measuring mathematical attitudes among elementary school students in Indonesia. According to the EFA output, data on students' mathematical attitudes reveal four factors: efforts to clearly understand problems, efforts to take logical actions, efforts to convey information clearly and concisely, and efforts to seek better solutions. The CFA results validate the EFA conclusions, showing that the measurement model aligns well with the data and confirming the validity and reliability of the mathematics attitude questionnaire. Additionally, expert reviewers assessed content validity, leading to the final 32 items, which were refined for clarity, construct alignment, and cultural relevance for Indonesian students. However, we recognize some limitations in this study. The primary limitation is the small dataset, which impacts the validity of the conclusions. The analysis is based on a sample of only 187 respondents from three different regions in Indonesia, affecting the ability to generalize the findings to other groups. Future research should aim to replicate these results with a larger, randomly selected sample from more diverse cultural backgrounds to enhance generalizability. Additionally, since the data were collected online, this may have left out individuals without internet access. Future studies should explore various data collection methods to include a broader range of participants.

Conflicts of Interest The authors declare no conflict of interest.

References

[1] M. W. Browne & R. Cudeck (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005.

[2] I. G. N. Darmawan (2020). The changes in attitudes of 15-year-old Australian students towards reading, mathematics and science and their impact on student performance. *Australian Journal of Education*, *64*(3), 304–327. https://doi.org/10.1177/0004944120947873.

[3] V. A. DeBellis & G. A. Goldin (2006). Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics*, *63*(2), 131–147. https://doi.org/10.1007/s10649-006-9026-4.

[4] S. S. Dewanti, Y. Ayriza & F. A. Setiawati (2020). The application of item response theory for development of a students' attitude scale toward mathematics. *The New Educational Review*, *60*(2), 108–123. https://doi.org/10.15804/tner.20.60.2.09.

[5] J. R. dos Reis-Junior, D. Bassi-Dibai, D. N. Morais, A. Pontes-Silva, L. P. Mendes, J. S. Pinheiro, C. A. Fidelis-de Paula-Gomes & A. V. Dibai-Filho (2022). Translation, cross-cultural adaptation, and validation of the Athlete Fear Avoidance Questionnaire (AFAQ) into Brazilian Portuguese. *BMC Musculoskeletal Disorders*, *23*(1), Article ID: 974. https://doi.org/10.1186/s12891-022-05951-0.

[6] F. Erika & A. Rahmadani (2021). Development of 21st century skills-based stereochemistry learning tools to train students' argumentation skills. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, *7*(4), 822–833. https://doi.org/10.33394/jk.v7i4.4088.

[7] F. Fadillah & M. Tarigan (2022). Validity and reliability study of IST 2000R Indonesian version. *Journal of Educational, Health and Community Psychology*, *11*(4), 874–889. https://doi.org/10.12928/jehcp.v11i4.24693.

[8] G. Fogarty, P. Cretchley, C. Harman, N. Ellerton & N. Konki (2001). Validation of a questionnaire to measure mathematics confidence, computer confidence, and attitudes towards the use of technology for learning mathematics. *Mathematics Education Research Journal*, *13*(2), 154–160. https://doi.org/10.1007/BF03217104.

[9] R. K. Gable & M. B. Wolf (2012). *Instrument Development in the Affective Domain: Measuring Attitudes and Values in Corporate and School Settings* volume 36 of *Evaluation in Education and Human Services*. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-1400-4.

[10] M. Ghardallou, C. Zedini, J. Sahli, T. Ajmi, H. Khairi & A. Mtiraoui (2022). Psychometric properties of a French version of the Jefferson Scale of Empathy. *International Journal of Medical Education*, *13*, 205–214. https://dx.doi.org/10.5116/ijme.62d2.8497.

[11] D. Goretzko, K. Siemund & P. Sterner (2024). Evaluating model fit of measurement models in confirmatory factor analysis. *Educational and Psychological Measurement*, *84*(1), 123–144. https://doi.org/10.1177/00131644231163813.

[12] M. Gúlburnu & K. Yildirim (2021). Development and implementation of mathematics attitudes scale for the primary and secondary students. *Pegem Journal of Education and Instruction*, *11*(4), 177–184. https://doi.org/10.47750/pegegog.11.04.17.

[13] J. F. Hair Jr, M. Lds Gabriel, D. da Silva & S. Braga Junior (2019). Development and validation of attitudes measurement scales: Fundamental and practical aspects. *RAUSP Management Journal*, *54*(4), 490–507. https://doi.org/10.1108/RAUSP-05-2019-0098.

[14] L. E. Hart (1989). *Affect and Mathematical Problem Solving: A New Perspective*, chapter Describing the affective domain: Saying what we mean, pp. 37–45. Springer, New York, USA. https://doi.org/10.1007/978-1-4612-3614-6_3.

[15] B. Hemmings, P. Grootenboer & R. Kay (2011). Predicting mathematics achievement: The influence of prior achievement and attitudes. *International Journal of Science and Mathematics Education*, *9*(3), 691–705. https://doi.org/10.1007/s10763-010-9224-5.

[16] R. Hidayat, W. I. W. Idris, H. Qudratuddarsi & M. N. A. Rahman (2021). Validation of the mathematical modeling attitude scale for Malaysian mathematics teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(12), Article ID: em2047. https://doi.org/10.29333/ejmste/11375.

[17] E. T. Horata, S. Kalkan, G. Taşkın & J. S. Tedla (2022). Validity and reliability of the Turkish version of the questionnaire for physical therapy students' attitudes toward their profession and education (Q-PTSAPE). *Türk Fizyoterapi ve Rehabilitasyon Dergisi*, *33*(3), 219–226. https://doi.org/10.21653/tjpr.1096839.

[18] M. Igbaria & J. J. Baroudi (1993). A short-form measure of career orientations: A psychometric evaluation. *Journal of Management Information Systems*, *10*(2), 131–154. https://doi.org/10.1080/07421222.1993.11518003.

[19] Ð. Kadijević (2008). TIMSS 2003: Relating dimensions of mathematics attitude to mathematics achievement. *Zbornik Instituta za Pedagoska Istrazivanja*, *40*(2), 327–346. https://doi.org/10.2298/ZIPI0802327K.

[20] H. Kamalimoghaddam, R. A. Tarmizi, A. F. M. Ayub & W. M. W. Jaafar (2016). Confirmatory model of mathematics self-efficacy, problem solving skills and prior knowledge on mathematics achievement: A structural equation model. *Malaysian Journal of Mathematical Sciences*, *10*(S), 187–200.

[21] S. Katagiri (2004). *Mathematical Thinking and How to Teach It*. Meijitosyo, Tokyo.

[22] U. S. Kesmodel (2018). Cross-sectional studies – what are they good for? *Acta Obstetricia et Gynecologica Scandinavica*, *97*(4), 388–393. https://doi.org/10.1111/aogs.13331.

[23] R. Kline (2016). *Principles and Practice of Structural Equation Modeling*. The Guilford Press, New York 4 edition.

[24] C. León-Mantero, J. C. Casas-Rosal, C. Pedrosa-Jesús & A. Maz-Machado (2020). Measuring attitude towards mathematics using Likert scale surveys: The weighted average. *PLOS One*, *15*(10), Article ID: e0239626. https://doi.org/10.1371/journal.pone.0239626.

[25] J. R. McGinnis, S. Kramer, G. Shama, A. O. Graeber, C. A. Parker & T. Watanabe (2002). Undergraduates' attitudes and beliefs about subject matter and pedagogy measured periodically in a reform-based mathematics and science teacher preparation program. *Journal of Research in Science Teaching*, *39*(8), 713–737. https://doi.org/10.1002/tea.10042.

[26] L. Muhamad Safiih & M. A. Nor Azreen (2016). Confirmatory factor analysis approach: A case study of mathematics students' achievement in TIMSS. *Malaysian Journal of Mathematical Sciences*, *10*(S), 41–51.

[27] T. C. Mutohir, T. Lowrie & S. Patahuddin (2018). The development of a student survey on attitudes towards mathematics teaching-learning processes. *Journal on Mathematics Education*, *9*(1), 1–14. https://doi.org/10.22342/jme.9.1.4193.1-14.

[28] A. Nasrullah & M. Marsigit (2016). Keefektifan problem posing dan problem solving ditinjau dari ketercapaian kompetensi, metode, dan sikap matematis. *PYTHAGORAS Jurnal Matematika dan Pendidikan Matematika*, *11*(2), 123–135. https://doi.org/10.21831/pg.v11i2.11180.

[29] A. Palacios, V. Arias & B. Arias (2014). Attitudes towards mathematics: Construction and validation of a measurement instrument las actitudes hacia las matemáticas: Construcción y validación de un instrumento para su medida. *Revista de psicodidáctica*, *19*(1), 67–91. https://doi.org/10.1387/RevPsicodidact.8961.

[30] C. Primi, A. Bacherini, C. Beccari & M. A. Donati (2020). Assessing math attitude through the Attitude Toward Mathematics inventory – Short form in introductory statistics course students. *Studies in Educational Evaluation*, *64*, Article ID: 100838. https://doi.org/10.1016/j.stueduc.2020.100838.

[31] J. Russo, P. Kalogeropoulos & A. Roche (2023). Exploring underachieving students' views of, and attitudes towards, mathematics across stage of schooling. *Asian Journal for Mathematics Education*, *2*(2), 240–257. https://doi.org/10.1177/27527263231177435.

[32] G. W. Schneider, O. Lage, J. Fairclough, V. D. Raventos & M. De Los Santos (2021). The brief attitudes survey for interprofessional collaborative learning: the design, reliability, and validation of a new instrument. *Cureus*, *13*(12), Article ID: 20238. https://doi.org/10.7759/cureus.20238.

[33] S. Suherman & T. Vidákovich (2022). Adaptation and validation of students' attitudes toward mathematics to Indonesia. *Pedagogika*, *147*(3), 227–252. https://doi.org/10.15823/p.2022.147.11.

[34] M. Szczygieł (2023). The psychometric properties of the Mathematics Attitude Scale for Adults (MASA). *Current Psychology*, *42*(21), 18447–18458. https://doi.org/10.1007/s12144-022-02980-9.

[35] A. Vaisi-Raygani, A. Dalvandi, K. Nourozi, A. Ebadi, M. Rahgozar, R. Jalali, N. Salari & A. Abdi (2021). Validation of the persian version of the compassionate care assessment tool. *International Journal of Nursing Sciences*, *8*(1), 95–101. https://doi.org/10.1016/j.ijnss.2020.12.004.

[36] M. W. Watkins (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807.

[37] R. Wen & A. K. Dubé (2022). A systematic review of secondary students' attitudes towards mathematics and its relations with mathematics achievement. *Journal of Numerical Cognition*, *8*(2), 295–325. https://doi.org/10.5964/jnc.7937.

[38] S. Wijayanto (2008). *Structural Equation Modeling dengan Lisrel 8.8*. Graha Ilmu, Yogyakarta.

[39] Y. T. Wu & C. C. Tsai (2005). Information commitments: evaluative standards and information searching strategies in web-based learning environments. *Journal of Computer Assisted Learning*, *21*(5), 374–385. https://doi.org/10.1111/j.1365-2729.2005.00144.x.

[40] M. Yaşlıoğlu & D. T. Yaşlıoğlu (2020). How and when to use which fit indices? A practical and critical review of the methodology. *Istanbul Management Journal*, *88*, 1–20. https://doi.org/10.26650/imj.2020.88.0001.

[41] C. Zhang, X. Liu, T. Hu, F. Zhang, L. Pan, Y. Luo & Z. Wang (2021). Development and psychometric validity of the perioperative anxiety scale-7 (PAS-7). *BMC Psychiatry*, *21*(1), Article ID: 358. https://doi.org/10.1186/s12888-021-03365-1.