The New Educational Review

Sintha Sih Dewanti
Indonesia
Yulia Ayriza
Indonesia
Farida Agus Setiawati
Indonesia

The Application of Item Response Theory for Development of a Students' Attitude Scale Toward Mathematics

DOI: 10.15804/tner.2020.60.2.09

Abstract

Mathematics Education Study Program students should have a tendency to accept mathematics because it will support the achievement of mathematical competence in the fields of work, knowledge, and management. This study aims to construct, validate, and analyze the characteristics of attitude scale items, and determine students' attitudes towards mathematics. Aspects of student attitudes towards mathematics, were taken as namely: intrinsic motivation, enjoyment, anxiety, self-confidence, and value. The results of factor analysis show good model fit with the items measuring unidimensionality. Analysis of item characteristics was done using polytomous item response theory with a Partial Credit Model (PCM). The difficulty level of grains is at intervals of $-2.52 \leq d \leq 2.58$. Students' attitudes towards mathematics are in the interval $-0.67 \leq \theta \leq 2.36$.

Key words: *student attitudes towards mathematics, instrument construction, Item Response Theory, polytomous items, Partial Credit Model*

Introduction

Attitude is a concept associated with affective domains such as beliefs, emotions, values (McLeod, 1992; Zan, Brown, Evans & Hannula, 2006; Goldin, Rösken, & Törner, 2009). Attitudes can be described as tendencies that are associated with individuals and regularly form thoughts, feelings, and behaviors that are related to psychology (Dursun, 2015). Some researchers such as McLeod (1992) consider attitude to be a series of emotional reactions, while other researchers such as Di Martino & Zan (2010) argue that several components interact with each other in attitude forming.

In mathematics education, early studies of attitudes began to emerge in the mid-20th century (Dutton, 1951). The assumption is that it is not only cognitive factors that play a role in mathematics learning. To this end, students of the Mathematics Education Study Program as prospective mathematics educators who are outstanding, creative, superior, professional, and globally competitive, must also pay attention to non-cognitive factors that will maximize their mathematical competence. Zan, Brown, Evans & Hannula (2006) suggested that the most important research about mathematical attitudes is on understanding the mutual relationship between attitude and achievement. Several other studies show that there is a strong relationship between various domains of attitude (enjoyment of mathematics, motivation to carry out mathematical activities, confidence in mathematics, and perception of mathematical values) and mathematical achievement (Anttonen, 1969; Atkinson & Raynor, 1974; Minato, 1983; Minato & Yanase, 1984; Foire, 1999; Bouchey & Harter, 2005; Samuelsson & Granstrom, 2007).

Generally, students of the Mathematics Education Study Program have a positive tendency towards mathematics at the beginning when they choose this study program. High school graduates who choose to continue their studies in the Mathematics Education Study Program certainly already have a sense of interest in mathematics compared to other subjects. But in the course of being a student, these tendencies can change. Changes in attitude can occur at any time because of the functional value of attitude, and the process that changes this has become a major focus in social psychology (Petty, Wheeler, & Tormala, 2003). It is hoped that students' attitudes towards mathematics will not turn negative. A negative attitude is considered one of the main factors that influence poor achievement in mathematics (Schoenfeld, 1989; Zan, Brown, Evans & Hannula, 2006). Therefore, measuring the attitudes of students of the Mathematics Education Study Program towards mathematics is considered necessary so that students can continue to excel in mathematics.

Research Problem

The relationship between attitude and achievement has been widely studied using instruments that measure attitudes towards mathematics and student achievement. However, the basis for concluding the relationship between attitudes and achievements is not yet strong, and the instruments used for this goal often do not go through the process required to develop correct instruments (Ma & Kishor, 1997; McLeod, 1992).

Research Focus

The subjects of this study were students of the Mathematics Education Study Program who had different characteristics from elementary and middle school students. Therefore, it is necessary to research the development of instruments with a scale of students' attitudes towards mathematics. The purpose of this study is to: 1) construct an attitude scale; 2) validate the attitude scale; 3) analyze the characteristics of the attitude scale items; and 4) discover students' attitudes towards mathematics.

Methodology of Research

Research Sample

The subjects of this study were 227 undergraduate Mathematics Education Study Program students of whom 127 (55.95%) came from public universities and 100 (44.05%) came from private universities. The sample consisted of 119 (52.42%) male students and 108 (47.58%) female students. There are 72 (31.72%) 1st semester students, 61 (26.87%) 3rd semester students, 68 (29.96%) 5th semester students, 7 (3.08%) 7th semester students, with 19 students (8.37%) people from higher than 8th semester.

Research Instrument Development

The scale of students' attitudes towards mathematics consists of 5 aspects, which are: intrinsic motivation, enjoyment, anxiety, self-confidence, and value. Each aspect was measured by 3 or 4 indicators. Indicators of intrinsic motivation are interest, desire in career choices, and enthusiasm to pursue mathematics beyond the mandatory level. The indicators of enjoyment are enjoying mathematics, participating in mathematical discussions, challenge in solving new problems, and happy feelings in math class. Indicators of anxiety are feelings of fear, nervousness, confusion, and feelings of tension. Self-confidence indicators are that the student can perform well in mathematics, can learn mathematics easily, and can be a good problem solver. Value indicators are the usefulness of mathematics, the relevance of mathematics, and the value of mathematics in life.

The attitude scale was validated by 8 validators consisting of mathematics education lecturers, mathematics teachers, psychology lecturers, and linguists. The validator chose answers by paying attention to the appropriateness of aspects, indicators, and statements. All items are said to be valid because there was a coefficient value of Aiken's $V \geq 0.75$ with a validator of 8 people and a choice of 4 answers (Aiken, 1985).

Instrument reliability related to measurement errors is indicated by Cronbach's Alpha coefficient of 0.816 (high reliability). High reliability indicates that there is a small error in obtaining measurement results. Item reliability is indicated by Corrected Item-Total Correlation. Corrected Item-Total Correlation in each item is more than 0.3, from which it can be concluded that each item statement is reliable.

Results of Research

In analyzing data using item response theory, the first thing to do is to test the dimensions of the empirical data. The testing process is carried out by exploratory factor analysis using the principal component method. The output shows that the Kaiser-Meyer-Olkin Measure of Sampling Adequacy = 0.790 shows KMO> 0.5, which means that this study has sufficient data. Also, the significance of Bartlett's test shows that the correlation matrix is not an identity matrix, so the data form a correlation matrix with a close relationship between variables. Then based on anti-images correlation, the Measures of Adequate Sampling (MSA) was > 0.5 so all data are eligible for factor analysis.

The number of factors can be determined by selecting factors that have an Eigenvalue greater than 1. Based on the Eigenvalue, 5 factors were formed. This can also be seen in the scree plot to determine the exact number of components, presented in Figure 1.

To facilitate interpretation, a rotation is carried out to obtain a simpler loading structure. The Varimax method was chosen to get a loading structure that has a strong relationship with only one factor. The results of the factor rotation are presented in Table 1. The results of the factor rotation indicate that each item measures 1 dimension so that subsequent data analysis uses unidimensional item response theory.

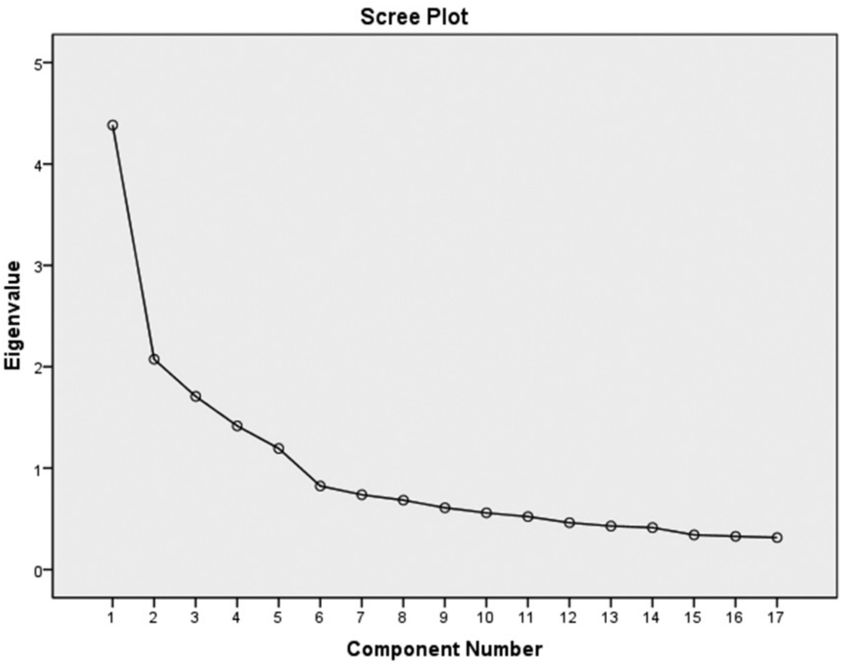

Figure 1. Scree plot of the main component analysis

Table 1. Rotated component matrix

	Component				
	1	2	3	4	5
A1				0.742	
A2				0.801	
A3				0.728	
B1	0.825				
B2	0.788				
B3	0.682				
B4	0.774				
C1		0.742			
C2		0.686			
C3		0.764			
C4		0.715			
D1					0.731
D2					0.586
D3					0.815
E1			0.822		
E2			0.864		
E3			0.757		

The validity of the construct theory of attitudes towards mathematics is evidenced by confirmatory factor analysis. The t-value output is presented in Figure 2.

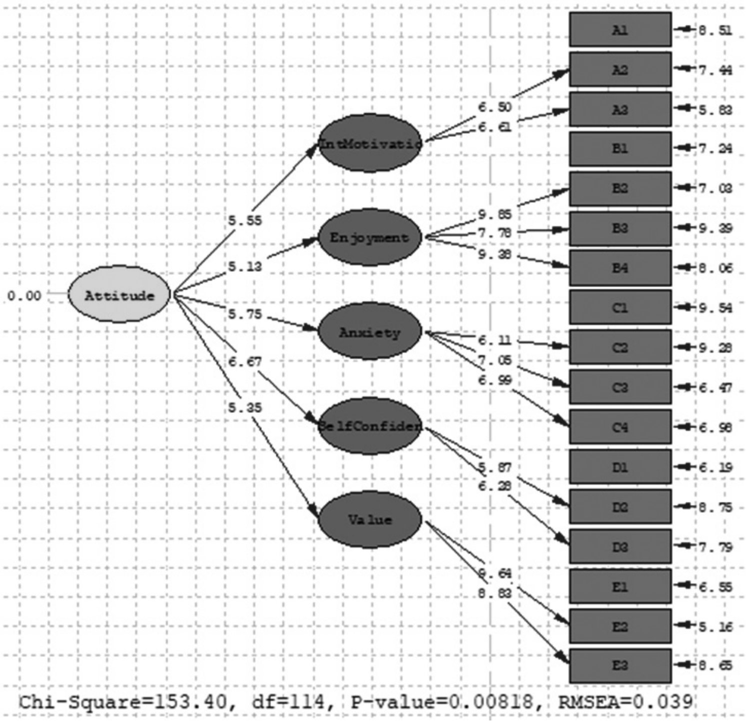

Figure 2. T-value of the construct theory of attitude towards mathematics

Figure 2 shows all the paths are significant, but the p-value = 0.008 for the model compatibility test is not significant. For this reason, it is necessary to pay attention to modification indices to find suggestions for improving the model. After modification, the standardized solution is presented in Figure 3.

To find out the characteristics of polytomous items, researchers used a Partial Credit Model (PCM) with the help of R software. The first step is to test the suitability of the model if the data analysis can use PCM. The results of the model compatibility test are presented in Table 2.

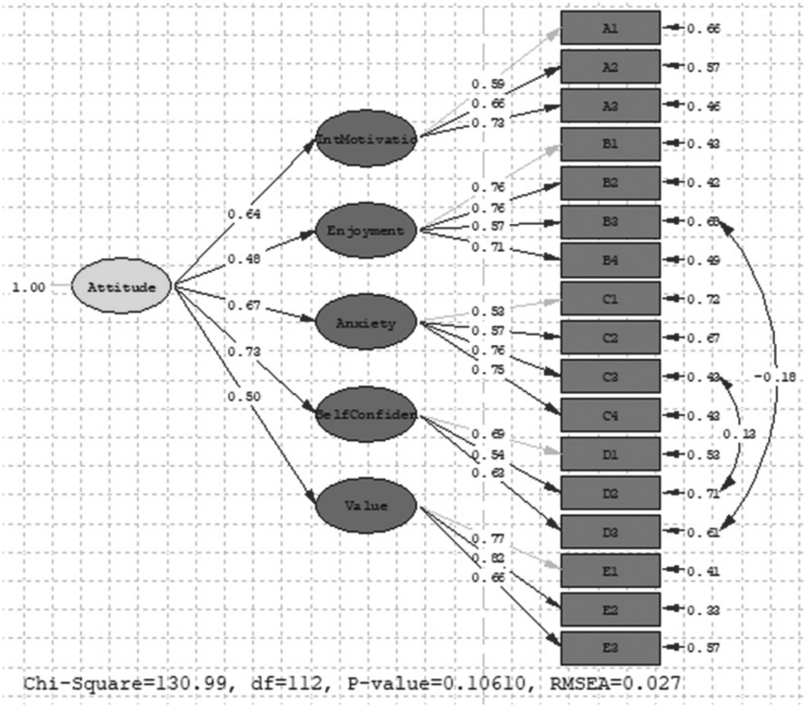

Figure 3. The standardized solution of the construct theory of attitude to mathematics after modification

Table 2. Model Match Test Output

Item	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
A1	231.569	226	0.385	1.02	1.013	0.25	0.18
A2	225.346	226	0.500	0.993	0.995	-0.05	-0.03
A3	201.128	226	0.882	0.886	0.87	-1.24	-1.46
B1	217.182	226	0.651	0.957	0.982	-0.45	-0.17
B2	194.622	226	0.936	0.857	0.878	-1.66	-1.43
B3	231.064	226	0.394	1.018	1.034	0.23	0.4
B4	199.263	226	0.900	0.878	0.888	-1.51	-1.37
C1	232.509	226	0.369	1.024	1.044	0.3	0.53
C2	221.735	226	0.568	0.977	1.013	-0.22	0.17
C3	183.638	226	0.982	0.809	0.818	-2.41	-2.39
C4	185.679	226	0.977	0.818	0.828	-2.17	-2.06
D1	204.705	226	0.842	0.902	0.91	-1.19	-1.11
D2	231.335	226	0.390	1.019	1.001	0.22	0.04
D3	256.819	226	0.078	1.131	1.071	1.31	0.76
E1	207.82	226	0.802	0.916	0.941	-0.91	-0.64
E2	214.295	226	0.702	0.944	0.974	-0.64	-0.28
E3	220.587	226	0.589	0.972	0.993	-0.26	-0.04

The above output shows that there was a p-value> 0.05 for each item, which means that all items are fit to be analyzed using PCM which is a 1PL polytomous model that considers the power of different items constant. Also, it is necessary to see the person fit to determine the characteristics of respondents who are suitable for this instrument to be used on. Of 227 respondents, there were 19 people (8.37%) who did not fit the instrument construct.

The threshold is the intersection for each category that shows the minimum ability to be able to reach certain points. As there are 5 answer choices on the attitude scale, then there will be a maximum of 4 thresholds. The threshold for each item is presented in Table 3.

Table 3. The threshold for each item

Item	Location	Threshold 1	Threshold 2	Threshold 3	Threshold 4
A1	0.326	-0.358	-0.336	0.687	1.309
A2	0.352	-0.604	-0.256	0.747	1.520
A3	0.375	-0.691	-0.028	0.478	1.743
B1	0.448	-0.054	-0.717	0.893	1.671
B2	0.305	-1.193	0.222	0.509	1.681
B3	0.445	0.047	-0.701	0.956	1.477
B4	-0.010	-1.850	-0.669	0.855	1.624
C1	0.287	-0.707	-0.332	0.807	1.379
C2	0.401	-0.417	-0.481	0.600	1.901
C3	0.635	0.044	0.552	1.308	NA
C4	0.230	-0.877	-0.297	0.723	1.369
D1	0.859	0.201	0.465	1.910	NA
D2	0.496	-0.017	0.256	0.322	1.424
D3	0.514	0.183	-0.279	0.488	1.665
E1	0.123	-1.401	-0.022	0.339	1.577
E2	0.437	-0.534	-0.314	1.067	1.526
E3	0.224	-0.651	-0.154	0.324	1.376

Of the 17 items, there are 15 items that have 4 thresholds and 2 items that have 3 thresholds. Generally, threshold 1 < threshold 2 < threshold 3 < threshold 4, because in response to "strongly agree" it should have a higher tendency of approval than responding to "agree". There are 13 items that have a threshold that is getting bigger from one category to the next. The Characteristic Curve items of several items that have a threshold like this are presented in Figure 4.

PCM does not require the steps to complete the test items to be sequential, nor for them to have the same difficulty (De Ayala, 1993). This resulted in the fact that

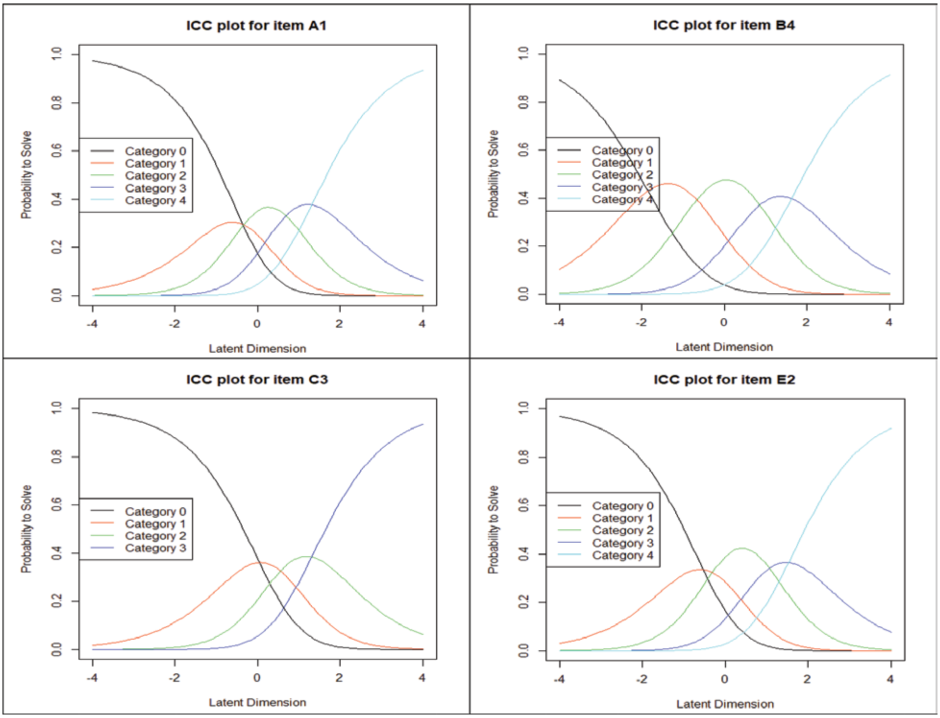

Figure 4. Items with threshold 1 < threshold 2 < threshold 3 < threshold 4

the threshold for PCM scoring from one category to the next is not always greater. There are 4 items, namely B1, B3, C2, and D3 which have a threshold 2 < threshold 1 < threshold 3 < threshold 4, as presented in Figure 5.

The level of difficulty in PCM is the level of difficulty answering the upper-level category for each item and is presented in Table 4.

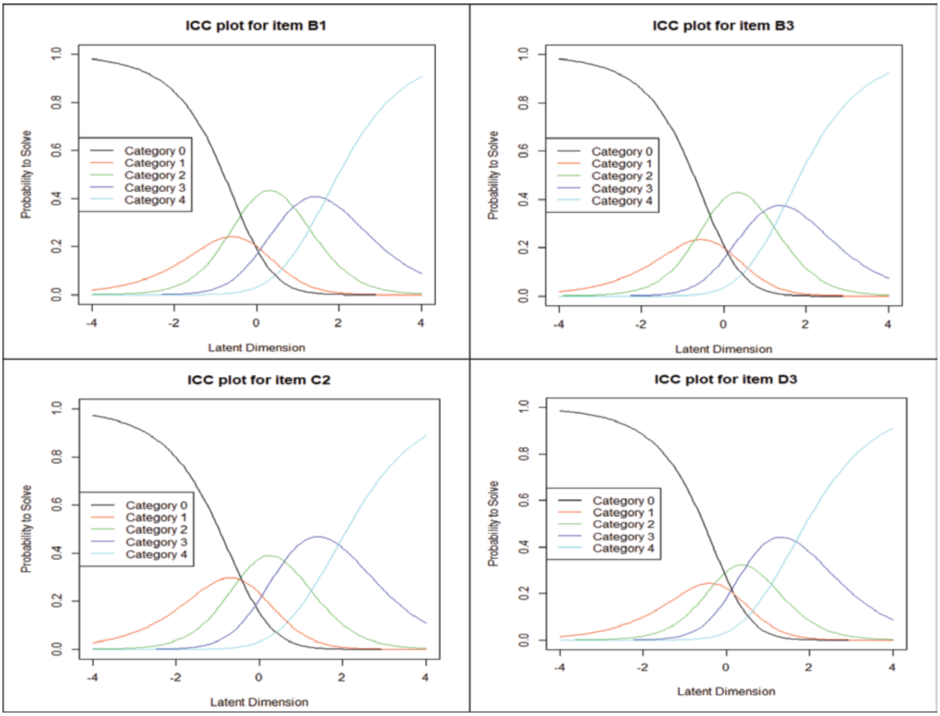

Figure 5. Items with threshold 2 < threshold 1 < threshold 3 < threshold 4

Table 4. Difficulty level for each category in each item

Itm.Ctgr	Diff.	Itm.Ctgr	Diff.	Itm.Ctgr	Diff.	Itm.Ctgr	Diff.	Itm.Ctgr	Diff.
A1.c2	-0.694	B1.c3	0.122	B4.c4	-0.040	C4.c2	-1.173	D3.c4	2.057
A1.c3	-0.007	B1.c4	1.793	C1.c1	-0.707	C4.c3	-0.450	E1.c1	-1.401
A1.c4	1.302	B2.c1	-1.193	C1.c2	-1.040	C4.c4	0.919	E1.c2	-1.423
A2.c1	-0.604	B2.c2	-0.971	C1.c3	-0.233	D1.c1	0.201	E1.c3	-1.083
A2.c2	-0.860	B2.c3	-0.463	C1.c4	1.147	D1.c2	0.666	E1.c4	0.493
A2.c3	-0.113	B2.c4	1.218	C2.c1	-0.417	D1.c3	2.576	E2.c1	-0.534
A2.c4	1.408	B3.c1	0.047	C2.c2	-0.898	D2.c1	-0.017	E2.c2	-0.848
A3.c1	-0.691	B3.c2	-0.654	C2.c3	-0.299	D2.c2	0.239	E2.c3	0.220
A3.c2	-0.719	B3.c3	0.302	C2.c4	1.602	D2.c3	0.561	E2.c4	1.746
A3.c3	-0.241	B3.c4	1.779	C3.c1	0.044	D2.c4	1.985	E3.c1	-0.651
A3.c4	1.502	B4.c1	-1.850	C3.c2	0.596	D3.c1	0.183	E3.c2	-0.805
B1.c1	-0.054	B4.c2	-2.519	C3.c3	1.904	D3.c2	-0.096	E3.c3	-0.480
B1.c2	-0.771	B4.c3	-1.664	C4.c1	-0.877	D3.c3	0.392	E3.c4	0.896

The distribution of difficulty levels is presented in Figure 6.
The characteristics of students is presented in Figure 7.

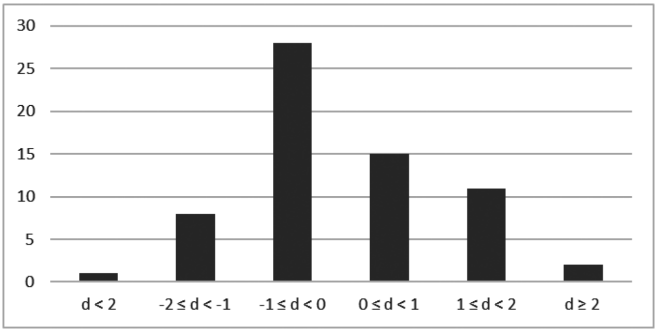

Figure 6. The difficulty level of categories for each item

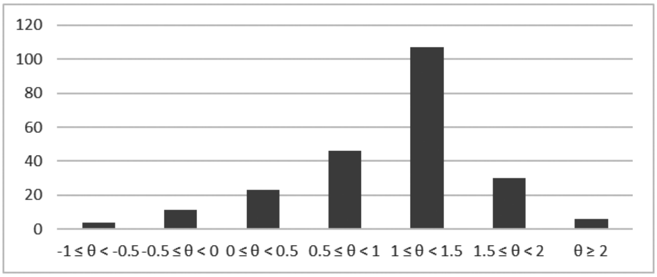

Figure 7. Characteristics of students

Discussion

This research aims to get the right instruments to measure the attitudes of students of the Mathematics Education Study Program towards mathematics. The development of the instrument was carried out following 12 steps from developing an overall plan to reporting the results (Downing, 2006: 4-23). The purpose of a test is important in determining what content is appropriate, and what is not, what is suitable for the test, in choosing the format of the test item and how to respond, and in planning data collection efforts to evaluate the validity and interpretation of test scores, and the quality of other technical tests (Linn, 2006: 28).

Good instruments must be made in such a way that they are easy to use and the results are accurate for the purpose of the assessment. Validity refers to how far the test score provides accurate information for making decisions based on that score (Messick, 1995: 13; Amy, 1999: 3). A statement of validity cannot be made about the instrument, but must refer to conclusions made when using scores from the scale (Finch & French, 2019: 172). Therefore, the validation process involves gathering evidence to show the scientific basis for interpreting the score as planned. Validity is support for the interpretation of test scores against the intended use of tests based on evidence and theory (Mardapi, 2017: 32).

Validity is defined as a framework of three interrelated aspects, namely criteria validity, content validity, and construct validity (Croker & Algina, 1986: 217; Amy, 1999: 3-5). Criteria validity means how far the score from the assessment or test is associated with certain criteria. Content validity means that the assessment should include a representative content domain. Construct validity means how far a test score correlates with the theoretical characteristics of the measured attribute. The third proof of validity is carried out in this study, namely through the validator assessment analyzed with Aiken's coefficient, Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA). Proof of validity is not compartmentalized but is used together with other measures to build arguments that show to what extent valid conclusions can be made with scores derived from these instruments (Finch & French, 2019: 172).

The validity of the contents is determined using the agreement of experts in the field of study. The results of validation show all items are valid because of the coefficient value of Aiken's V ≥ 0.75 with 8 validators, 4 rating categories, and a significance level of 0.05 (Aiken, 1985). Although this type of validity assessment has some interesting theoretical properties, this validity will have problems with subjectivity (Kane, 2006). Construct validation rests on the foundation of theoretical expectations, and the extent to which the scale fits these expectations. In other words, the construct validity addresses the extent to which the instrument behaves according to how the theory suggests it should. Proof of construct validity can be demonstrated by analyzing results of empirical measurement, namely by Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) (Retnawati, 2017: 20). The EFA results show that this instrument explains 63.38% of students' attitudes towards mathematics with 5 factors. The rotation results grouped the items into 5 factors by the theory that attitude consists of 5 aspects, namely: intrinsic motivation, enjoyment, anxiety, self-confidence, and value.

The CFA results show that each path is significant from both the latent variable (exogenous variable to endogenous variable) and the observed variable, but the p-value = 0.008 <0.05. This shows that the model compatibility test is not significant, so it needed to be modified. Modification was done by correlating the error covariance between the 2 observed variables, namely B3 – D3, and C3 – D2. This meant a review of the attitude scale indicators needed to be done. Indicator B3 is "a challenge to solve a new problem", while indicator D3 is "can be a good problem solver". The two indicators measure related things, namely attitudes towards problem-solving. To be a good problem solver, students must have the challenge to solve new problems. The C3 indicator is "confusion", while the D2 indicator is "can learn mathematics easily". These two indicators also measure related things. Students who can learn mathematics easily will certainly not experience confusion. This relationship can also be reversed, students who do not experience confusion will be able to learn mathematics easily.

A Partial Credit Model (PCM) is very suitable for analyzing multipoint scale personality response scales (Masters & Wright, 1996, Embretson & Reise, 2000). The results of the analysis showed PCM fit, meaning that the items on the attitude scale followed the 1-PL (Parameter Logistic) model by taking into account the level of difficulty and assuming constant power differences (Masters, 1982). Generally, higher category scores indicate higher ability than lower category scores. PCM is suitable for items that are scored in a tiered category, but the difficulty index in each step does not need to be ordered. The response items in this study can be correct at certain steps but can be wrong at other steps.

The threshold shows the meeting point of two category probability lines in one item. The individual's probability of responding to category x at this stage is the difference between the level trait (θ) and the threshold (δ_{ij}). In other words, the category intersection parameter can be considered as the level of difficulty of the stages relating to the transition from one category to the next category, and there is a difficulty step *mi* (intersection) for items with *mi* + 1 answer categories (Embretson & Reise, 2000). The value δ_{ij} does not always have to be sequential in item i because it is a relative magnitude of two adjacent probabilities (De Ayala, 1993; Muraki, 1992). The threshold can also be interpreted as a point on a latent nature scale, where for two consecutive categories the response curves intersect. The threshold is a point where two categories have the same probability to be chosen because of the associated level of a trait (Linacre, 2006).

Conclusions

Based on the theory, a scale to measure students' attitudes toward mathematics was compiled consisting of 5 aspects, namely: intrinsic motivation, enjoyment, anxiety, self-confidence, and value. The results of factor analysis show that the attitude scale produces a model that fits the requirements for a unidimensional measure. The results of the analysis of grain characteristics identified 65 categories of 17 items with a level of difficulty (d) at intervals of $-2.52 \leq d \leq 2.58$, and most were located at intervals of $-1 \leq d < 0$ of 43.08%. Of 227 students, there were 19 (8.37%) people who did not fit the construct of this instrument. Student attitude toward mathematics is described as student ability (θ) located at intervals of $-0.67 \leq \theta \leq 2.36$, and most student's ability is at interval $1 \leq \theta < 1.5$ (47.14%). This instrument can be used in class assessments or large-scale assessments. The results of the analysis of item characteristics can show the level of difficulty in each item of each category in detail, so the instrument can be used to measure student attitudes based on their ability level.

References

Aiken, L.R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*, 131–142.

Amy, B. (1999). *Traditional and modern concepts of validity*. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation.

Anttonen, R.G. (1969). A longitudinal study in mathematics attitude. *The Journal of Educational Research*, 62, 467–471.

Atkinson, J.W., & Raynor, J. (1974). *Motivation and achievement*. Washington, DC: Winston.

Bouchey, H.A., & Harter, S. (2005). Reflected appraisals, academic perceptions, and math/science performance during early adolescence. *Journal of Educational Psychology*, *97*(4), 673–686.

Croker, L., & Algina, J. (1986) *Introduction to classical and modern test theory*. New York: Holt Rinehart and Winston INC.

De Ayala, R.J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development*, *25*, 172-189.

Di Martino, P., & Zan, R. (2010). 'Me and maths': Towards a definition of attitude grounded on students' narratives. *Journal of Mathematics Teacher Education*, *13*(1), 27–48.

Downing, S.M. (2006). Twelve steps for effective test development. In S.M. Downing, & T.M. Haladyna (Eds.). *Handbook of test development*. New Jersey: Lawrence Erlbaum Associates, Inc.

Dursun, Ş. (2015). Investigation of high school students' attitude and anxiety levels towards

Mathematics in terms of some variables. *Educational Research and Reviews*, *10*(13), 1773–1780.

Dutton, W.H. (1951). Attitudes of prospective teachers toward arithmetic. *Elementary School Journal*, 42, 84-90.

Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Finch, W.H., & French. B.F. (2019). *Educational and psychological measurement*. New York & London: Routledge, Taylor & Francis.

Foire, G. (1999). Math-abused students: Are we prepared to teach them? *Mathematics Teacher*, *92*, 403–407.

Goldin, G.A., Rösken, B., & Törner, G. (2009). No longer a hidden variable in mathematics teaching and learning processes. In J. Maaß, W. Schlöglmann (Eds.), *Beliefs and attitudes in mathematics education: New research results* pp 1–18. Rotterdam: Sense

Kane, M.T. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (4th ed.), (pp. 17–64). Washington, DC: The National Council on Measurement in Education & the American Council on Education.

Linacre, J.M. (2006). *Winstep: Rasch- model computer programs*. Chicago: Winsteps.com.

Linn, R.L. (2006). The standards for educational and psychological testing: Guidance in test development. In S.M. Downing, & T.M. Haladyna (Eds). *Handbook of test development*. New Jersey: Lawrence Erlbaum Associates.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, *28*, 26–47.

Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan* [Measurement, assessment, and evaluation of education]. Yogyakarta: Parama Publishing.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47* (2), 149–174.

Masters, G.N., & Wright, B.D. (1996). The partial credit model. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory*. New York: Springer.

McLeod, D.B. (1992). Research on affect in mathematics education: A reconceptualization. In D.A. Grouws (Ed.), *Handbook of research on mathematics learning and teaching* (pp. 575–596). New York: Macmillan.

Messick, S. (1995). Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practice*, *14*(4), 5-8.

Minato, S. (1983). Some mathematical attitudinal data on eighth-grade students in Japan measured by a semantic differential. *Educational Studies in Mathematics*, *14*, 19–38.

Minato, S., & Yanase, S. (1984). On the relationship between students attitude towards school mathematics and their levels of intelligence. *Educational Studies in Mathematics*, *15*, 313–320

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16* (2), 159-176.

Petty, R.E., Wheeler, S.C., & Tormala, Z.L. (2003). Persuasion and attitude change. In T.

Millon & M. Lerner (Eds.), *Handbook of psychology: Personality and social psychology* (pp. 353-382). Hoboken, NJ: Wiley.
Retnawati, H. (2017). *Validitas reliabilitas dan karakteristik butir: Panduan untuk peneliti, mahasiswa, dan psikometrian* [Reliability validity and item characteristics: A guide for researchers, students, and psychometrics]. Yogyakarta: Parama Publishing.
Samuelsson, J., & Granström, K. (2007). Important prerequisite for students' mathematical achievement. *Journal of Theory and Practice in Education*, *3*(2), 150–170.
Schoenfeld, A. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education, 20*(4), 338–355
Zan, R., Brown, L., Evans, J., & Hannula, M.S. (2006). Affect in mathematics education: An introduction. *Educational Studies in Mathematics*, *63*,113–121.