JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

Analysis of CTT and IRT on Non-Test Measurement Instruments: Mathematical Mindset Scale

Priarti Megawanti[1], Nurimani[2], Meyta Dwi Kurniasih[3]

[1]Department of Mathematics Education, Universitas Indraprasta PGRI, Indonesia
[2]Department of Mathematics Education, STKIP Kusuma Negara, Indonesia
[3]Department of Mathematics Education, FKIP, Universitas Muhammadiyah Prof. Dr. HAMKA, Indonesia
*priartimegawanti@gmail.com

ARTICLE INFO

Keywords:
mathematical mindset
item response theory
classical test theory
non-test instrument development

ABSTRACT

Purpose – This study aims to determine empirically the appropriate number of items that meet all item selection criteria, from classical test theory to item response theory.

Methodology – This study used the Mathematical Mindset Scale (MMS) as an instrument for a non-test. It was conducted on 259 Mathematics Education students in Indonesia to determine the appropriate number of items that meet all item selection criteria, from classical test theory to item response theory. This quantitative research has been analyzed, including sample adequacy tests, construct validity and reliability, the goodness of fit, model suitability to polytomous data, assumption tests (unidimensional, invariance, local independence), characteristic curve graphs, information function graphs, and standard error graphs.

Findings – After several trials, 11 items passed the classical tests and item response theory criteria. It is also known from this study that non-test instruments can be less than 20 as long as they meet the analysis criteria at each stage.

Contribution – This research contributes to helping increase insight and knowledge for researchers in the fields of education and psychology, especially those interested in non-test measurement tools and IRT and classical analysis.

Received 8 December 2024; Received in revised form 15 December 2024; Accepted 12 June 2025
Jurnal Eduscience (JES) Volume 12 No. 3 (2025)
Available online xx June 2025
©2025 The Author(s). Published by LPPM Universitas Labuhanbatu. This is an open-access article under the **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY - NC - SA 4.0)**

INTRODUCTION

Not all researchers have enough time to develop the instruments used in their research. Some researchers want to modify or adapt from articles that are kind enough to permit their instruments to be used by other researchers. However, some researchers want to be able to create their instruments. In the end, questions such as how many items are right for a non-test instrument are something that novice researchers often question. Therefore, an empirical study of non-test measuring instruments and the item selection process is needed based on the analysis of classical test theory and Item Response Theory (IRT). In research using non-test measuring instruments, respondents are tired of filling out questionnaires, which is challenging (Rizdanti & Akbar, 2022). This results in the expected answers being inconsistent with the reality felt or experienced by respondents. Of course, this can have an impact on the research results.

Although it should not be underestimated, in its preparation, non-test measuring instruments are generally attempted to be made with sentences as simple as possible (Aiken, 1996; Azwar, 1995, 2024). Thus, respondents with any level of ability can answer the statement items according to the values of life they adhere to or as they experience directly. Respondents are expected to answer honestly when filling out non-test measuring instruments so that researchers can conduct the analysis process and provide accurate conclusions. If respondents answer carelessly or dishonestly, the results cannot describe the reality. In several studies, some respondents tend to answer questionnaires because they feel there is no benefit related to the research being conducted (Arthur et al., 2021).

Generally, instrument development research, both tests and non-tests, outlines the steps that must be taken before the item selection stage. The first stage is to collect theories to compile constructs and determine dimensions or aspects of the latent variables to be studied later (Azwar, 1994; Furr, 2014; Mardapi et al., 2011). Researchers must prepare as many items as they can compile based on the constructs and aspects they have obtained at the theory collection stage. The large number of items aims to anticipate if, in one dimension or aspect, all items are invalid at the construct validation stage. Content validation by experts is also an important stage to minimize items with poor wording (Almanasreh et al., 2019; Azwar, 2024; Lynn, 1986).

After the items have been improved according to the experts' suggestions, item testing is carried out during the trial stage (Aka, 2019; Gall et al., 2003). If, at the trial stage, there are still many remaining items of the instrument, some respondents may get bored, so they tend to fill it out quickly. Researchers can overcome this obstacle by explaining to respondents that the questionnaire can be answered periodically with a specific time tolerance range. This method can be one solution so respondents can fill out the questionnaire objectively. Conversely, if, at the trial stage, the researcher prepares too few statement items, several obstacles will be faced later. One of them is that respondents may complete the questionnaire faster with a few items if the statements are easy for respondents to understand. However, there may be several invalid items at the construct validation stage, and researchers will have to re-arrange the statement items from the beginning if too many items have to be removed or replaced. The number of items should be calculated according to the dimensions or aspects to be measured (Hutajulu et al., 2021).

One of the non-test measurement tools is the Mathematical Mindset Scale (MMS), which consists of mindset and mathematical mindset. Several references list mindset and mathematical mindset instruments with different numbers of items. This often confuses researchers who want to use the measurement tool directly without adapting or modifying it. The mindset measurement tool, better known as the Implicit Theory of Intelligence Scale (ITIS), was developed by Dweck and colleagues (2022b) consisting of 10 items to make it easier for respondents to fill out the questionnaire so that it can shorten the time for interpreting the questionnaire and further handling related to the results of the interpretation. In addition, Rammstedt et al. (2021) compared mindset measurement tools consisting of 3 and 1 item. A short instrument containing only 1 item "for extra quick assessment" was used during PISA 2018 (OECD, 2020). The research results conducted by Rammstedt et al. showed that the mindset measurement tool with three items had better reliability than 1 item. However, if there is not much measurement time available, then a mindset measuring instrument with 1 item can be considered.

Meanwhile, MMS was developed by Im and Park (2023) by compiling nine items based on seven positive norms proposed by Boaler (2016). In addition, Saefudin et al. (2023) developed a mathematical mindset measuring instrument consisting of 20 items. This study aims to provide steps for analyzing and selecting items based on classical test theory analysis and IRT. With empirical data from MMS, a non-test measuring instrument, it is hoped that it can help novice researchers research developing measuring instruments. Thus, each researcher will later be able to find answers regarding the number of items appropriate for the research being carried out.

Before discussing the analysis of non-test items, because this study uses MMS as a measuring instrument that is being tested, it is necessary to discuss the mathematical mindset. As a new concept proposed by the OECD in PISA 2018, mindset and mathematical mindset are important. The concept of a mathematical mindset explains how mathematics is viewed not just as numbers but as something that can be understood widely by anyone who studies it (Rahmah, 2018). A mathematical mindset is how students are introduced to the fact that mathematics can be events in the universe or how humans solve their daily problems (Boaler, 2016). A mathematical mindset also implicitly explains adaptive and stable human abilities related to mathematical skills, intelligence, and talent (Saefudin et al., 2023). The more often someone makes mistakes in learning something, the more the nerves in the human brain will increase (Boaler, 2016; Carol S Dweck, 2008). Likewise, when someone studies mathematics, every process, in the form of difficulties and failures, will help them understand mathematics better. However, some people believe that not everyone possesses mathematical intelligence. Some believe that mathematical intelligence is something that is fixed.

A person gifted with mathematical genius will not need to study or try hard to understand mathematics (Chestnut et al., 2018). If someone shows that they work hard, it will be considered clear evidence that they are not talented (Carol S. Dweck, 2022a). Mathematical intelligence is considered to be something that is genetically inherited (Carol S Dweck, 2008; Mills & Mills, 2018; Rammstedt et al., 2021; Saefudin et al., 2023). Some people believe this will avoid math problems they cannot do because it will make them look stupid (Ayebo & Mrutu, 2019; Carol S. Dweck, 2022b; Saefudin et al., 2023). They can be classified as having a fixed math mindset (Daly et al., 2019).

They are afraid of being considered stupid and will choose answers they think they can do. They are terrified when asking questions or failing exams because looking stupid is scary for them (Carol S. Dweck, 2022a). If educators and parents have a fixed mindset, they tend to differentiate their students and children with the label of genius or vice versa (Heyder et al., 2020; Ramirez et al., 2018). They will judge from the speed and results rather than the length of the process taken by the student (Cutler, 2020). Without realizing it, what they do fosters anxiety in learning mathematics, decreased motivation to learn, and unhealthy competition in schools (Carol S. Dweck, 2022b). Students with a fixed math mindset will not easily accept criticism and input from others. In addition, they view failure as tangible evidence that they are not talented (Mills & Mills, 2018). Meanwhile, educators with a fixed mindset tend to take all criticism and input from others to heart, making it difficult for them to get up and learn from their mistakes because they see the criticism and input as a sign that they are bad and will always be that way (Meierdirk & Fleischer, 2022).

On the other hand, there are groups of people who will choose complex math problems because they feel challenged to be able to solve them (Shen et al., 2016). They do not care about grades and consider mistakes in learning math as a reason for them to continue learning. They think that intelligence can be developed no matter how difficult it is and no matter how stupid they are at first (Boaler, 2016; Carol S Dweck, 2008; Mills & Mills, 2018; Saefudin et al., 2023). They believe that everyone starts from not being able to. Likewise, mathematical intelligence can be mastered as long as they are willing to continue learning and practicing continuously (Carol S. Dweck, 2022a; Mills & Mills, 2018; Suh et al., 2011). They can be said to have a growth math mindset (Daly et al., 2019).

Mathematics educators should have a growth mindset (Meierdirk & Fleischer, 2022). They will be open to accepting students' shortcomings and mistakes so that students will feel comfortable making mistakes but still want to try. Educators with a growth mindset tend to have high standards for student achievement and help them achieve it (Carol S. Dweck, 2022a). They are willing to listen to their students rather than immediately judging them as failures (Meierdirk & Fleischer, 2022). They teach that correcting mistakes is the best way to become an expert. These educators will encourage students to ask lots of questions and understand that failure is not something to be ashamed of (Stohlmann, 2022).

Based on the explanation above, a mathematical mindset can be explained as a person's perspective regarding mathematics. This perspective can be divided into two, i.e., growth math mindset and fixed math mindset. Someone who believes that mathematics is something anyone can master by studying seriously has a growth math mindset. Meanwhile, someone who believes that mathematics is a talent or intelligence inherited without making much effort has a fixed math mindset. Based on the results of research from several research results from Dweck (2022b), Saefudin et al. (2023), and Chen et al. (2021), several aspects that form a mathematical mindset can be formulated, namely (1) Belief in mathematical ability, (2) Effort, (3) Resilience, (4) Focus, (5) Challenge, (6) Learning from Critics, (7) Learning from Mistakes, and (8) Learning from Others.

METHODOLOGY

Research Design and Instrument

This study is an empirical study. Since this study aims to determine the number of items on a non-test measuring instrument that empirically meets the criteria, the design of this study is instrument development with classical test theory analysis and IRT. The measuring instrument used in this study is the MMS, which has 43 items. The MMS is a non-test measuring instrument that modifies the Implicit Theory of Intelligence Scale (ITIS) developed by Dweck and colleagues (Carol S. Dweck, 2022b). The MMS in this study uses a Likert scale with five answer choices and includes an undecided option. This differs from the ITIS, which uses six answer choices and does not include undecided, undecided, or neutral options. In addition, the ITIS only consists of 1, 3, 6, 8, and 10 items, while the MMS in this study initially consisted of 43 items with eight aspects. The answer choices for positive statements start from strongly disagree with a score of 1, disagree with a score of 2, undecided with a score of 3, agree with a score of 4, and strongly agree with a score of 5. However, for negative statements, a score of 1 is given for the answer choice strongly agree, a score of 2 for the answer choice agree, undecided still gets a score of 3, a score of 4 for disagree, and a score of 5 for the answer choice strongly disagree. Before the data is processed, the respondent's answer choices for negative statements must be adjusted according to their scores to avoid errors when analyzing later.

Participant and Data Collection

This study used active students who chose the Mathematics Education study program as respondents. Two hundred fifty-nine students were willing to fill in all MMS items. All respondents studying at state and private universities in Jakarta had an equal opportunity to be selected randomly. This number is a sample of the total population of Mathematics Education students in Jakarta recorded in PDDikti 2023 of 2,222 people. Although the number of samples was only 259 people, this number met the criteria for the number of samples at the trial stage, according to Azwar (2024), namely the number of items to be tested multiplied by 5 or 10. Therefore, the number of items to be tested is 43, and the result of the multiplication of 5 is 215 respondents. Furthermore, the answers to these 259 students will be analyzed to obtain MMS items that meet the criteria.

Data Analysis

The analysis starts with the Kaiser-Meyer-Olkin (KMO) test to determine the adequacy of the number of respondents who fill out the research instrument. Azwar (2024) explains that the number of samples can be determined from the number of instrument items, which is 5 to 10 times the number. Because the number of MMS items is 43, a minimum sample of 215 people is needed. If the KMO value exceeds 0.5, the analysis can be continued to the Confirmatory Factor Analysis (CFA) test (Arlinwibowo et al., 2024; Retnawati, 2016). Conversely, if it is not above 0.5, the researcher must add more samples to meet the KMO test requirements.

After adequate KMO value, the analysis is continued by conducting a CFA test. The CFA test is conducted to determine the loading factor value of each item. The loading factor value limit, according to several researchers, can vary. Santoso (2014) explained that the loading factor generally ranges from 0.7 to 0.4. However, the criteria of 0.7 or 0.5 are considered too difficult to achieve (Azwar, 2024). Some others consider 0.4 an acceptable loading factor limit (Joseph F. Hair et al., 2010; Retnawati, 2016). This study used a loading factor of 0.4 in the first trial, 0.4,9 in the second trial, and 0.5 in the third. This ensures that the selected items meet the classical analysis criteria and IRT. Thus, in the second trial, items with a loading factor value of less than 0.4 must be discarded or replaced so that later, only items with a loading factor value above 0.49 remain. The item elimination stage will continue until valid items are obtained and meet the Goodness of Fit (GoF) criteria. Furthermore, items that meet the loading factor limit value requirements are included to estimate reliability values. Reliability testing uses Construct Reliability (CR). The minimum value criteria for CR is 0.7 (Joe F. Hair et al., 2011).

In GoF testing, there are 25 criteria. However, in empirical research practice, these criteria do not all have to be met (Haryono, 2016; Simanjuntak & Hamimi, 2019). According to Hair et al. (2010), if researchers only use 4 to 5 GoF criteria, it is considered sufficient to assess the feasibility of a model, with the condition that each GoF criterion, namely absolute fit indices, incremental fit indices, and parsimony fit indices are represented. A GOF test is needed in the CFA test to determine whether the Model being tested is fit. The most frequently used GOF test criteria always consider the representation of the three GOF groups: absolute, incremental, and parsimonious. GoFs that are included in the absolute group are Chi-Square (X2), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI). Absolute GoF provides information about how good the Model is overall. Meanwhile, incremental GoF measures the extent to which the tested Model is better than the initial Model, where the initial assumption is that the Model has no relationship between variables. Some incremental GoFs are Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Non-Normed fit Index (NNFI), Normed Fit Index (NFI), Incremental Fit Index (IFI), and Relative Fit Index (RFI). The last is parsimonious GoF, which balances Model fit and simplicity. This criterion aims to obtain a simple model. Some parsimonious GoF are Parsimonious Normed Fit Index (PNFI), Parsimonious Comparative Fit Index (PCFI), Parsimony Goodness-of-Fit Index (PGFI), and Chi-Square/df (X2/df). Table 1 shows the cutoff values of several GoF testing criteria.

Table 1. Goodness of Fit Criteria

No.	**Criteria**	***Cut off Value***
1	X^2 (Chi-Square)	< a f (smaller than X^2 table)
2	Significance probability (p)	> 0,05
3	GFI	> 0,90
4	AGFI	> 0,90
5	CFI	> 0,90
6	NNFI/TLI	> 0,90
7	RMSEA	< 0,08
8	RMR	< 0,05

Source: (Haryono, 2016; Santoso, 2014)

FINDINGS

Dweck's concept of a growth mindset inspired researchers in education and psychology to prove that mindset significantly influences human potential development. After Dweck compiled the ITIS as a standard instrument to measure a person's mindset, several researchers used the ITIS instrument with 8 items (Apiola & Sutinen, 2020; Limeri et al., 2020; Midkiff et al., 2018). Meanwhile, Rammstedt et al. (2021) used the ITIS instrument by comparing the reliability estimates of the ITIS, consisting of 3 and 1 items.

In addition to researchers who use ITIS, some researchers, like Chen et al. (2021) are trying to develop ITIS. Their research produced 18 items that met the Confirmatory Factor Analysis and Goodness of Fit criteria. Research that developed ITIS into a mathematical mindset was conducted by Saefudin et al. (2023) and produced 20 items. However, this study did not explain the steps for developing and selecting items to produce the 20 items. Another study that developed the mathematical mindset scale was conducted by Im and Park (2023), who compiled nine items based on the seven positive norms proposed by Boaler (2016). Based on several studies that have been conducted related to mindset and mathematical mindset, this study attempts to develop ITIS and explain the steps for selecting items to produce items that meet the criteria classically and IRT.

Based on the analysis results, there are 3 MMS models. The first Model consists of 43 items and eight aspects. The second Model consists of 25 items and 7 aspects. The last Model is MMS, which has 13 items and five aspects. The overall KMO result in the first Model is 0.8. The CFA analysis results produce a Chi-Square (X2) of 1,610.177 and df 849,000, which is X2/df, then the result is 1.8966 and not more than 2, which means fit. In addition, the RMSEA value is 0.059, which means less than 0.08, and this can also be said to be fit. However, other GoF criteria are not yet fit. Meanwhile, the results of the loading factor for each item in the first Model can be seen in Figure 1. The letter B in Figures 1, 2, and 3 is for the Belief aspect, R for Resilience, E for Effort, LC for Learning from Critics, LM for Learning from Mistakes, C for Challenge, F for Effort, F for Focus, and LO for Learning from Others.

Based on Figure 1, it can be seen that the items that have a loading factor above 0.4 are B2, B4, R1, R3, R4, R5, E1, E2, E3, E4, E5, LC1, LC2, LC3, LC4, LM1, LM3, LM4, C3, C4, C5, LO2, LO4, LO5, and LO6. Twenty-five items can be included in the calculation of the CR estimate, which is 0.908, which is higher than 0.7. This means that the first Model is highly consistent. However, because the GoF criteria have not been met, Model 1 is improved in the second Model by eliminating items with a loading factor below 0.4.

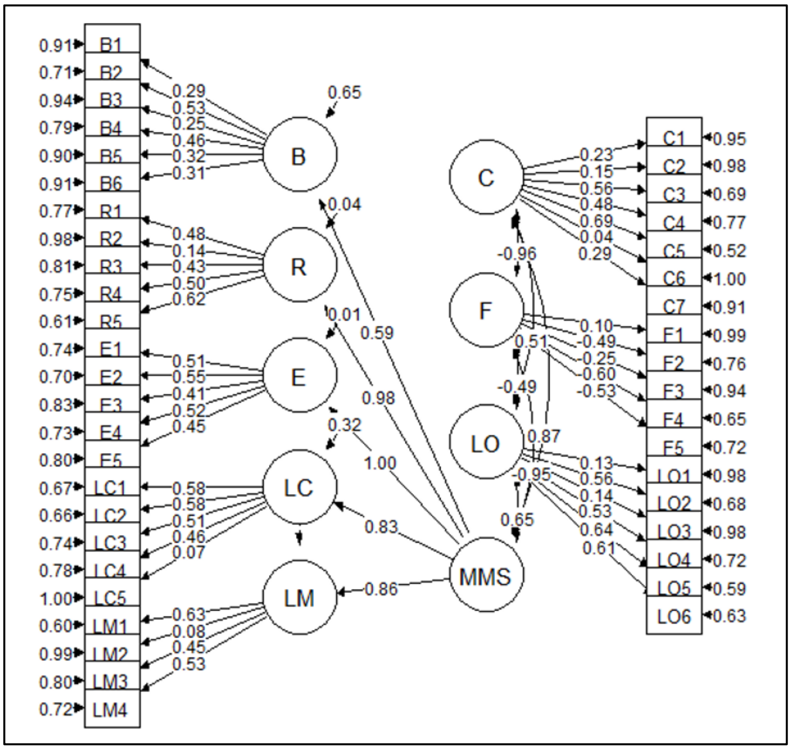

Figure 1. Model 1 MMS Likert with 43 Items

In Figure 1, it can also be seen that there is one aspect, namely Focus, all of which have loading factors below 0.4. The focus aspect determines the respondent's Focus based on their mathematical mindset tendencies. Someone with a growth math mindset will focus on the process, not the results. Conversely, someone with a fixed math mindset will focus on the results. Because this aspect has the same measurement objective as the effort aspect, these two aspects are not included in the MMS model analysis. The analysis of the second MMS model with 25 items showed an overall KMO value of 0.84. This value is above 0.5, meaning the number of samples is sufficient.

Meanwhile, the GoF criteria in this second Model have a higher RMSEA than the first, 0.083, and X2/df is 2.764. Both criteria indicate that the second Model is not yet fit, although it is close, likewise with other GoF criteria. However, there are no items with a loading factor below 0.4 in this second Model. Thus, based on the loading factor value of each item, the CR value of the second MMS model can be found to be 0.9097. This value shows that the second MMS model's consistency level is relatively high. Next, the third MMS model will be tried again by reducing the items below 0.49.

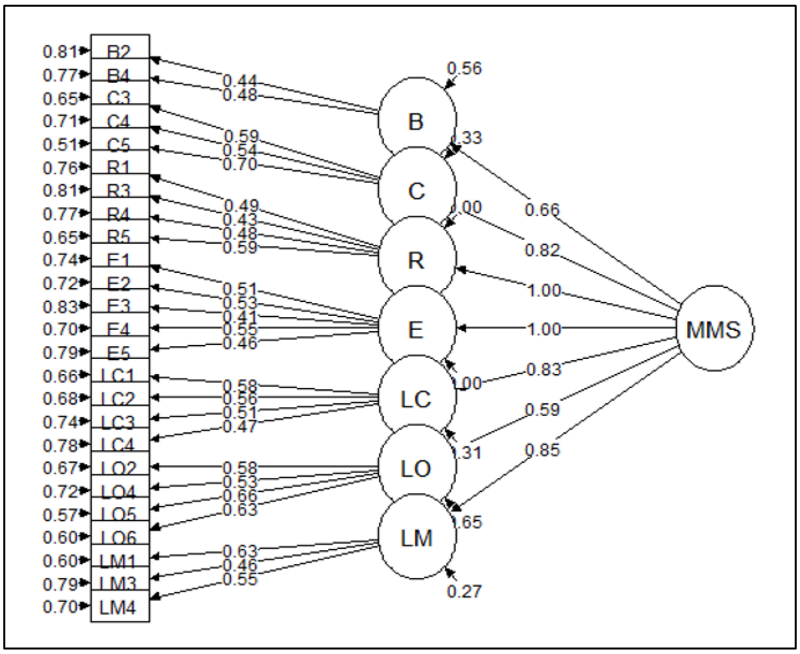

Figure 2. Model 2 MMS Likert with 25 Items

After eliminating items with a loading factor below 0.49 and trying to find a model that meets all the requirements, the third MMS model with 11 items was obtained. This Model has an overall KMO value of 0.9 and a CR of 0.8771. All items in the third MMS model have a loading factor above 0.4 with the GoF criteria in Table 2. Some GoF criteria are not yet fit, but almost all of the fit criteria have represented the three GoF types: absolute, incremental, and parsimonious.

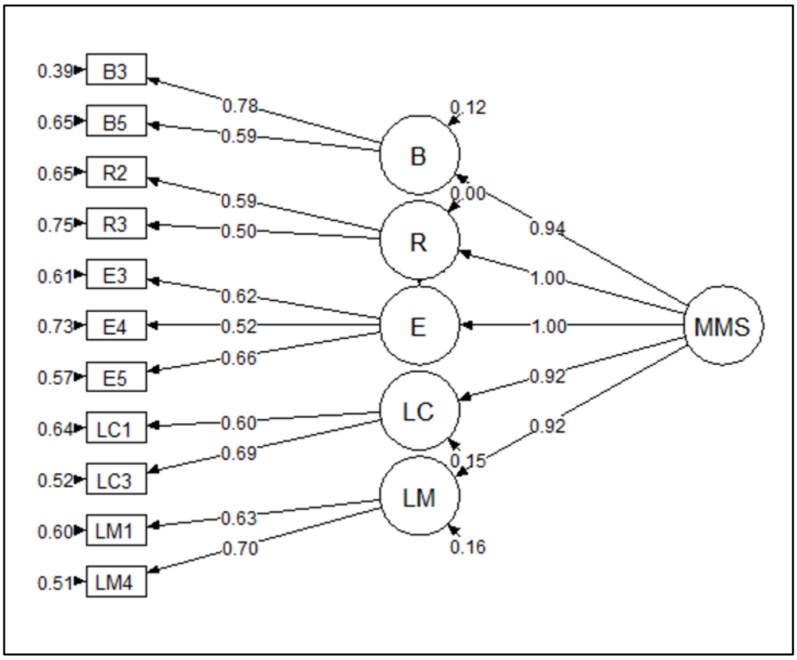

Figure 3. Model 3 MMS Likert with 11 Items

After the GoF is met, the analysis continues to item analysis with item response theory. The first thing that needs to be done is to prove the suitability of the Model to be analyzed (Arlinwibowo et al., 2024; Retnawati, 2014). Based on the characteristics contained in the MMS, the appropriate Model is the Graded Response Model (GRM) because the MMS uses a Likert scale to measure graded responses with several categories, from strongly disagree to agree strongly. Therefore, the appropriate Model is the GRM, and the following analysis process analyzes the MMS item parameters in the third Model. Estimating item parameters and checking model suitability are item calibrations (Retnawati, 2014).

In IRT analysis, there are three assumption tests: unidimensional, local independence, and invariance. The following will present the analysis results based on the three assumption tests. In the unidimensional test, MMS is expected to measure only one dimension, namely mathematical mindset. Based on the results of the scree plot in Figure 4 show that there is only 1 component, namely the first component, that has an eigenvalue above 1. In contrast, the other components have much lower Eigenvalues and are below line 1. This means that MMS only measures one dimension. In this case, it should be understood that the dimensions in question do not make up MMS, such as Belief, Effort, Resilience, and several other dimensions mentioned previously. However, the analysis of the unidimensional test aims to find out that the measuring instrument will only measure one ability (Hambleton et al., 1991). This differs from MMS, which only aims to discover a person's mindset tendencies. The statement items in MMS have the same purpose, so only one dimension or component is needed to measure mindset using MMS.

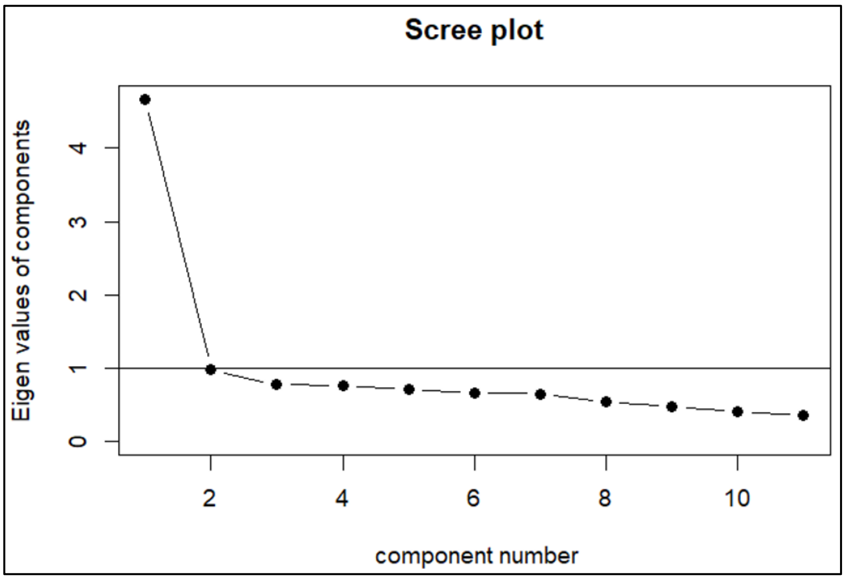

Figure 4. Scree plot of Model 3 MMS Likert

After the unidimensional test is fulfilled, the following assumption test is conducted, namely local independence. The local independence test aims to ensure that the respondent's answer to a statement item does not affect the respondent in answering other statement items (Retnawati, 2014). Based on the results of the local independence test, it is known that all items meet the test. This means that the 11 MMS items in the 3rd Model are not dependent on other items.

The third assumption test is parameter invariance. The purpose of the test is to determine whether the characteristics of the statement items do not depend on the distribution of respondent ability parameters and whether the parameters that characterize the respondents do not depend on the characteristics of the statement items (Retnawati, 2014). The test will be proven if the results of the item parameter estimates do not differ, even though they are tested on groups with different levels of ability (Retnawati, 2014). Figure 5 shows the results of the invariance test analysis of parameter a (discriminative power) of female and male Mathematics Education students. Based on the figure, each point is relatively close to the slope line 1. This indicates no variation in the estimated parameter of the discriminative power in the male and female student groups. Meanwhile, Figure 6 shows the results of the invariance test of parameter b (difficulty level) of the male and female student groups. Based on the figure, each point is relatively close to the slope line 1. This proves there is no variation in parameter b (difficulty level) in these two groups. After the three assumption tests are met, the analysis of the item response theory is continued with the analysis with characteristic graphs for each item, the value of the instrument information, and the standard error of the instrument.

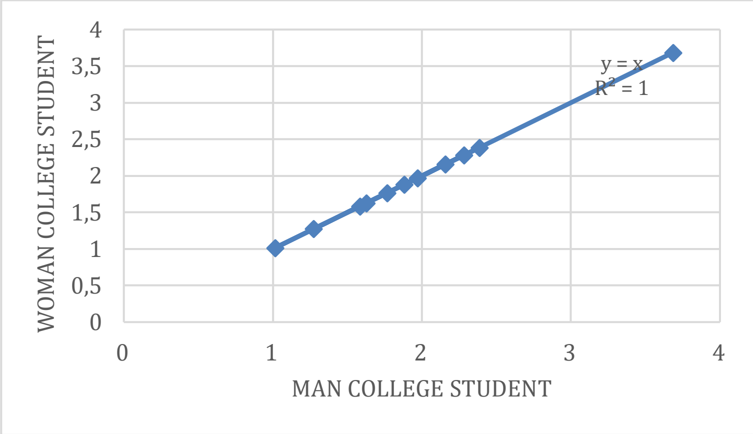

Figure 5. Scatter Plot of Invariance Test Parameter Based on Gender

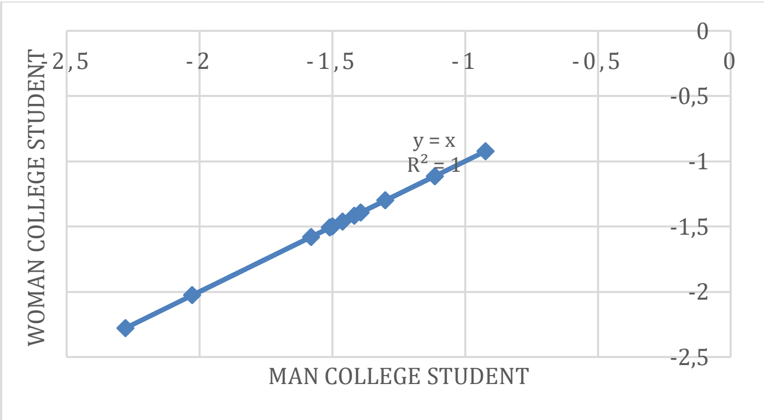

Figure 6. Scatter Plot of Invariance Test Parameter b Based on Gender

The item parameter data with the GRM model in Table 3 shows the values of a (discrimination power) and b (difficulty level). Since there are five MMS answer choices, the parameter b (difficulty level) has b1, b2, b3, and b4 as the intersection of each answer choice. An item can be considered good if the value of a (discrimination power) is between 0 and 2 (Hambleton et al., 1991). Based on Table 3, a characteristic graph can be made for each item. In parameter b4, several items do not have a value (Na), which indicates that respondents do not widely choose the answer choice with a score of 5. Respondents may have difficulty determining the answer choices for items B3, B5, R2, E5, and LM1.

Table 3. Items Parameter MMS for Discrimination Power(a) and Difficulty Level (b)

Items	a	b1	b2	b3	b4
B3	3.686328	-1.96854	-1.39347	-0.53769	Na
B5	1.275006	-2.88156	-2.26342	-0.93519	Na
R2	2.388053	-1.66158	-1.20064	0.091442	Na
R3	1.627754	-2.85199	-2.13321	-1.22637	-0.10992
E3	1.014512	-3.439	-3.01889	-2.0316	-0.62161
E4	1.584693	-1.97211	-1.80258	-1.10114	0.421107
E5	1.970682	-2.61953	-1.45372	-0.10562	Na
LC3	2.281342	-2.41278	-1.70762	-1.32	-0.22735
LC1	1.880576	-2.21673	-1.95581	-1.4781	-0.38379
LM1	2.156091	-2.45109	-1.55408	-0.37844	Na
LM4	1.765999	-2.3429	-2.07358	-1.4234	-0.15892

The GRM analysis model considers the difficulty and discriminant indices (Arlinwibowo et al., 2024). The difficulty level parameter (b) is a characteristic curve intersection. The intersection graph in Figure 7 makes it easy to compare with other items. Figure 7 represents the item probability curve for LC1, LM1, LM4, E3, E4, E5, LC3, R3, B3, B5, and R2, which were analyzed using GRM. Each plot shows the probability of choosing each category on a particular item based on the ability level (θ).

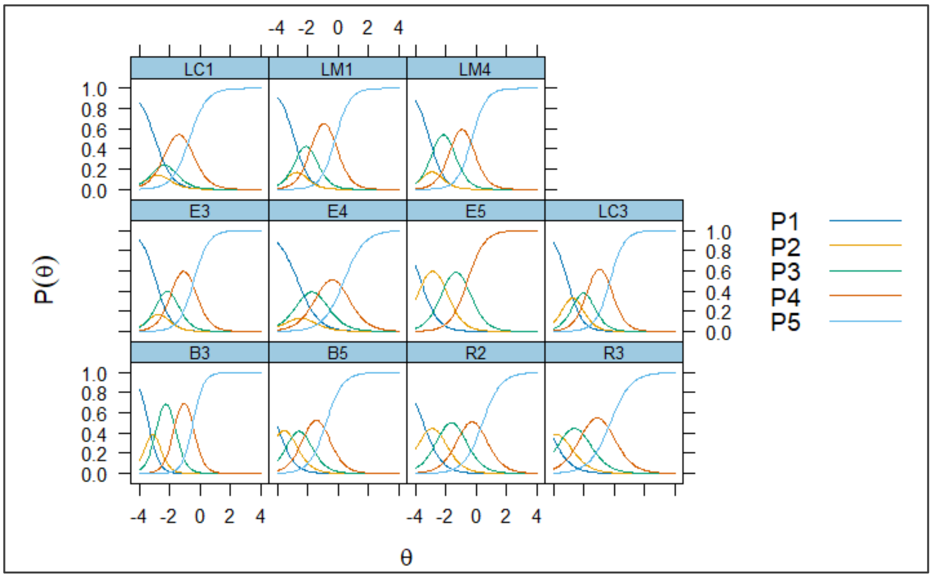

Figure 7. Items Probability Curve for Model 3 MMS Likert

The distribution curve in Figure 7 shows the most appropriate respondent ability for each item. P1, P2, P3, P4, and P5 show the probability of respondents' ability to choose each option. In items with a positive statement form, P1 is strongly disagree, P2 is disagree, P3 is undecided, P4 is agree, and P5 is strongly agree. Conversely, in items with a negative statement form, P1 is strongly agree, P2 is for agree, P3 is undecided, P4 is for disagree, and P5 is strongly disagree.

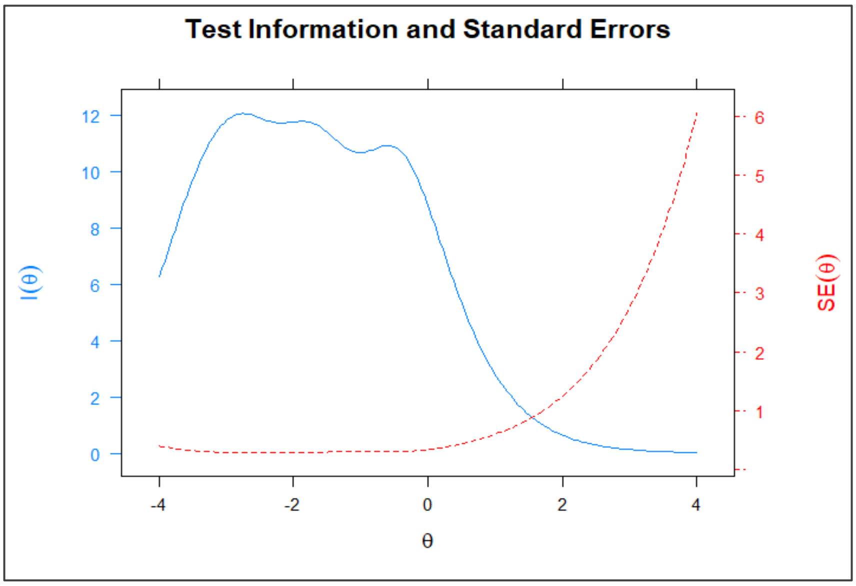

Figure 8. Information Function and Standard Error Curve for Model 3 MMS Likert

The function graph in Figure 8, symbolized by the blue line, shows how much information the test provides at various levels of ability or trait (*θ*). The Standard Error (SE) graph in Figure 8, indicated by the dashed red line, shows the uncertainty or level of measurement error at various levels of θ. The SE graph is inversely proportional to the information function graph, where the higher the information function graph, the vice versa. High information at a value of θ indicates that the test is more accurate or informative in measuring ability at that level.

DISCUSSION

The MMS statement items were initially made as many as 43 with eight aspects, i.e., Belief, Challenge, Effort, Focus, Resilience, Learning from Others, Learning from Critics, and Learning from Mistakes. Based on the analysis and selection of each stage, several aspects and items had to be eliminated. Several items were re-examined to determine whether they were suitable if included in other aspects that were not eliminated. Aspects that had to be removed had understandings and theories that were close to or even the same as other aspects that had not been eliminated. For example, the Challenge and Resilience aspects have the same goal of measuring how much someone is willing to face challenges and difficulties when studying mathematics. Likewise, effort and Focus determine a person's orientation when studying mathematics. Both Effort and Focus aim to measure the math mindset in terms of effort orientation and focus, namely more on the process or results. Someone with a growth math mindset will be oriented and focused on the process (Boaler, 2016; C.S. Dweck, 2014). On the other hand, someone with a fixed math mindset will be oriented towards results (Saefudin et al., 2023). Therefore, the Focus aspect must be eliminated, so the items in this aspect that have a loading factor above 0.4 or even 0.5 are moved to the Effort aspect.

The next aspect that was eliminated was the Learning from Others aspect. The statement items from the Learning from Others aspect that could still be maintained were then grouped into the Learning from Mistakes and Learning from Critics aspects. This was done by considering which items were more suitable to be included in Learning from Mistakes and which were more appropriate in the Learning from Critics aspect. Thus, the Challenge, Focus, and Learning from Others aspects were eliminated, and the items originally in the three aspects were combined into other aspects with more or less the same measurement objectives. Thus, the third MMS model has 11 items that meet the criteria for classical test theory analysis and IRT. The items can be seen in Table 4.

Table 4. Item Statements of MMS Likert Third Model with 11 Items

No.	Aspects	Statements
1	Belief (B3)	I realize I have no talent in mathematics, but I will do my best to master it.
2	Belief (B5)	By continuing to try, I will be able to do mathematics.
3	Resilience (R2)	I want to master mathematics even though it is difficult because it is challenging.
4	Resilience (R3)	If I had to study hard to master mathematics, I would not have been good at it.
5	Effort (E3)	Asking questions in class during math lessons shows that I am stupid.
6	Effort (E4)	I will continue to ask anyone until I understand mathematics.
7	Effort (E5)	Mastering mathematics is not easy, but I have the opportunity to learn it slowly.
8	Learning from Critics (LC3)	Criticism from others challenges me to prove that I can do mathematics, too.
9	Learning from Critics (LC1)	Criticism from others, namely that everyone has different talents, makes me realize that I do not need to force myself to study mathematics.
10	Learning from Mistakes (LM1)	Bad math grades are evidence that I will not be able to master mathematics.
11	Learning from Mistakes (LM4)	I can correct all my mistakes when answering math problems.

Meanwhile, the interpretation of each item depicted in the ICC curve (Figure 7) is as follows. The LC1 curve shows this item is more suitable for measuring participants with low to medium ability levels (around -3 to 1). Most participants with very low ability will likely choose the lowest category (P1), while participants with medium ability will move to a higher category. While the LM1 curve is similar to LC1, the middle category curves (P3 and P4) are more distributed at a higher θ (around -2 to 1), indicating that this item may assess participants with slightly higher ability than LC1. The LM4 curve is more evenly distributed along the ability range from -2 to 2. This item is suitable for assessing participants with medium ability because the higher categories or answer choices are only selected by respondents with higher ability levels. The E3 curve is similar to LC1 and LM1. The E3 curve shows respondents with low to medium ability tend to choose lower categories. Medium ability around -1 will choose categories P3 and P4. The E4 curve has a wider pattern than E3, with categories P4 and P5 being more dominant at mid-to-high ability (around 0 to 3). This indicates that this item is sensitive to participants with above-average ability.

The E5 curve shows a rightward shift in the curve for the mid-to-high category, indicating that participants with higher ability (around one and above) are more likely to choose the higher category (P4 or P5). The LC3 curve is similar to E5 but wider, indicating sensitivity across a broader range of abilities. Participants with abilities ranging from -2 to 2 can choose a wide range of categories on this item. The R3 curve also shows a fairly wide distribution, with the highest category only achieved at θ above 2. This item is suitable for participants with mid-to-high ability. The B3 curve is suitable for participants with lower ability since the highest category is already achieved at θ close to 0. This is similar to what was seen in the previous figure. The distribution of the B5 curve indicates that this item is more suitable for participants with medium ability, with the highest category only being achieved at θ close to 2, similar to item E5. The R2 curve has a similar distribution of curves to B5, where the highest category is achieved at θ above 2. This indicates that this item suits participants with medium to high ability.

Items such as LC1 and B3 can identify low ability, while E5 and R2 help measure participants with higher abilities. In general, items that are suitable for respondents with low ability are LC1 and B3. Meanwhile, items that are suitable for respondents with medium ability are items LM1, LM4, E3, and E4. Finally, items suitable for high-ability respondents are E5, LC3, R3, B5, and R2. Meanwhile, in Figure 8, it can be seen that the highest information is around the θ range between -2 to 1. This means that MMS is most informative for participants with an ability level in that range. The SE graph is low at the same ability interval as the high information, around -2 to 1. This means that measurements in that range are relatively more accurate and have lower errors. While outside this range, for example, at θ more than two or less than -2, the SE graph will increase. This will indicate that the test is less accurate for participants with abilities outside that range.

CONCLUSION

Reducing the number of aspects and items helps increase the GoF value because it can make the measuring instrument model simpler to meet the criteria of several GoF analyses, especially the Parsimonious criteria that emphasize model simplicity. The elimination of these three aspects is not without careful consideration. After conducting CFA several times, reducing MMS aspects can help improve the quality of the instrument, and several important GoF criteria can be adequately met. This certainly cannot be generalized for all instruments and all research conditions. Many factors can influence the results of this study. However, from the results of this study, it is known that non-test instruments can be less than 20 as long as they meet the analysis criteria at each stage. In addition, the number of respondents is more than 200, which greatly helps increase the reliability coefficient and loading factor value, even though the number of items in an instrument is few. In the future, other studies can be carried out that can improve the shortcomings of this study. In addition, studies that can prove the influence of a mathematical mindset on other variables are also needed. This research contributes to helping increase insight and knowledge for researchers in the fields of education and psychology, especially those interested in non-test measurement tools and IRT and classical analysis.

REFERENCES

Aiken, L. R. (1996). Rating Scales and Checklists: Evaluating Behavior, Personality, and Attitudes. John Wiley & Sons.

Aka, K. A. (2019). Integration Borg & Gall (1983) and Lee & Owen (2004) models as an alternative model of design-based research of interactive multimedia in elementary school. Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/1318/1/012022

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. Research in Social and Administrative Pharmacy, 15(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

Apiola, M., & Sutinen, E. (2020). Mindset and Study Performance: New Scales and Research Directions. ACM International Conference Proceeding Series. https://doi.org/10.1145/3428029.3428042

Arlinwibowo, J., Retnawati, H., & Hadi, S. (2024). Aplikasi Teori Respon Butir dengan R dan R Studio. Cahaya Harapan.

Arthur, W., Hagen, E., & George, F. (2021). The Lazy or Dishonest Respondent: Detection and Prevention. Annual Review of Organizational Psychology and Organizational Behavior, 8, 105–137. https://doi.org/10.1146/annurev-orgpsych-012420-055324

Ayebo, A., & Mrutu, A. (2019). An Exploration of Calculus Students' Beliefs about Mathematics. International Electronic Journal of Mathematics Education, 14(2), 385–392. https://doi.org/10.29333/iejme/5728

Azwar, S. (1994). Seleksi Aitem dalam Penyusunan Skala Psikologi. Buletin Psikologi, 2(2), 26–33. https://doi.org/ISSN: 0854-7106

Azwar, S. (1995). Human Attitude: Theory and Measurement. Pustaka Pelajar.

Azwar, S. (2024). Penyusunan Skala Psikologi. Pustaka Pelajar.

Boaler, J. (2016). A Mathematical Mindset. Jossey-Bass. https://doi.org/10.1088/2058-7058/32/12/32

Chen, S., Ding, Y., & Liu, X. (2021). Development of the Growth Mindset Scale: Evidence of Structural Validity, Measurement Model, Direct and Indirect Effects in Chinese Samples. Current Psychology. https://doi.org/10.1007/s12144-021-01532-x

Chestnut, E. K., Lei, R. F., Leslie, S. J., & Cimpian, A. (2018). The myth that only brilliant people are good at math and its implications for diversity. Education Sciences, 8(2). https://doi.org/10.3390/educsci8020065

Cutler, C. S. (2020). Preservice Teachers' Mathematical Mindsets During Pandemic-Induced Pivot to Online Learning. Frontiers in Education, 5(November), 1–6. https://doi.org/10.3389/feduc.2020.595264

Daly, I., Bourgaize, J., & Vernitski, A. (2019). Mathematical Mindsets Increase Student Motivation: Evidence from the EEG. Trends in Neuroscience and Education, 15, 18–28. https://doi.org/10.1016/j.tine.2019.02.005

Dweck, C S. (2014). Mindsets and math/science achievement. rpforschools.net. http://www.rpforschools.net/articles/Mindsets/Dweck 2008 Mindsets and Maths Science Achievement.pdf

Dweck, Carol S. (2022a). Mindset (Updated). Ballantine Books.

Dweck, Carol S. (2022b). Self-theories: Their role in motivation, personality, and development. Routledge - Taylor & Francis Group, LLC.

Dweck, Carol S. (2008). Mindsets and Math / Science Achievement. The Opportunity Equation: Transforming Mathematics and Science Education for Citizenship and the Global Economy, 1–17. www.opportunityequation.org

Furr, R. M. (2014). Scale Construction and Psychometrics for Social and Personality Psychology. Scale Construction and Psychometrics for Social and Personality Psychology, 3–15. https://doi.org/10.4135/9781446287866

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). Educational Research: An Introduction. In Qualitative Voices in Educational Research. Pearson Education. https://doi.org/10.4324/9781003008064-1

Hair, Joe F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM : Indeed, a Silver Bullet. Journal of Marketing Theory and Practice, 19(2), 139–151. https://doi.org/10.2753/MTP1069-6679190202

Hair, Joseph F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis (Seventh). Pearson Prentice Hall.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory (D. S. Foster (ed.)). Sage Publication.

Haryono, S. (2016). SEM Methods for Management Research with AMOS, LISREL, PLS. Intermedia Personalia Utama.

Heyder, A., Weidinger, A. F., Cimpian, A., & Steinmayr, R. (2020). Teachers' Belief that Math Requires Innate Ability Predicts Lower Intrinsic Motivation Among Low-Achieving Students. Learning and Instruction. https://doi.org/10.1016/j.learninstruc.2019.101220

Hutajulu, J. M. J., Djunaidi, A., & Triwahyuni, A. (2021). Properti Psikometri Beck Hopelessness Scale pada Populasi Non-Klinis Indonesia. Intuisi: Jurnal Psikologi Ilmiah, 13(1), 24–37. https://doi.org/10.15294/intuisi.v13i1.28037

Im, S., & Park, H. J. (2023). A Mathematical Mindset Scale using the positive norms. Psychology in the Schools, 60(8), 2901–2918. https://doi.org/10.1002/pits.22904

Limeri, L. B., Carter, N. T., Choe, J., Harper, H. G., Martin, H. R., Benton, A., & Dolan, E. L. (2020). Growing a growth mindset: characterizing how and why undergraduate students' mindsets change. International Journal of STEM Education, 7(1). https://doi.org/10.1186/s40594-020-00227-2

Lynn, M. R. (1986). Determination and quantification of content validity. In Nursing Research (Vol. 35, Issue 6, pp. 382–386). http://ijoh.tums.ac.ir/index.php/ijoh/article/view/26

Mardapi, D., Kumaidi, & Kartowagiran, B. (2011). Pengembangan Instrumen Pengukur Hasil Belajar Nirbias dan Terskala Baku. Jurnal Penelitian Dan Evaluasi Pendidikan, 15(2), 326–341. https://doi.org/http://dx.doi.org/10.21831/pep.v15i2.1100

Meierdirk, C., & Fleischer, S. (2022). Exploring the mindset and resilience of student teachers. Teacher Development, 26(2), 263–278. https://doi.org/10.1080/13664530.2022.2048687

Midkiff, B., Langer, M., Demetriou, C., & Panter, A. T. (2018). An IRT analysis of the growth mindset scale. In Springer Proceedings in Mathematics and Statistics (Vol. 233). Springer International Publishing. https://doi.org/10.1007/978-3-319-77249-3_14

Mills, I. M., & Mills, B. S. (2018). Insufficient evidence: Mindset intervention in developmental college math. Social Psychology of Education. https://doi.org/10.1007/s11218-018-9453-y

OECD. (2020). Growth mindset. In PISA 2018 Results (Volume III) (pp. 199–209). OECD. https://doi.org/10.1787/bd69f805-en

Rahmah, N. (2018). The Nature of Mathematics Education. Al-Khwarizmi, 1(2), 1–10. https://doi.org/10.24256/jpmipa.v1i2.88

Ramirez, G., Hooper, S. Y., Kersting, N. B., Ferguson, R., & Yeager, D. (2018). Teacher Math Anxiety Relates to Adolescent Students' Math Achievement. AERA Open, 4(1), 1–13. https://doi.org/10.1177/2332858418756052

Rammstedt, B., Gruning, D. J., & Lechner, C. M. (2022). Measuring Growth Mindset: Validation of a Three-Item and a Single-Item Scale in Adolescents and Adults. European Journal of Psychological Assessment, 1, 1–12. https://doi.org/10.1027/1015-5759/a000735

Rammstedt, B., Grüning, D. J., & Lechner, C. M. (2021). Measuring Growth Mindset: A Validation of a Three-item and Single-item Scale in Youth and Adults. Center for Open Science. https://doi.org/10.31234/osf.io/rs43g

Retnawati, H. (2014). Teori respons butir dan penerapannya (Item Response Theory and its aplication). Nuha Medika.

Retnawati, H. (2016). Analisis Kuantitatif Instrumen Penelitian (Quantitative Analysis of Research Instruments). Parama Publishing.

Rizdanti, S., & Akbar, S. (2022). Hubungan Religiusitas dengan Tingkat Stres Dalam Menyusun Skripsi di Fakultas Kedokteran Universitas Islam Sumatera Utara. Jurnal Kedokteran STM (Sains Dan Teknologi Medik), 5(2), 94–100. https://doi.org/10.30743/stm.v5i2.318

Saefudin, A. A., Wijaya, A., Dwiningrum, S. I. A., & Yoga, D. (2023). The characteristics of the mathematical mindset of junior high school students. Eurasia Journal of Mathematics, Science and Technology Education, 19(1). https://doi.org/10.29333/ejmste/12770

Santoso, S. (2014). Basic concepts and applications of SEM with AMOS 22. Jakarta: Elex Media Komputindo.

Shen, C., Miele, D. B., & Vasilyeva, M. (2016). The Relation between College Students' Academic Mindsets and Their Persistence during Math Problem Solving. Psychology in Russia: State of the Art, 9(3), 38–56. https://doi.org/10.11621/pir.2016.0303

Simanjuntak, M., & Hamimi, U. K. (2019). Penanganan Komplain dan Komunikasi Word-of-Mouth. Jurnal Ilmu Keluarga Dan Konsumen, 12(1), 75–86. https://doi.org/DOI: http://dx.doi.org/10.24156/jikk.2019.12.1.75

Stohlmann, M. (2022). Growth mindset in K-8 STEM education: A literature review since 2007. Journal of Pedagogical Research, 6(2), 149–163. https://doi.org/10.33902/JPR.202213029

Suh, J. M., Graham, S., Ferranone, T., Kopeinig, G., & Bertholet, B. (2011). Developing persistent and flexible problem solvers with a growth mindset. In Motivation and Disposition: Pathways to Learning Mathematics (Issue 2006, pp. 169–184). NCTM 2011 Yearbook.