Design and Validation of Mathematical Literacy Instruments for Assessment for Learning in Indonesia

Suciati[*]
Yogyakarta State University/
Universitas Borneo Tarakan,
INDONESIA

Sudji Munadi
Yogyakarta State University,
INDONESIA

Sugiman
Yogyakarta State University,
INDONESIA

Wiwin Dwi Ratna Febriyanti
Universitas Borneo Tarakan,
INDONESIA

Abstract: This study aims to design mathematical literacy instruments that have evidence of content and construct validity and are reliable for use as an assessment for learning. The research involved eight experts as instrument validators and 273 eighth-grade students of junior high school in Yogyakarta Province. The results showed that the ten mathematical literacy items developed had the V Aiken coefficient index calculated from 0.781 to 0.906 (> 0.75). The results of adequacy testing of samples with KMO and Bartlett show Chi-Square in the Bartlett test of 608,608, the p-value <0.05 and KMO value of 0.781 (> 0.5). The results of testing of the measurement model with Confirmatory Factor Analysis (CFA) produce a Root Mean Square Error of Approach (RMSEA) value of 0.049 (≤ 0.08), chi-s Square of 33.92 (<2df), the p-value of 0.05004 (≥ 0.05). Nine out of the ten items developed had t-value> 1.96, Standardized Loading Factor (SLF) was greater than the critical limit (> 0.3), and Construct Reliability (CR) of 0.78 (> 0.7). It can be concluded that the developed mathematical literacy instrument can measure what must be measured and nine items significantly reflect the construct or latent variable, as well as the level of consistency of a good score.

Keywords: *Instruments, mathematics literacy, content validity, construct validity, construct reliability.*

Introduction

Science and technology are developing rapidly in the 21st century. Every person is required to have the ability to think more creatively and be able to accept rapid technological developments in this century (Maskur et al., 2020). One of the abilities needed to compete in the 21st century is Mathematical Literacy.

Mathematical Literacy is the ability to formulate, use, and interpret mathematics in various contexts, including mathematical reasoning, using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena to assist individuals in making constructive and reflective decisions (Owens, 1988; Organisation for Economic Co-operation and Development [OECD], 2016; Ojose, 2011; Stacey & Turner, 2015; Steen et al., 2007). This understanding implies mathematical literacy not only in the mastery of the material but also in the use of reasoning, concepts, facts, and mathematical tools in solving everyday problems. Someone who has sensitivity in sorting out mathematical concepts that are relevant to the problems they face will have good mathematical literacy too. This ability requires a person to understand, analyze, interpret, evaluate, and synthesize information obtained from the problem at hand and then be modeled into a mathematical model and determined the solution to solve the problem using mathematical concepts effectively.

Mathematical literacy according to the OECD (2013) in PISA 2015 draft, mathematics framework consists of three aspects, namely process, content, and context. Process aspects relate to what individuals do to link the context of a problem with mathematics and its problem solving, and the abilities that underlie these processes. The content aspect is the target that will be measured in the assessment. The context aspect is where the problem is placed. Again, the context aspect is where the assessment located. Problems in mathematical literacy are problems in interesting real-world contexts and require the use of real-life data in modeling problems (Kula et al., 2018)

The development of mathematical literacy is a demand that must be done by getting used to it in the learning process (del Prado Hill et al., 2016). Assessment is one of the important aspects of the education process especially learning

---

[*] **Corresponding author:**
Suciati, Yogyakarta State University, Indonesia. ✉ cauchy_my@yahoo.com

(Friyatmi et al., 2020; Kuger et al., 2016; Matters, 2009; Bishop et al., 1996). Thus, the development of mathematical literacy in the learning process also requires assessment. An assessment system that can train, familiarize, and develop students' mathematical literacy, one of which is an assessment system that is integrated with the learning process is Assessment for Learning.

Assessment for Learning is done during the learning process takes place and is usually used as a basis for improving teaching and learning process by providing feedback on the learning process of students, monitoring progress, and determining the progress of learning (Kemendikbud, 2017; Black & Wiliam, 2009; Bennett & Gitomer, 2009; Black & Lee, 2003; Burton et al., 2018; Mardapi, 2017). Thus, it is necessary to develop test instruments in the form of questions that are integrated between aspects and indicators of mathematics literacy with indicators of achievement of competence in learning. Mathematical literacy instruments are developed by subject competencies based on the applicable curriculum in Indonesia so that information from the assessment results can be used to improve the learning process.

A good instrument that can be used as a basis for decision making is an instrument that meets several criteria, including validity and reliability (Anderson, 2010). Validity is a fundamental criterion that must be considered in developing tests. American Educational Research Association (AERA) (2014) states that validity refers to the extent to which evidence and theory support the interpretation of test scores for the use of the proposed test. Based on the source of the evidence used there are three types of validity, namely 1) content validity, 2) criteria-related validity, and 3) construct validity (Retnawati, 2015; Allen & Yen, 1979; Azwar, 2007; Crocker & Algina, 1986; Nunnally & Bernstein, 1978). The evidence is accumulated to support the interpretation of test scores and to prove validity. Based on the evidence, a measurement result can be said to be valid or not.

Content validity is validity estimated through testing the feasibility or relevance of the test content through rational analysis by a competent panel of expert judgment (Azwar, 2007). The contents of measurements are adjusted to the measurement specifications in the form of measuring instruments and the type of understanding measured (Naga, 2012). The validity of the contents is more focused on the material in the measuring instrument.

The second type of validity is criterion validity. This validity is used when the test score can be associated with other criteria that have been standardized. Allen and Yen (1979) mention that this validity has a correlation coefficient between test scores and criterion scores. Azwar (1996) added validation procedures based on these criteria will produce one of two types of validity, predictive validity or concurrent validity. Predictive validity and concurrent validity are distinguished by the waiting time for the acquisition of the two data scores.

In addition to validity, reliability is also one indicator of an instrument said to be good or not. Stanley (1971) defines reliability as the consistency of measurement to one another. Consistent measurement results shown using the same measuring instrument for different people or times will get the same results.

The development of mathematical literacy instruments that have evidence of validity and good reliability coefficients will produce a set of instruments that are feasible to use to measure students' mathematical literacy achievements. So the information obtained can be used to improve the learning process. The mathematical literacy instrument developed has differences with the mathematical literacy instrument developed by PISA. The question indicators developed are integration between the PISA mathematics literacy indicators and the learning objectives of mathematics in the Indonesian national curriculum. The current curriculum in Indonesia is the 2013 curriculum. The curriculum is a new curriculum that is centered on students (Ichsa et al., 2019). The learning objectives used are learning objectives in terms of the systems of the linear equations in two variables. Therefore the mathematical literacy instrument developed is suitable for Assessment for Learning in eighth-grade students of junior high school in Indonesia.

Methodology

This study aims to design mathematical literacy instruments that can be used for Assessment for Learning and have evidence of content and construct validity and reliability. The research begins by developing an instrument construct based on theoretical studies. The study was conducted in several stages, namely 1) designing mathematical literacy instruments, 2) testing instruments, 3) analyzing the results of trials.

Sample and Data Collection

Quantitative data obtained from the results of this study consisted of two types of data. The first set of data is the results of expert assessments (validators). Validation is done by technique Delphi and involved 8 experts consisting of 4 mathematics education experts and 4 education assessment experts. The Delphi technique steps that are used refer to Hsu and Sandford (2007). The assessment is done by checking the compatibility of the items with the concepts, writing techniques, and language used. Validators provide comprehensive assessment decisions using five categories, namely irrelevant (1), less relevant (2), quite relevant (3), relevant (4), and very relevant (5).

The second set of data is collected from students. Data collection sourced from students is divided into two, namely readability trials and limited trials. In the legibility test, there were 31 students involved. Students are asked to read

and understand everything about the test that must be completed in the mathematics literacy test instrument given. Students are asked to provide input or suggestions in parts where students do not understand or other input related to the length of the question sentence, typeface, font size, image/table display, and clarity of the stimulus provided.

The second trial is a limited trial involving 273 students. Quantitative data collected are in the form of response patterns or answers of test participants after working on 10 items of mathematical literacy questions. Scoring is given by considering the steps done by the students to complete the questions based on criteria combination. 2 scores (full credit) is for all correct steps, 1 score is for partial correct steps, and 0 score is for incorrect steps. The entire instrument trial involved eighth-grade students of Junior high school in Yogyakarta Province.

The sample size in the trial is determined by considering the objectives of the trial and the analysis used in this case is Confirmatory Factor Analysis (CFA). According to Gorsuch (1983), the minimum sample size of respondents in the CFA is 200 respondents. Therefore, as many as 273 respondent subjects (students) have fulfilled the sample size requirements. The sampling technique uses purposive cluster sampling based on regencies in Yogyakarta Province and student abilities (high, medium, low).

Analyzing of Data

Data analysis is carried out in stages. The first stage is the analysis to obtain evidence of content validity based on validator ratings. The analysis was performed by calculating the magnitude of the content validity coefficient (V) index using the V Aiken formula as follows (Aiken, 1985):

V = ∑ s / [n(c-1)]   ……………. (1)

S = r - lo
lo = The lowest validity rating (in this case = 1)
c = The highest validity rating (in this case = 5)
r = score given by rater
n = the number of raters

The second stage of analysis, namely analysis to obtain evidence of construct validity and reliability. The analysis was carried out based on the test participants' response patterns. The analysis used was CFA with the help of Lisrel 8.50 software. The analysis is carried out in stages, 1) testing the adequacy of the sample using the KMO and Bartlett test results, 2) testing the measurement model, 3) analyzing the construct validity based on t-value and Standardized Loading Factor (SLF), 4) estimate the construct reliability coefficient. The construct reliability coefficient is estimated by the equation (Retnawati, 2015):

$$CR = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 + \left(\sum_{i=1}^{n} \delta_i\right)}$$ $i = 1,2,3,…n$ ……………. (2)

CR   = Construct Reliability
λ = Standardize Loading factor
δ = Error of each indicator

The results of the construct reliability coefficient estimation are further classified based on the level of reliability. The reliability level classification according to Guilford (Istiyono, 2018) is as follows:

Table 1. Level of Reliability

No	Reliability Coefficient	Level of Reliability
1	0,80 - 1,00	Very high
2	0,60 - 0,80	High
3	0,40 - 0,60	Enough
4	0,20 – 0,40	Low
5	0,00 – 0,20	Very low

Design of Mathematical Literacy Instruments

Mathematical literacy assessment instruments consist of instrument constructs, instrument lattices, math literacy questions, and scoring guidelines. The development of mathematical literacy instruments is carried out by the steps of developing written test instruments.

1. Setting test objectives

Determining the purpose of the test is very important because the form and how the test is developed depends on the purpose of the test. The developed mathematical literacy assessment instrument is a test that is used as an Assessment for Learning to measure the achievement of mathematics literacy for eighth-grade students of SMP on the content change and relationships material systems of linear equations in two variables.

2. Establishing competencies and the material being tested

Because what is being developed is a formative test that measures the achievement of mathematical literacy on the material systems of linear equations in two variables, the scope of the material includes the basic competencies of the systems of linear equations in two variables. For this purpose, the core competencies and basic competencies of the lessons refer to the Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 37 of 2018 (Rules of Change). Based on the Basic Competence (KD), the construct of the mathematical literacy assessment instrument on the material systems of linear equations in two variables.

The construct of the mathematical literacy instrument construct begins with defining mathematical literacy conceptually and operationally, to obtain an indicator of mathematical literacy. These indicators are integrated with KD and continuum learning to produce indicators of achievement of mathematical literacy competencies in the material systems of linear equations in two variables.

3. Arranging the grain distribution matrix

After obtaining an indicator of achievement in mathematics literacy competence the next step is to determine the test distribution matrix. The matrix contains components of aspects of the mathematical literacy process, components of the context of mathematical literacy (personal, work, social, and scientific), and the number of instrument items.

Table 2. Matrix Distribution of Mathematics Literacy Tests on Material Systems of Linear Equations in Two Variables

Components Process	Components Process			Number of Items
	Formulate The Situation Mathematically	Use Concepts, Procedures, Facts, And Reasoning	Interpreting And Evaluating Mathematical Results	
Personal	1	1	1	3
Occupational	1	0	1	2
Societal	0	1	1	2
Scientific	1	1	1	3
Total	3	3	4	10

4. Arranging instrument blueprints

The instrument blueprint is a guideline for assembling items into tests. The instrument lattice contains the item specifications in the form of BC, the content of mathematical literacy, aspects of mathematical literacy, the context of mathematical literacy, indicators of competency achievement, problem indicators, the form of questions, and the number of questions. Indicators for items are arranged using verbs that correspond to the depth of the questions that measure aspects of formulating, using and interpreting concerning indicators of achievement of competence and levels of achievement in mathematical literacy.

Mathematical literacy items use the context of real life, and the problems given are non-routine. Therefore, it is necessary to load stimulus questions for students to think about and be a source of information for solving problems. The stimulus questions are adjusted to the context or situation where the problem is placed, namely the work, social, scientific, and personal context. The stimulus given is presented in the form of discourse, pictures, tables, graphs, and so on. Blueprints developed in the initial stages are presented in Table 3.

Table 3. Test Blueprints of Mathematical Literacy on Material Systems of Linear Equations in Two Variables

Aspect	Indicator	Context	Level	Item Number
Formulate the situation mathematically	Understanding the symbolic language needed to present it mathematically	Personal	2	1
	Presenting the situation mathematically using appropriate variables and models	Occupational	2	3
	Translating problems into mathematical language	Scientific	3	4
Use concepts, procedures, facts, and reasoning	Implementing strategies in finding mathematical solutions	Societal	3	7
	Identifying the completion of linear equations in two-variable based on a given contextual problem	Personal	4	8
	Solving mathematical models of problems related to linear equations of two variables	Scientific	4	5
Interpreting and evaluating mathematical results	Reinterpreting mathematical results in the real context	Personal	5	2
	Evaluating the suitability of mathematical solutions to the context of the problem	Societal	5	9
	Reinterpreting mathematical results in the real context	Scientific	5	6
	Explaining the reasons why the results or conclusions are in line with the context of the problem	Occupational	6	10

5. Writing questions

The item writing is done by describing the problem indicators into questions whose characteristics are by the details in the grid that has been made. The questions are written are problems that will be given to students during the learning process. The problem meets the criteria of a problem in mathematics, which is routine, involves several steps of completion and focuses on reasoning.

The questions are written also pay attention to the rules that apply. Among other things, it needs to be written clearly and firmly (not convoluted). The language used is adapted to the development of students so that it is easy to understand. The type and size of letters also need to be considered so that students feel comfortable. The stimulus is given, whether in the form of discourse, pictures, tables, or graphs must be displayed clearly and functioning.

Mathematical literacy instruments that have been designed in the early stages are then explored with Focus Group Discussion (FGD). FGD was conducted to obtain input, suggestions, and responses from participants that were useful for improving the design of the assessment model. The revised assessment instrument design based on input, suggestions, and responses to the results of the FGD subsequently became products that were ready to be provided by students for readability testing.

The legibility test involved 31 junior high school students in class VIII. The legibility test results showed that 1) some items were not understood by students because the formulation of the problem sentences was not stated clearly and firmly, 2) the size of the picture in item number 3 was too small so that some numbers could not be read clearly, 3) items questions number 1, 2, and 3 do not limit the answers students must give, 4) item number 2 is interesting because it is like an android game. Based on the results of the readability test, the researcher revised the appearance of the image, the size of the writing, simplified the formulation of the sentence to make it easier for students to understand and added the answer limits to questions number 1, 2, and 3. The revised instrument design was then ready to be validated by experts at a later stage.

Findings and Discussion

Contents Validity of Mathematical Literacy Instruments

Mathematical literacy instruments that have been developed in the initial stages were further validated to determine the appropriateness of the instrument in measuring the achievement of students' mathematical literacy. Validation was done by technique Delphi and involved 8 experts consisting of 4 mathematics education experts and 4 education assessment experts. The experts examined the items and provided an assessment of the mathematical literacy instruments that researchers developed based on the assessment sheets provided. The assessment sheet contains aspects of the assessment specified in the rules of preparation of the test.

The results of the experts' assessment were qualitative and quantitative. Qualitative assessment results were given by experts in the form of suggestions and revisions to improvements to the instrument construct formulation of indicators, subject matter, and the form of scoring guidelines. Quantitative assessment results were given by experts after the researchers made revisions based on advice and input provided by experts. The assessment was done by

checking the compatibility of the items with the concepts, writing techniques, and language used. The experts gave a comprehensive assessment decision using five categories, irrelevant, less relevant, quite relevant, relevant, and very relevant. Based on the results of the experts' validation, all the experts stated that the instrument was suitable for use after the revision.

The feasibility of the mathematical literacy assessment instrument was also supported by the content validity coefficient (V_c) for the item was computed using the Aiken formula (1). The results of the V_c calculation were then compared with the coefficient V in the table (V_{Table}) in the Right-Tail Probabilities (p) table for selected Values of the Validity Coefficient (V) (Aiken, 1985). Items that have been developed were declared valid, if $V_c \geq V_{Table}$. V_{Table} can be seen by adjusting the number of rater and the number of categories used in the assessment sheet. Because of the many rater or experts who assessed 8 people and the number of categories was five, then obtained V_{Table} of 0.75. The recapitulation of the Aiken index V calculation results on the mathematical literacy instrument with 10 items is presented in Table 4 below.

Table 4. Content Validity Coefficient (V) for the Item of Mathematical Literacy Instruments

Item Number	ΣS	n	c-1	V_{Table}	V_c	Decision
1	28	8	4	0,75	0,875	Valid
2	28	8	4	0,75	0,875	Valid
3	28	8	4	0,75	0,875	Valid
4	25	8	4	0,75	0,781	Valid
5	29	8	4	0,75	0,906	Valid
6	27	8	4	0,75	0,844	Valid
7	26	8	4	0,75	0,813	Valid
8	28	8	4	0,75	0,875	Valid
9	29	8	4	0,75	0,906	Valid
10	26	8	4	0,75	0,813	Valid

Based on Table 4, it appears that the V coefficient on each item is in the range of 0.781 to 0.906. This shows that all mathematics literacy items have good evidence of content validity in terms of conformity with the indicator because the value of the V coefficient exceeds the Aiken index V table value of 0.75. The results of the content validity provide evidence that the developed mathematical literacy instrument can measure what should be measured so that it is feasible to use it for the next stage which is tested to obtain evidence of the quality of the instrument.

Constructing Validity of Mathematical Literacy Instruments

Mathematical literacy assessment instruments consisting of 10 items that have obtained evidence of eligibility based on content validity were then tested. The instrument trials involved 4 schools in Yogyakarta Province with 273 students taking the test. All test participants were asked to work on the whole math literacy items in 70 minutes. The test participant's responses or answers after working on 10 items of math literacy questions were scored using a scoring guide that had been developed. Scoring is given by considering the steps done by the students to complete the questions based on criteria combination. 2 scores (full credit) is for all correct steps, 1 score is for partial correct steps, and 0 score is for incorrect steps.

The scoring results of all test participants' answers were analyzed to prove that the construct developed could be empirically confirmed. The analysis used is Confirmatory Factor Analysis (CFA) with the help of Lisrel 8.50 software. The analysis begins with testing the adequacy of the sample. Analysis of sample adequacy using KMO and Bartlett test results. The results of the analysis are presented in Table 5.

Table 5. KMO and Bartlett Test Results

Kaiseer-Meyer-Olkin Measure of Sampling Adequacy		0,781
Bartlett's test of sphericity	Approx. Chi-Square	608,608
	df	45
	Significance	0,000

The results of the analysis of the adequacy of the sample show the Chi-Square value in the Bartlett test of 608.608 with a degree of freedom of 45, a value of p <0.05 and a KMO value of 0.781. A group of data is said to meet the assumption of sample sufficiency if the KMO value is> 0.5 (Hair, as cited in Hidayati et al., 2018). This means that a sample size of 273 used in the instrument trials was sufficient. Therefore, the analysis can be continued by testing the measurement model.

Testing the measurement model using factor analysis

The factor analysis model used is confirmatory factor analysis with two-level latent variables or second order. The measurement model based on the results of the analysis is presented in Figure 1.
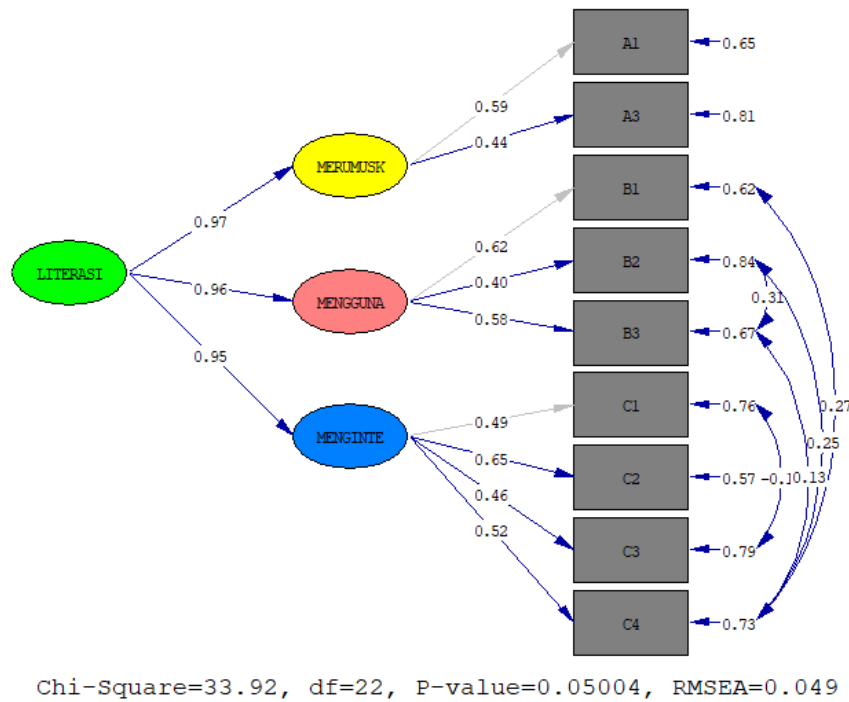

Chi-Square=33.92, df=22, P-value=0.05004, RMSEA=0.049

Figure 1. Path of the Second Order Diagram of CFA Mathematics Literacy

Based on the results of the analysis in the first iteration, the coefficient was obtained loading factor for item number 2 (A2) on the indicator formulating 0.14. Because the coefficient loading factor was smaller than the criterion limit (> 0.3), then the item was decided to be removed from the measurement model. The results of tests conducted on the measurement model with 3 indicators and the remaining 9 items produce a value Root Mean Square Error of Approximation (RMSEA) of 0.049 (≤ 0.08), chi-s Square of 33.92 (<2df), the p-value 0.05004 (≥ 0.05). Based on the criteria of goodness of fit (GoF) index (Brown, 2015; Ghozali & Fuad, 2015; Kline, 2014), all three of them have met the models fit criteria. The values of the other match indices are presented in Table 6 below.

Table 6. Testing Model Fit Based on Goodness of Fit Index

Goodness of Fit Index	Criteria	Estimated Results	Decision
Chi-square	< 2df	33,92 (df= 22)	*fit*
Significance (p-value)	≥ 0.05	0,05004	*fit*
RMSEA	≤ 0.08	0,049	*fit*
Goodness of Fit Index (GFI)	≥ 0.90	0,97	*fit*
Adjusted Goodness of Fit Index (AGFI)	≥ 0.90	0,93	*fit*
Normed Fit Index (NFI)	≥ 0.90	0,93	*fit*
Comparative Fit Index (CFI)	≥ 0.95	0,97	*fit*
Incremental Fit Index (IFI)	≥ 0.95	0,97	*fit*

Based on Table 6, 8 model fit criteria were used, indicating all model match criteria were met. This means that the measurement model developed was supported by data. Thus, it can be concluded that the measurement model of mathematical literacy assessment instruments as a whole shows a good match.

After the measurement model developed empirically was stated to have a good match based on the data, then the next step was to prove the construct validity. The proof was done using a size Standardized Loading Factor (SLF). The Loading Factor is a correlation coefficient that shows the closeness of the relationship between latent variables with manifest variables. According to Kline (2014), the loading factor of 0.3 is a reasonable criterion for showing prominent charge indications for samples of at least 100.

Analysis results from second Order CFA in terms of t-value and the SLF values presented in Table 7 show that the 9 items analyzed have values t-value > 1.96 and have an SLF value greater than the critical limit (> 0.3). This means that the nine items significantly reflect construct or latent variables and have good proof of validity. The analysis results are presented in Table 7 below.

Table 7. t-value, Standardized Loading Factor, and Instrument Validity

Latent Variable	Indicator/ Item	t-value	SLF (λ)	Decision
Mathematical Literacy	Formulate the situation	9,44	0,97	Valid
	Use concepts, procedures, facts, and reasoning	8,23	0,96	Valid
	Interpreting and evaluating mathematical results	7,23	0,95	Valid
Formulate the situation	A1	*	0,59	Valid
	A3	4,64	0,44	Valid
Use concepts, procedures, facts, and reasoning	B1	*	0,62	Valid
	B2	5,04	0,40	Valid
	B3	6,42	0,58	Valid
Interpreting and evaluating mathematical results	C1	*	0,49	Valid
	C2	6,49	0,65	Valid
	C3	5,12	0,46	Valid
	C4	5,64	0,52	Valid

In addition to validity, reliability is also an indicator of whether an instrument is said to be good or not. Based on the coefficients loading factor and the unique error index for each item, the reliability coefficient can be estimated from the developed measurement model. Reliability is a coefficient that shows the level of consistency of the measurement results score. The calculation results Construct Reliability Mathematical literacy instruments are presented in the following Table 8.

Table 8. Construct Reliability of Mathematical Literacy Instruments

Item	SLF (λ)	$(\Sigma\lambda)^2$	Error (δ)	$\Sigma\delta$	Construct Reliability	Level of Reliability
A1	0,59		0,65			
A3	0,44		0,81			
B1	0,62		0,62			
B2	0,40		0,84			
B3	0,58	22,56	0,67	6,44	0,78	High
C1	0,49		0,76			
C2	0,65		0,57			
C3	0,46		0,79			
C4	0,52		0,73			

The reliability of a construct was estimated using formula (2). Based on Table 8, it appears that the instrument has high reliability. A high and low -reliability coefficient of a test is influenced by several factors related to the test, namely (1) number of items, (2) homogeneity of test material, (3) homogeneity of grain characteristics, and (4) variability score.

Conclusion

The developed mathematical literacy assessment instrument consists of 9 breakdown items that measure the achievement of mathematical literacy in three processes namely, 1) formulating the problem into a two-variable linear equation model, 2) using mathematical concepts, procedures, facts and reasoning in solving problems related to SPLDV, and 3) interpret, apply, and evaluate mathematical results. The instrument has evidence of content and constructs validity. The content validity is seen from the Aiken coefficient V index. Aikens coefficient V index of mathematical literacy instruments developed was in the range of 0.781 to 0.906 with a critical limit of 0.75. This provides evidence that the developed mathematical literacy instrument can measure what should be measured so that it is feasible to use to measure the achievement of students' mathematical literacy on the material system of two-variable linear equations.

The feasibility of the instrument is also evaluated from the construct validity. The construct that was developed was empirically confirmed by the analysis of the model fit. The analysis shows that the measurement model developed is

supported by the data. The items developed significantly reflect construct or latent variables and have good proof of validity. This is indicated by the magnitude of the coefficient loading factor > 0.3. Thus, it can be concluded that the developed mathematical literacy instrument can measure what should be measured, and the nine items significantly reflect construct or latent variables, as well as the level of consistency of good measurement results.

The results of the estimated coefficient of reliability with construct reliability indicate that the developed instrument has a high reliability of 0.78. This means that developed mathematical literacy instruments have a high degree of consistency in the measurement results.

Suggestions

This research resulted in nine items of mathematics literacy tests. The test item measures the ability of mathematical literacy only on algebraic content in the material system of two-variable linear equations. Therefore, further development can be done on other content and material. Assessment activities will produce good information if the instruments used to meet three criteria, namely valid, reliable, and objective. This study only proves the validity and estimates the reliability coefficient of the developed instrument. Further research can be done by doing an item analysis to estimate item parameters and capability parameters.

Acknowledgment

References

American Educational Research Association (2014). *Standards for educational and psychological testing*. AERA.

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and psychological measurement, 45*(1), 131-142.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole Publishing Company.

Anderson, L. W. (2003). *Classroom assessment: Enhancing the quality of teacher decision making*. Lawrence Erlbaum Associates.

Azwar, S. (1996). *Tes prestasi: Fungsi dan pengembangan pengukuran prestasi belajar* [Achievement tests: Function and development of learning achievement measurements]. Pustaka Pelajar.

Azwar, S. (2007). *Validitas dan reliabilitas* [Validity and reliability]. Pustaka Pelajar.

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment, and professional support. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 43-61). Springer.

Bishop, A. J., Clements, M. K., Clements, K., Keitel, C., Kilpatrick, J., & Laborde, C. (1996). *International handbook of mathematics education (Vol. 4)*. Kluwer Academic Publishers.

Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGraw-Hill Education (UK).

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*(1), 5. https://doi.org/10.1007/s11092-008-9068-5

Brown, T. A. (2015). *Confirmatory factor analysis for applied research.* Guilford Publications.

Burton, M., Silver, E. A., Mills, V. L., Audrict, W., Strutchens, M. E., & Petit, M. (2018). Formative assessment and mathematics teaching: Leveraging powerful linkages in the US context. In D. Thompson, M. Burton, A. Cusiie & D. Wright (Eds.), *Classroom assessment in Mathematics: Perspectives from around the globe* (pp. 193-205). Springer International Publishing.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Holt Rinehard and Winston Inc.

del Prado Hill, P., Friedland, E. S., & McMillen, S. (2016). Mathematics-literacy checklists: A pedagogical innovation to support teachers as they implement the common core. *Journal of Inquiry and Action in Education, 8*(1), 23-38. https://doi.org/10.1598/JAAL.55.1.6

Friyatmi, Mardapi, D., Haryanto, & Rahmi, E. (2020). The development of computerized economics item banking for classroom and school-based assessment. *European Journal of Educational Research, 9*(1), 293-303. https://doi.org/10.12973/eu-jer.9.1.293.

Ghozali, I., & Fuad (2015). *Structural equation modeling: Teori, konsep, dan aplikasi dengan Program Lisrel 9.10* [Structural equation modeling: Theories, concepts, and applications with the Lisrel 9.10 Program]. Diponegoro University Publisher Agency.

Gorsuch, R. L. (1983). *Factor analysis.* Psychology press.

Hidayati, K., Budiyono, & Sugiman (2018). Development and validation of student's responsibility scale on mathematics learning using subject scaling model. *International Journal of Instruction*, *11*(4), 499-512. https://doi.org/10.12973/iji.2018.11431a

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1-8.

Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher order thinking skills from elementary to master students in environmental learning. *European Journal of Educational Research, 8*(4), 935-942. https://doi.org/10.12973/eu-jer.8.4.935

Istiyono, E. (2018). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern* [Development of assessment instruments and analysis of physics learning outcomes with classical and modern test theory]. UNY Press.

Kemendikbud (2017). *Panduan penilaian oleh pendidik dan satuan pendidikan sekolah menengah pertama* [Guidance assessment by educators and middle school education units]. Ministry of Education and Culture.

Kline, P. (2014). *An easy guide to factor analysis*. Routledge.

Kuger, S., Klieme, E., Jude, N., & Kaplan, D. (2016). *Assessing contexts of learning: An international perspective*. Springer.

Kula Unver, S., Hidiroglu, C. N., Tekin Dede, A., & Bukova Guzel, E. (2018). Factors revealed while posing mathematical modelling problems by Mathematics student teachers. *European Journal of Educational Research, 7*(4), 941-952. https://doi.org/10.12973/eu-jer.7.4.941

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

Maskur, R., Sumarno, S., Rahmawati, Y., Pradana, K., Syazali, M., Septian, A., & Kinarya Palupi, E. (2020). The effectiveness of problem based learning and aptitude treatment interaction in improving mathematical creative thinking skills on Curriculum 2013. *European Journal of Educational Research*, *9*(1), 375-383. https://doi.org/10.12973/eu-jer.9.1.375

Matters, G. (2009). A problematic leap in the use of test data: From performance to inference. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 209-225). Springer. https://doi.org/10.1007/978-1-4020-9964-9.

Naga, D. S. (2012). *Teori sekor pada pengukuran mental* [Theory of scores on mental measurement]. Nagarani Citrayasa.

Nitko, A. J., & Brookhart, S. M. (2007). *Educational assessment of students.* Pearson Education, Inc.

Nunnally, J. C., & Bernstein, I. (1978). *Psychometric testing* (3rd ed.). McGraw-Hill.

Ojose, B. (2011). Mathematics literacy: Are we able to put the mathematics we learn into everyday use. *Journal of Mathematics Education, 4*(1), 89-100.

Organisation for Economic Co-operation and Development (2013). *PISA 2015 Draft Mathematics Framework*. OECD.

Organisation for Economic Co-operation and Development (2016). *Program for international students' assessment (PISA) results from PISA 2015*. OECD.

Owens, J. E. (1988, August). *Curriculum and evaluation standards for school mathematics: Report of the national council of teachers of mathematics' commission on standards for school mathematics*. Paper presented at the Summer Workshop of The Association of Teacher Educators. The University of Alabama.

Permendikbud (2018). *Peraturan menteri pendidikan dan kebudayaan nomor 37 tahun 2018 tentang perubahan atas peraturan menteri pendidikan dan kebudayaan nomor 24 tahun 2016 tentang kompetensi inti dan kompetensi dasar pelajaran pada kurikulum 2013 pada pendidikan dasar dan menengah* [Ministry of education and culture regulation number 37 of 2018 concerning amendment to the minister of education and culture regulation number 24 of 2016 concerning core competencies and basic competencies of learning in the 2013 curriculum in primary and secondary education]. Ministry of Education and Culture.

Retnawati, H. (2015). *Validitas reliabilitas dan karakteristik butir* [Validity, reliability, and item characteristics]. Parama Publishing.

Stacey, K., & Turner, R. (2015). The evolution and key concepts of the PISA mathematics frameworks. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy* (pp. 5-33). Springer international Publishing.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). American Council and Education.

Steen, L. A., Turner, R., & Burkhardt, H. (2007). Developing mathematical literacy. In W. Blum, P.L. Galbraith, H.-W. Henn & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI Study* (pp. 285-294). Springer.