Available online at JDR Website: http://journal.unublitar.ac.id/jdr
Journal of Development Research, Volume 8, Number 2, November 2024, Page 110
DOI: https://doi.org/10.28926/jdr.v8i2.398

Development of a Mathematical Literacy Test Instrument Based on Local Wisdom in Blitar Raya for Fifth-Grade Elementary Students

Rida Nanda Kirana[1,*], Surayanah[2], Siti Mas'ula[3]

Universitas Negeri Malang, Indonesia

E-mail: rida.nanda2101516@students.um.ac.id[1], surayanah.fip@um.ac.id[2], siti.masula.fip@um.ac.id[3]

*Corresponding Author

Abstract

This research aims to create and develop a mathematics literacy test instrument based on the local wisdom of Blitar Raya that is valid according to the opinions of material and instrument experts, and that is practical for educators and students in terms of validity, reliability, difficulty level, and the quality of distractors. The simplified eight-stage R&D model by Borg & Gall serves as the foundation for this research methodology. Data collection was obtained through unstructured observations, interviews, questionnaires for student needs analysis, student response surveys, and student feedback. The results of the expert validation of the instrument yielded 100% (very valid), and the expert validation of the content resulted in 93.75%. (very valid). In the small-scale test, 36 questions were found to be valid and 14 questions invalid, with 50 questions deemed reliable resulting in a reliability index of ri = 0.93. The difficulty index reached 0.61, categorized as moderate, and the discrimination power obtained an index of 0.60, categorized as good. Based on the practicality questionnaire results from teachers, it was 93.75%, and from students, it was 91.25%. In the large-scale test, 28 questions were valid and 8 invalid, with 36 questions reliable resulting in a reliability index of ri = 0.88. The difficulty index was 0.68, categorized as moderate, and the discrimination power obtained an index of 0.42, categorized as good. Based on the practicality questionnaire results from teachers, it was 90.62%, and from students, it was 94.6%.

Keywords: *elementary school, instruments test, mathematical literacy, mathematics*

Introduction

Mathematics is an important and interesting subject that must be studied from the elementary education level. In addition, the goal of mathematics subjects in elementary and secondary schools is to enhance students' abilities to think logically, analytically, systematically, critically, creatively, and collaboratively (Pradipta et al., 2020). Quoted from the Ministry of Education and Culture's website, mathematics in the Independent Curriculum at Elementary Schools includes several elements, one of which is the number element. Number concepts are fundamental material that students must master in order to learn more complex mathematics at the next level (Ramadhani, 2022). In addition, according to Ardianto (2020), the use of numbers, such as arithmetic operations in society, is very beneficial in solving various issues related to daily life. In the context of real-life applications of mathematics, students must possess the ability to formulate, use, and interpret mathematics, which is referred to as mathematical literacy skills (Habibi, H., & Suparman, 2020). Based on that presentation,

in mathematics learning, the concept of numbers aligns with the context of mathematical literacy skills.

Regarding mathematical literacy tests, education in Indonesia has two aspects of measurement, namely international and national standards. The international standards are *PISA (Program for International Student Assessment)* and *TIMSS. (Trend in International Mathematics and Science Study).* The latest TIMSS survey results from 2015, which involved fifth-grade elementary school students, revealed that Indonesian students scored only 397 points in mathematics, falling below the TIMSS Scale Centerpoint standard of 500 points. The low results of mathematical literacy tests are also evidenced by the PISA test, where the average OECD math score is 379, which is below the international average score of 490. (OECD, 2018). The results of the national test through the Minimum Competency Assessment (AKM) indicate that students' reading skills have only reached 50% proficiency. Meanwhile, students' ability in mathematics to answer correctly has only reached 33% (Detikedu, 2021). By looking at the results of national and international tests that have been explained, we must realize how important mathematical literacy test instruments are to help students understand and master mathematical literacy. Your mathematical literacy can train students' thinking patterns to formulate, use, and interpret various mathematical problems that arise in everyday life.

Through observations and interviews with fifth-grade teachers at several elementary schools in Blitar City, the researcher collected data for needs analysis. Based on these findings, the researchers identified several common threads related to the unidentification of the level of mathematical literacy mastery. This is due to a number of factors; first, the learning activities used do not aim to enhance students' mathematical literacy skills, as teachers do not provide opportunities for students to apply their analytical abilities to the problems presented. As a result, teachers do not give students the chance to understand concepts well. In addition, the second factor is due to the fact that the written test used as a tool to measure student learning outcomes does not include indicators of mathematical literacy, thus it cannot measure or assess the extent of students' mastery of mathematical literacy skills.

Furthermore, the researchers also conducted interviews and discussions with mathematics teachers at SDN 2 Sentul regarding the measurement of mathematical literacy skills. The results show that the questions used by teachers to assess students' learning outcomes do not meet the standards for each indicator of mathematical literacy. The questions provided only test students' procedural knowledge in using mathematical formulas and have not yet honed their mathematical literacy skills. One reason for the suboptimal mathematical literacy skills of students is the fact that the questions used in mathematics learning outcome tests are not effective in enhancing their mathematical literacy abilities (Aeni, 2020). In addition, regarding information on local wisdom-based learning, it is known that the implementation of learning has connected with local wisdom, but it has not been planned in writing. The researcher found that there is a discrepancy between the implementation of learning and its evaluation process. This is reinforced by research conducted by Zainuddin et al., (2024), which found that discussions with elementary school supervisorls in Blitar City revealed that cultural values, especially local wisdom, have not been incorporated into the curriculum. Based on this, there is a need for a mathematical literacy test instrument based on the local wisdom of Blitar Raya.

The development of a mathematical literacy test instrument based on local wisdom from Blitar Raya can incorporate the local wisdom of the entire Blitar region, which has not been included in the school curriculum until now. The instrument based on mathematical literacy integrated with local wisdom is important because it has unique characteristics that other instruments measuring students' understanding do not possess. This is in line with Farhatin et al (2020) statement, that schools can incorporate local wisdom values into learning through assessment processes that use instruments oriented towards local wisdom. In addition, the test

tools created can introduce the local wisdom of Blitar to the younger generation, so that it can be preserved through education to protect local wisdom from the influences of foreign cultures caused by technological advancement and free markets.

Based on the needs assessment data administered to fifth-grade students at SDN 2 Sentul, it was found that 100% of the students are aware of the local wisdom of Blitar Raya. As many as 88% of the students have studied mathematics related to local wisdom. A total of 96% of the students believe that mathematics learning based on the local wisdom of Blitar Raya can help them understand the material better. All students feel that learning based on local wisdom in mathematics will make the learning process more interesting and meaningful; however, 80% of the students find it difficult to connect the benefits of mathematics with real-life contexts, and 88% of the students acknowledge that the math problems have not tapped into their abilities to formulate, employ, and interpret mathematics in everyday life.

Based on the description of the needs above, a research development of a mathematical literacy test instrument was conducted with the title "Development of a Mathematical Literacy Test Instrument Based on Local Wisdom of Blitar Raya for Grade V Elementary School." The aim of this research is to produce a mathematical literacy test for fifth-grade elementary school students that is validated both logically and empirically, as well as practically. Logical validation is obtained through material questionnaires and expert assessments of the instrument. Empirical validation is obtained through tests of validity, reliability, distinguishing power, difficulty level, and the quality of distractors, as well as practical assessments from teachers and students.

Research Methods

The R&D (*Research and Development)* approach is applied in this research. R&D is a research methodology used to create and evaluate educational products (Okpatrioka, 2023). Due to time and budget constraints, the Borg & Gall research paradigm has been condensed into 9 steps. (1) potential and issues; (2) data collection; (3) initial product development; (4) product validation; (5) product revision I; (6) small-scale testing; (7), product revision II; (8) large-scale testing; and (9) final product. The research took place at UPT SDN Sentul 2 with a series of trials involving small and large groups. Ten students from class VA participated in the small group trial, while twenty-five students from class VB took part in the large group trial. The teachers of classes VA and VB at UPT SDN Sentul 2, as well as material and instrument experts, are among the specialists participating in this research. Student needs analysis questionnaire, logical validation questionnaire, empirical validation questionnaire, and student response questionnaire related to the local wisdom-based mathematical literacy test instrument of Blitar Raya are used to collect quantitative data.

Next, the obtained data will be processed, and the logical validation questionnaire will be measured using a Likert scale. Then, the results of the questionnaire will be processed using the formula referring to Akbar (2017).

$$Vah = \frac{Tse}{Tsh} x100\% \quad (1)$$

Description:
Vah = Expert validation
Tse = Total score obtained
Tsh = Total score that is expected.

The results of the processing are then categorized according to Akbar's (2017) criteria, as shown below.

Table 1. Product Validity Criteria

Assessment	Interpretation criteria	Test decision

85,01 – 100,00	*Very Valid*	*Can be used without revision*
70,01 – 85,00	*Quite Valid*	*Can be used but requires minor revisions*
50,01 – 70,00	*Less Valid*	*May be used with major revisions*
01,00 – 50,00	*Not Valid*	*Cannot be used*

Based on Table 1, it can be concluded that if the test instrument product receives a percentage rating above 71%, minor revisions can be made. However, if the test instrument product receives a percentage rating below 71%, major revisions can be carried out based on suggestions and critiques from instrument experts and content specialists. In addition, validity, reliability, discriminative power, difficulty level, and distractors are all determined by empirical validation. Microsoft Excel 2016 is used to calculate validity. Meanwhile, reliability testing, discriminating power, and trick power are calculated using SPSS Version 25, and the quality of the distractors is assessed with the help of Anates Version 4. Table 2 presents the criteria for item analysis in the following manner.

Table 2. Criteria for Item Analysis

Characteristics	Category	Interpretation criteria
Validity of test items	Valid	$r_{xy(\text{calculate})} > r_{xy(table)}$
	Invalid	$r_{xy(\text{calculate})} < r_{xy(table)}$
Reliability	Reliable	>0,70
	Unreliable	<0,70
Level difficulty	Difficult	P ≤ 0,30
	Moderate	0,30 < P ≤ 0,70
	Easy	P > 0.70
Differentiating factor.	Very good	0.71 – 1.00
	Good	0.41 – 0,70
	Satisfactory	0.21 – 0,40
	Bad	0.00 – 0,20
	Very bad	–
The quality of distractors	Excellent	++
	Good	+
	Not good	–
	Bad	– –
	Very bad	– – –

The next stage is the practicality test obtained from the practicality questionnaire that has been distributed to teachers and students. The practicality test is interpreted using a formula that refers to Hamdunah (2015) as follows.

Table 3. Product Practicality Criteria

Assessment	Interpretation criteria
80% < P ≤ 100%	Very Practical
60% < P ≤ 80%	Practical
40% < P ≤ 60%	Quite Practical
20% < P ≤ 40%	Less Practical
0 ≤ P ≤ 20%	Not Practical

Based on Table 3 above, theoretically, an item is considered valid if the practical percentage is ≥ 61.

Results and Discussions

The logical validity of the local wisdom based mathematical literacy test instrument in Blitar Raya

The following will present a recap of the results of logical validity that includes all aspects that have been validated by content experts and instrument experts. Both experts are lecturers at Malang State University in their respective fields. The outline for the development of test instruments on logical validity is modified from Setiawan et al. (2017); Widhiyani (2019); and Suarti. (2022). The overall recap of logical validity can be seen as follows.

Table 4. Recap of Expert Material Validation Results

Aspect	Question Item	Score	Percentage	Remarks
Presentation	The use of language is easily understood by students.	4	100%	Very valid
	The presentation of questions is appropriate and based on the local wisdom of Blitar Raya.	4	100%	Very valid
	The alignment of question categories with mathematical literacy.	3	75%	Quite valid
	As a practical and efficient instrument.	3	75%	Quite valid
Content quality	The accuracy of concepts.	4	100%	Very valid
	The accuracy of questions.	3	75%	Quite valid
	Measuring test instruments based on mathematical literacy.	4	100%	Very valid
Construction	The alignment of questions with test instruments based on the local wisdom of Blitar Raya.	4	100%	Very valid
	The presentation of questions follows the sequence of materials.	4	100%	Very valid
	Providing motivation for student learning.	4	100%	Very valid
Usage	The effectiveness of use.	4	100%	Very valid
	The practicality of using test instruments.	4	100%	Very valid
Total		48	1.125%	
Average			93,75%	
Category				Very valid
Description				Without revision

After adjustments were made according to the suggestions and input from subject matter experts, the results of the expert validation calculations, which have a percentage of 93.75% based on Table 5 above, are considered very valid and can be used without modification. The suggestions and input from the validator are as follows: 1) the indicators are the result of the development of the core competencies (KD), so the indicators should align with the learning objectives, 2) standardizing the reference questions used to ensure they are in accordance with the indicators or learning objectives, and 3) double-checking the numbering on the questions. The recommendations from the content expert are in line with Syahrita (2021), that the alignment of learning development with the indicators corresponds to the learning objectives, thus producing valid test items.

Overall, the test instrument product received valid scores and test decisions without revisions from subject matter experts, allowing it to be pilot tested. Four factors included in the criteria for evaluating material experts are presentation, content quality, construction, and utilization. The presentation aspects include 4 questions, namely, (1) the use of language that is easily understood by students, (2) the presentation of questions that is appropriate and based on the local wisdom of Blitar Raya, (3) the alignment of question categories with mathematical literacy, and (4) serving as a practical and efficient instrument. The validator provided a very valid category percentage of 87.5% in this case. One of the prerequisites for a strong mathematical literacy exam is to establish a connection between question indicators and the capacity to formulate, apply, and understand mathematical literacy process problems in real-world situations (Auliya, 2021).

The quality aspect of the content includes 3 questions: (1) the accuracy of the concepts, (2) the accuracy of the questions, and (3) measuring the test instruments based on mathematical literacy. The validator gave a percentage of 91.7% in the very good category for content quality. When a student faces real-world math problems, it is very important for them to understand the concepts and the accuracy of the problems (Romansyah & Nurhamdiah, 2018). The construction aspects include 3 questions, namely, (1) the alignment of the questions with the test instruments based on local wisdom of Blitar Raya, (2) the presentation of the questions in accordance with the sequence of the material, and (3) providing motivation for student learning. According to Sukmawati et al., (2022), through an evaluation process based on mathematical literacy skills that connects elements of the surrounding culture, it aims to monitor students' learning outcomes and foster a greater appreciation for their culture. This is intended so that children are capable and motivated to learn mathematics. Lastly, on the aspect of usage, there are 2 questions: (1) the effectiveness of use, and (2) the practicality of using the accuracy test instrument. In terms of usage, the validator provided a percentage of 100% with a category of very valid. The test instrument products have met the aspects of usability. Instruments are considered good if they offer convenience in terms of preparation, use or execution, interpretation of results, and storage (Setiawan et al., 2017).

Furthermore, the local wisdom-based mathematical literacy test instrument of Blitar Raya was validated by instrument experts. In the expert validation questionnaire for the instrument, there are 3 aspects: the aspect of the question grid which contains 1 statement, the aspect of the items which includes 2 main aspects, namely content validity and construct validity that consist of 5 questions, and the aspect of the scoring rubric and answer key which contains 3 statements. The results of the expert validation of the instrument are presented as follows.

Table 5. Summary of Expert Validation Results for the Instrument

Aspect	Question Item Guidelines	Score	Percentage	Remarks

Question framework	Can be used as a guide for designing question items	4	100%	Very valid.
Question items	Content Validity			
	The questions created align with the literacy-based math test instruments	4	100%	Very valid
	The questions created encompass local wisdom of Blitar Raya	4	100%	Very valid
	The questions created do not contain misconceptions	4	100%	Very valid
	Construct Validity			
	Framing statements as commands or questions that require responses	4	100%	Very valid
	Images or similar items are meaningful (clear descriptions or related to the issues being asked)	4	100%	Very valid
Assessment rubric and answer key.	Alignment of answer keys with the questions	4	100%	Very valid
	Appropriate weighting	4	100%	Very valid
	Final scoring is clear and detailed	4	100%	Very valid
Total		36	100	
Average			100%	
Category				Very valid
Description				Without revision

Table 5 above shows that the results of the expert validation calculations reaching 100% indicate that the developed mathematical literacy test instrument is highly valid and can be used without revisions after modifications were made according to the suggestions and feedback from subject matter experts. Based on the recommendations and input from the validators, the indication is that the mathematical literacy test instrument meets the necessary requirements. There are three components that serve as criteria for expert assessment of the material in the mathematical literacy test instrument, namely the question grid, the test items, and the assessment rubric along with the answer key. One question included in the test blueprint component is: Can the test instrument be used as a guideline for designing the writing of test items In the aspect of the test blueprint, it received a score of 100%, categorized as very valid. Regarding the analysis of test items according to Fitrianawati (2015), the quality of each component of the questions will improve when the question grid is created in accordance with the standards. In other words, the questions will serve as a tool to measure what needs to be measured and to describe every aspect of student learning or the subject matter that needs to be evaluated.

Next are the aspects of the items in the questions that contain 2 indicators, namely content validity and construct validity. Content validity includes 3 questions: (1) Are the questions created in accordance with the test instruments based on mathematical literacy? (2) Do the

questions incorporate local wisdom from Blitar Raya (3) Do the questions avoid misconceptions meanwhile, the construct validity indicator contains 2 questions, namely, (1) the creation of phrases that are commands or questions with the required responses, and (2) images or similar items that are meaningful. This false belief is related to how well students understand various fields of study, and the number of mistakes students make while completing assignments can serve as a measure of their content mastery. This aligns with Suarti (2022) opinion that when students use arithmetic to solve problems involving the interpretation of mathematical ideas, but they do not fully understand the concepts, this is known as misconceptions in mathematical problem-solving. Lastly, the assessment rubric and answer key contain 3 questions, (1) the alignment of the answer key with the questions, (2) appropriate weighting, and (3) clear and detailed final scoring. Based on the percentage results, it can be concluded that the assessment and weighting of the scores for the multiple-choice mathematical literacy test instrument are good. According to Putri et al., (2022), in the scoring guidelines for objective tests of the multiple-choice type, both teachers and students are not influenced by subjective factors in the examination and scoring.

Recapitulation of the analysis of test item instruments for mathematical literacy based on local wisdom of Blitar Raya

Small and large-scale trials are used to evaluate the test instruments that have been refined by experts during the logical validation stage. To create a valid mathematics literacy test instrument, 50 multiple-choice questions were analyzed and underwent a series of tests, including tests measuring reliability, discrimination power, and empirical validity, as well as the quality of distractors. The results of the item analysis from the small and large-scale trials are summarized as follows.

Small-scale testing

Small-scale testing is conducted to minimize the occurrence of errors in test instrument products, and the students' answers are used to determine whether the questions are valid for use or should be discarded. The following data on the recap of empirical validity, reliability, discriminative power, and the quality of distractors. The results of the small-scale trial recap, indicate that there are 36 valid questions and 14 invalid questions, which means that the invalid questions will not be used or have been eliminated. In addition, the 50 test instrument questions are said to be reliable. The difficulty index of the test item is 0.62, placing it in the moderate category. Furthermore, the discrimination power received an index value of 0.60, which falls into the good category. Each result from the item analysis calculations in the small-scale trial will be explained in detail below.

The empirical validity of the test items is evaluated to ensure that the items created and tested are valid. The validity test was conducted using MS Excel with Pearson correlation, and the results were compared to the r-table at a significance level of 5%. The correlation r-table tested on 10 students was valued at 0.632. If $r_{xy} > r_{tabel}$, then the question is declared valid and used, whereas if $r_{xy} < r_{tabel}$, the question is declared invalid and discarded. Based on the table above, the results of the empirical validity calculation for 50 items showed that 36 questions are valid and suitable to proceed to field testing, while 14 questions are declared invalid and discarded. The valid and suitable questions for large-scale testing are questions number 1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 40, 42, 43, 48, and 50. Question number 50 has the highest coefficient of 0.84, while question number 3 has the lowest coefficient of -0.20. The components of mathematical literacy for the small-scale test include 15 items for the formulate indicator, 14 items for the employ indicator, and 7 items for the interpret indicator.

To ensure the consistency of the tests to be administered, a reliability test of the items is necessary. If a test consistently produces the same findings for the same group over several periods, then the test can be considered reliable (Ida, F. F., & Musyarofah, 2021). Reliability in this study is determined using the split-half model in SPSS version 25. There are fifty questions. A reliability coefficient value of the test instrument greater than 0.70 indicates that the test instrument is considered reliable, while a reliability coefficient value of the test instrument less than 0.70 indicates that the test instrument is not reliable. The following is an interpretation of the calculation results of the reliability coefficient of the test (r_i). Based on the research, 50 items were obtained with $r_i = 0.931$. Based on the local wisdom of Blitar Raya, the instrument for testing the mathematical literacy of elementary school students has a reliability coefficient value of more than 0.70 according to the calculation results. This means that the test is reliable and suitable for use in large-scale testing.

The difficulty level test is needed in the test instrument to provide information on the difficulty level of the administered test. Good instruments should be at a level that is neither too difficult nor too easy (Indrawati et al., 2022). This is supported by the opinion of Khasanah et al,. (2023), which states that the difficulty of a test should ideally be balanced or proportional. Six questions fell into the difficult category, 39 questions were categorized as medium, and five questions were classified as easy based on the difficulty assessment of 50 questions rated on a small scale. The average difficulty level of the small-scale trial is often calculated to be 0.62, placing it in the medium or balanced trial category.

To what extent each question can differentiate between high-achieving students who have mastered the material and low-achieving students who have not yet grasped the material, that is the basis for testing the discriminative power. To identify the items in the questions that are classified as very poor, poor, fair, good, or very good in terms of their discriminative power, this study calculates the discriminative power using *SPSS* version 25 *alpha model*. There are 19 questions that fall into the very good category, 17 questions in the good category, 3 questions in the poor category, and 11 questions in the very poor category, according to the results of the discrimination power test in the small-scale trial. The category is very good, and the average is 0.60. Students with low abilities and students with high abilities can be distinguished by questions that have a strong differentiating power (Akhmadi, 2021).

The effectiveness of the quality of distractors depends on how well the incorrect choices can mislead test takers who are unaware of the correct answers available (Iskandar & Rizal, 2018). The more test participants choose that distractor, the more it can be said that the distractor is good. In this study, the quality of distractors was assessed using Anates version 4. Based on the results, it was found that the quality of the distractors is classified as good; however, there are 13 items that have poor distractors in the invalid questions. The quality of distractors categorized as poor and very poor is due to the distractors being prominently visible and heterogeneous, making them unlikely to be chosen by test participants who have a limited understanding of the material.

Based on the recap of the results from the small-scale trial analysis, 36 valid questions were obtained and can be used in the large-scale trial. The valid questions include material on number sense covering questions 1, 2, 4, 6, 7, 8, and 9. The material on operations with whole numbers includes questions 10, 11, 12, and 13. The material on operations with whole numbers in a one-variable diagram includes questions 15 and 17. The material on money operations includes questions 18, 19, 22, 23, 24, and 26. The material on the greatest common divisor (FPB) includes questions 27, 28, 29, and 30. The material on the least common multiple (KPK) includes questions 31, 32, 33, and 34. The material on fractions with the same numerator includes questions 35 and 36. The material on mixed operations with fractions includes questions 38, 40, 42, 43, 48, and 50. The valid questions will be used in a large-scale trial.

Large-scale trial

The large-scale trial was conducted at SDN 2 Sentul Kota Blitar with 25 fifth-grade students as subjects. The results of the empirical validity, reliability, discriminative power, and quality of distractors. The results of the large-scale trial recap show that there are 28 valid items and 8 invalid items, which will not be used or have been rejected. Furthermore, the 36 test items have a reliability value of 0.88, indicating that the instrument is dependable. The difficulty index of the test item is 0.68, which places the test item in the moderate category. Furthermore, the discrimination power received an index value of 0.42, categorizing it as good. The results of the item analysis calculations from the large-scale trial will be presented below.

The validity test of the test items was conducted using *MS Excel with Pearson correlation,* the results of which were compared to the r-table at a significance level of 5%. The correlation r-table tested on 25 students was valued at 0.396. If $r_{xy} > r_{tabel}$, then the question is considered valid and used, whereas if $r_{xy} < r_{tabel}$, the question is deemed invalid and discarded. Based on the table above, the results of the empirical validity calculation on 36 items show that 28 items are valid, while 8 items are declared invalid and discarded. The valid questions in the large-scale trial are questions number 1, 2, 6, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 26, 27, 29, 31, 33, 34, 36, 38, 40, 42, 43, 48, and 50. Question number 25 has the highest coefficient of 0.88, while question number 23 has the lowest coefficient of -0.23. The components of mathematical literacy in large-scale tests include 12 items for the formulate indicator, 10 items for the employ indicator, and 6 items for the interpret indicator. This is in line with the research findings of Solichin (2017) that a test can be considered valid if it truly measures what needs to be measured using statistical analysis, so that the identified valid questions can be tested and used, and vice versa.

Reliability testing is measured with the help of SPSS version 25 using the split-half model. The interpretation of the calculation results of the reliability coefficient (r_i) is that if the value of the reliability coefficient of the test instrument is greater than 0.70, it means that the test instrument is considered reliable. However, if the value of the test instrument's coefficient is less than 0.70, it indicates that it is not reliable. The analysis results show that from 36 items, a reliability index (r_i) of 0.881 was obtained. This calculation indicates that the reliability coefficient of the local wisdom-based mathematical literacy test instrument for elementary school students in Blitar Raya is greater than 0.70, thus it is declared reliable. The reliability of a test is defined as a measure of the consistency of a subject's scores when assessed under different conditions using the same measuring instrument (Syahnita, 2021). The results of the statistical analysis indicate that the local wisdom-based mathematics literacy test instrument of Blitar Raya is reliable due to its ability to evaluate the consistency of the test items.

To determine the difficulty level of 36 questions that were widely tested, the item difficulty index in this study was calculated using *SPSS version 25.* As a result, 15 questions fall into the easy category. Twenty questions have a moderate level of difficulty, while one question is classified as difficult. The overall average difficulty level of the small-scale experiment falls into the moderate category with a score of 0.68. According to Alwi (2015), the test items indicate that the closer the score is to 1, the easier the question is, and the further away from 0, the more difficult the question becomes. The conclusion of this mathematical literacy test instrument is at a moderate level, thus categorized as good and suitable for use. To determine the items that have categorization of item discrimination as very poor, poor, fair, good, and very good, the item discrimination in this study was determined using SPSS version 25 and the alpha model. Eight items fell into the very good category, eighteen items fell into the good category, two items fell into the fair category, four items fell into the poor category, and four items fell into the very poor category, in accordance with the results of the discrimination test in the small-scale trial. With a sufficient category, the average index of item discrimination for mathematical literacy questions is 0.42. In line with Solichin's research

(2017) that a good test instrument has an interpretation range of 0.41 – 0.70, the test questions are deemed suitable for use.

A good distractor quality is one that is not the correct answer but makes test participants interested in choosing that option (Amelia, 2016). Among the test participants, at least 5% chose the very good example. According to Arikunto (2008), an exception will work better if more test participants choose it. On the other hand, a decoy is not used if no one chooses it. Revisions need to be made for two questions categorized as poor in the valid questions and four questions with misleading and very misleading options in the invalid questions, based on the recap results. Good and very good quality test items based on local wisdom of Blitar Raya in the mathematical literacy test instruments need to be maintained, while the less effective items can be improved by focusing on other responses that align with the correct answers. Meanwhile, all the questions that fall into the poor and very poor categories need to be improved.

Based on the recap of the results from the large-scale trial analysis, 28 questions were found to be valid and 8 questions were invalid. The valid questions contain material on number sense, including numbers 1, 2, 6, 7, 8, and 9. The material on operations with whole numbers includes numbers 10, 11, 12, and 13. The material on operations with whole numbers in a one-variable diagram includes number 17. The material on money operations includes numbers 19, 22, 24, and 26. The material on the greatest common divisor includes numbers 27 and 29. The material on the least common multiple includes numbers 31, 32, 33, and 34. The material on fractions with the same numerator includes numbers 35 and 36. The material on mixed operations with fractions includes numbers 38, 40, 42, 43, 48, and 50. The valid questions will be used in a large-scale trial. Each item in the test uses indicators of mathematical literacy, namely formulate, employ, and interpret.

Practicality test of local wisdom based mathematical literacy instruments in Blitar Raya

The source of data on practicality in this research was obtained from the recapitulation of practicality questionnaires distributed to teachers and students at SDN 2 Sentul after they completed and analyzed the local wisdom-based mathematical literacy test instruments of Blitar Raya. The results will be presented in the following description.

Practicality test of the questionnaire (teacher)

The results practicality assessment of teachers evaluated using small-scale and large-scale products based on local wisdom elements from Blitar Raya received a score of 92.18% and falls into the very practical category, with the decision that the trial is suitable for use without modifications. The comment from Teacher I is that the local wisdom-based mathematical literacy test of Blitar Raya has good and complete number elements, a format that is easy to understand, and can provide knowledge to students regarding the local wisdom of Blitar Raya. In addition, the indicators used have been adjusted to the Learning Outcomes (CP) of Phase C in grade V. Meanwhile, the comment from Teacher II is that the mathematical literacy test instrument is very well used as an evaluation tool for learning outcomes by linking it to local wisdom.

According to the research by Jamiel et al,. (2022), the usefulness of the developed product can be measured by how well it connects with the surrounding culture and inspires students to preserve it. Therefore, the practicality aspect of the local wisdom-based mathematical literacy test instrument developed in this study can provide knowledge about local wisdom to students. Regarding the review of the abilities that the local wisdom-based mathematical literacy test instrument aims to achieve, it is appropriately used at the elementary school level, in line with the opinion of Chesa & Binti Azizatun Nafi'ah (2022) that the indicators of literacy and numeracy evaluation instruments should be adjusted to the students'

ability levels as well as their educational stage. The aspects of the suitability of the local wisdom-based test instrument format include grids, usage instructions, answer keys, and scoring techniques. The instruments must be practical and formative, meaning they should include outlines, usage instructions, answer keys, and good and complete assessment techniques that are easy to use in any condition. (Pamungkas, 2018). Based on the elaboration of the discussion and the comments from the practicality questionnaire by the teacher, it can be concluded that the local wisdom-based mathematical literacy test instrument for the number elements of fifth-grade schools in Blitar Raya can be considered practical and suitable for use.

Practicality test of the questionnaire (students)

The following presents a recap of the results of the practicality test conducted with students in a small-scale trial filled out by 10 students and a large-scale trial filled out by 25 students at SDN 2 Sentul. Modified student practicality lift kits from Rismayanti, T. A., Anriani, N., & Sukirwan (2022) and Prafianti et al., (2022). Based on the student practicality questionnaire, the results showed that the practicality score for the small test was 93%, while the score for the large test was 94.6%. Both results fall into the very practical category. Based on that category, it can be concluded that the local wisdom-based mathematical literacy test instrument of Blitar Raya is very practical for use by both students and teachers. There are three aspects in the student practicality questionnaire, namely interest, material, and language.

In terms of interest, achieving results of 93% and 99% indicates that students are very interested in the local wisdom-based mathematical literacy test instruments of Blitar Raya. The sub-indicators of interest include students' motivation and enthusiasm for learning, clear test instrument guidelines, and students feeling excited and not bored while working on the local wisdom-based mathematical literacy test instruments of Blitar Raya. This aligns with the opinion of Eliza, F., Warlizasusi, J., & Warsah (2021) that generally, students feel bored when taking evaluation tests; however, this can be addressed by enhancing students' learning motivation through engaging test instruments, making students feel challenged in their work. Furthermore, the material aspect achieved results of 89% and 90%, indicating that students are very interested in the content of the local wisdom-based mathematical literacy test instrument of Blitar Raya. The sub-indicators of the material include questions that align with the material being studied, questions and statements that are easy to understand, questions that can be easily processed, and the allotted time for completion being appropriate. Lastly, the language aspect received scores of 91% and 95%. It can be concluded that the language used in the local wisdom-based mathematical literacy test instrument products of Blitar Raya is easy to understand and can attract students' interest. Language aspects are the most important factors in supporting the quality of a test instrument; in this regard, there needs to be alignment between test participants and their level of language comprehension (Sa’idah, N., Yulistianti, H. D., & Megawati, 2019).

Conclusion

Based on the discussion of the mathematical literacy test instrument product based on local wisdom of Blitar Raya, the number element for fifth-grade elementary school received a material expert validation score of 93.75% (very valid). This indicates that the product test instrument for mathematical literacy is suitable for use without modification. The validation result from the material expert is 100% (very valid). This shows that the content in the mathematical literacy test instrument product is appropriate for use without modification. The analysis results of the test items from the small-scale trial showed that there were 36 valid questions and 14 invalid questions, which will not be used or discarded. In addition, the 50 test instrument items are said to be reliable with a value of 0.93. The difficulty index of the test items is 0.62, placing them in the moderate category. Furthermore, the discriminative power

received an index value of 0.60, which falls into the good category. Based on the percentage results of the practicality questionnaire according to teachers, it is 93.75%, and according to students' practicality questionnaire, it is 91.25%, both categorized as very practical. There are 28 valid questions and 8 invalid questions in the large-scale test obtained. In addition, the 36 test questions have a reliability score of 0.88, indicating that the instrument is reliable. The difficulty index of the test item is 0.68, which places the test item in the moderate category. Furthermore, the discrimination power received an index value of 0.42, categorizing it as good. The percentage results of the practicality questionnaire according to teachers is 90.62%, and according to the students' practicality questionnaire is 94.6%, both categorized as very practical and suitable for use without revisions. It can be concluded that the local wisdom-based mathematical literacy test instrument of Blitar Raya, specifically the Number element, is very valid and highly practical based on the results of empirical and logical validation tests as well as item analysis. Therefore, it can be used to familiarize and train students with questions based on mathematical literacy and is beneficial in incorporating the local wisdom of Blitar Raya. Suggestions or recommendations for other development research include expanding the test instrument product to cover a broader scope of mathematical literacy, differing in material, grade levels, and local wisdom from other regions.

References

Aeni, D. N. (2020). Literasi Matematika Berdasarkan Kemampuan Matematika Siswa Kelas Vi Sdn Jember Lor 05. *Skripsi, Jember: Fakultas Keguruan Dan Ilmu Pendidikan Universitas Jember*, 11.

Akbar, S. (2017). Instrumen Perangkat Pembelajaran. In *Remaja Rosdakarya.*

Akhmadi. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 Sdn Plumbungan Menggunakan Program Anates. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, *6*(1), 799–806. https://doi.org/10.33752/ed-humanistics.v6i1.1464

Alwi, I. (2015). Pengaruh Jumlah Alternatif Jawaban Tes Obyektif Bentuk Pilihan Ganda Terhadap Reliabilitas, Tingkat Kesukaran Dan Daya Pembeda. *Jurnal Ilmiah Faktor Exacta*, *3*(2), 184–193. https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/19/18

Amelia, M. A. (2016). Analisis Soal Tes Hasil Belajar High Order Thinking Skills (Hots) Matematika Materi Pecahan Untuk Kelas 5 Sekolah Dasar. *Jurnal Penelitian*, *20*, 123–131.

Ardianto, W. (2020). Konsep Dasar Bilangan dan Pembelajarannya di Sekolah Dasar. Semarang: Qahar Publisher. *Semarang: Qahar Publisher*.

Arikunto, S. (2021). Dasar-dasar evaluasi pendidikan edisi 3. *Bumi Aksara*.

Auliya, L. (2021). Pengembangan Instrumen Tes Literasi Matematis Berbasis Budaya Betawi Untuk Siswa SMP/MTs. *Repository.Uinjkt.Ac.Id*, *4*(1), 1–23. https://repository.uinjkt.ac.id/dspace/handle/123456789/59007%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/59007/1/11150170000024_Lisa Auliya Water mark.pdf

Chesa, N., & Binti Azizatun Nafi'ah. (2022). Asesmen Kompetensi Minimum (Akm) Kelas Sekolah Dasar Sebagai Sarana Evaluasi Kemampuan Literasi Dan Numerasi Siswa. *Jurnal Pendidikan Dasar*, *13*(2), 67–86. https://doi.org/10.21009/jpd.v13i2.28482

Detikedu. (2021). *Asesmen Kompetensi Minimum 2021*.

Eliza, F., Warlizasusi, J., & Warsah, I. (2021). Manajemen Pembelajaran Berbasis E-Learning Masa Covid 19 . *(Doctoral Dissertation, IAIN Curup)*.

Farhatin, N., Pujiastuti, H., & Mutaqin, A. (2020). Pengembangan Bahan Ajar Matematika Berbasis Kearifan Lokal Untuk Siswa Smp Kelas Viii. *Prima: Jurnal Pendidikan Matematika*, *4*(1), 33. https://doi.org/10.31000/prima.v4i1.2082

Fitrianawati, M. (2015). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal,

Kompetensi Guru Dan Hasil Belajar Peserta Didik. *Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa*, *5*(3), 282–295.

Habibi, H., & Suparman, S. (2020). Literasi Matematika dalam Menyambut PISA 2021 Berdasarkan Kecakapan Abad 21. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *6*(1), 57–64.

Hamdunah. (2015). Praktikalitas pengembangan modul kontruktivisme dan website pada materi. *Lemma*, *II*(1), 35–42. https://ejournal.upgrisba.ac.id/index.php/jurnal-lemma/article/view/524

Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *AL-MUARRIB : Journal Of Arabic Education*, *1*(1), 34–44.

Indrawati, I., Yuniawatika, Y., & Suminah, S. (2022). Pengembangan Instrumen Tes Setipe PISA untuk Melatih Kemampuan Literasi Matematika Siswa Kelas V SD LAB UM. *Jurnal Pembelajaran, Bimbingan, Dan Pengelolaan Pendidikan*, *2*(7), 629–639. https://doi.org/10.17977/um065v2i72022p629-639

Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(1), 12–23. https://doi.org/10.21831/pep.v22i1.15609

Khasanah, I., Fuady, A., & Sunismi. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. *Mathema Journal*, *5*(2), 110–125.

OECD. (2018). *Draft Mathematics Framework. Paris: OECD Publishing.* PISA.

Okpatrioka, O. (2023). Research and development (R&D) penelitian yang inovatif dalam pendidikan. Dharma Acariya Nusantara. *Jurnal Pendidikan, Bahasa Dan Budaya*, *1*(1), 86–100.

Pamungkas, A. S. (2018). Pengembangan Instrumen Kecakapan Matematis Dalam Konteks Kearifan Lokal Budaya Banten Pada Materi Bangun Ruang Sisi Datar. *Jurnal Matematika Kreatif-Inovatif*, *9*(1), 17–27.

Pradipta, Sariyasa, & Lasmawan. (2020). Pengembangan Instrumen Kemampuan Berpikir Kreatif dan Literasi Matematika Pada Materi Geometri Peserta Didik Kelas IV Sekolah Dasar. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *10*(1), 21–30.

Prafianti, R. A., Anjarsari, E., & Ramadhani, A. (2022). Kepraktisan dan Keefektifan Lembar Kerja Siswa Materi Lingkaran Berbasis Pendekatan Matematika Realistik. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *7*(2), 265. https://doi.org/10.30998/jkpm.v7i2.12797

Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen Penilaian Hasil Pembelajaran Kognitif pada Tes Uraian dan Tes Objektif. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, *4*(2), 139–148. https://doi.org/10.36232/jurnalpendidikandasar.v4i2.2649

Ramadhani, W. P. (2022). Perbedaan penguasaan materi operasi bilangan bulat siswa SMP di desa dan di kota menggunakan teknik berhitung cepat. *Jurnal Abdimas Indonesia*, *2*(2), 216-222. https://doi.org/10.53769/jai.v2i2.239.%0A

Rismayanti, T. A., Anriani, N., & Sukirwan, S. (2022). Pengembangan e-modul berbantu kodular pada smartphone untuk meningkatkan kemampuan berpikir kritis matematis siswa. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *6*(1), 859–873.

Romansyah, F. ., & Nurhamdiah. (2018). Profil Pemahaman Konsep Siswa Sekolah DasarDalam Menyelesaikan Soal Luas Dan KelilingLingkaran. *Jurnal Pendidikan Tambusai*, *2*(6), 1703–1709. https://jptam.org/index.php/jptam/article/view/160/149

Sa'idah, N., Yulistianti, H. D., & Megawati, E. (2019). Analisis instrumen tes higher order thinking matematika SMP. *Jurnal Pendidikan Matematika*, *13*(1), 41–54.

Setiawan, H., Sa, C., Akbar, D., Artikel Abstrak, I., & Setiawan Pendidikan Dasar, H. (2017). Pengembangan Instrumen Asesmen Autentik Kompetensi Pada Ranah Keterampilan Untuk Pembelajaran Tematik Di Sekolah Dasar. *Jurnal Pendidikan: Teori, Penelitian, &*

Pengembangan, *2*(7), 874–882. https://doi.org/ttp://dx.doi.org/10.17977/jptpp.v2i7.9602

Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Dirāsāt: Jurnal Manajemen & Pendidikan Islam*, *2*(2), 192–213. www.depdiknas.go.id/evaluasi-proses-

Suarti, N. W. (2022). Pengembangan Instrumen Penilaian Tematik Berbasis HOTS untuk Peserta Didik Kelas IV Sekolah Dasar. *Indonesian Journal of Educational Development*, *2*(4), 534–548. https://doi.org/10.5281/zenodo.6178829

Sukmawati, D., Anggoro, B. S., & Pratiwi, D. D. (2022). Pengembangan Instrumen Evaluasi Literasi Matematis Berdasarkan Perspektif Multiple Intelligences Berbasis Etnomatematika Pada Budaya Jawa. *Jurnal Educatio FKIP UNMA*, *8*(4), 1215–1226. https://doi.org/10.31949/educatio.v8i4.3172

Syahnita, R. (2021). Strategi Pembelajaran dan Kemampuan Guru Kelas. *Modul Biokimia Materi Metabolisme Lemak, Daur Asam Sitrat, Fosforilasi Oksidatif Dan Jalur Pentosa Fosfat*, *4*, 6.

Widhiyani, I. A. N. T., Sukajaya, I. N., & Suweken, G. (2019). Pengembangan Soal Higher Order Thinking Skills Untuk Pengkategorian Kemampuan Pemecahan Masalah Geometri Siswa Smp. *Jurnal Pendidikan Dan Pembelajaran Matematika Indonesia*, *8*(2), 68–77. https://doi.org/10.23887/jppm.v8i2.2854

Zainuddin, M., Surayanah, S., & Lestariningsih, L. (2024). Pengembangan Panduan Keterampilan Proses Tema Daerahku dan Keindahan Alamnya Berbasis Kearifan Lokal Pada Guru SD di Blitar. *Briliant: Jurnal Riset Dan Konseptual*, *9*(1), 74. https://doi.org/10.28926/briliant.v9i1.1648