PLUSMINUS
Jurnal Pendidikan Matematika

Improving Students' Numeracy Skills Through the Development of Diagnostic and Evaluation Instruments for Numeracy Proficiency

Nofriyandi[1*], Dedek Andrian[2], Abdurrahman[3], Akbar Septiawan[4], Febri Loska[5], Siti Nurhalimah[6]
[1*,2,3,4,5,6] Mathematics Education, Universitas Islam Riau
Jalan Kaharudin Nst No. 113, Simpang Tiga, Pekanbaru, Riau, Indonesia
[1*]nofriyandi@edu.uir.ac.id; [2]dedekandrian@edu.uir.ac.id; [3]abdurrahman@edu.uir.ac.id;
[4]akbarseptiawan404@gmail.com; [5]febriloska@student.uir.ac.id; [6]sitinurhalimah951@student.uir.ac.id

ABSTRAK

Numerasi adalah kemampuan penting bagi siswa dan merupakan dasar untuk kesuksesan di masa depan. Penelitian ini bertujuan untuk mengembangkan instrumen dan mengevaluasi keterampilan numerasi siswa di 10 sekolah di Pekanbaru. Metode penelitian yang digunakan adalah metode R & D dan Survei. Populasi penelitian ini adalah seluruh siswa sekolah menengah pertama di Pekanbaru. Sampel penelitian terdiri dari 430 siswa yang dipilih dengan metode multiple random sampling. Teknik pengumpulan data yang digunakan dalam penelitian ini adalah pendekatan survei, dengan instrumen utama berupa tes pilihan ganda. Tujuh analisis data digunakan dalam penelitian ini: validitas konten, reliabilitas Cronbach Alpha, Analisis Faktor Konfirmatori (CFA), reliabilitas konstruk, dan deskriptif. Hasil penelitian menunjukkan bahwa analisis validitas konten terdapat empat item yang perlu direvisi karena kalimat yang ambigu. Reliabilitas Cronbach Alpha diperoleh sebesar 0,872 dengan kategori yang diterima. Analisis CFA menunjukkan bahwa 19 indikator sebagai konstruk merupakan kategori yang valid. Secara keseluruhan, keterampilan numerasi siswa dari dua belas kabupaten/kota di Provinsi Riau masih berada pada kategori yang kurang baik.

Kata Kunci: Tes Diagnostik; Kemampuan Numerasi; Budaya Melayu; Sekolah Menengah Atas; Pekanbaru

ABSTRACT

Numeracy is a crucial skill for students and serves as a foundation for future success. This study aims to develop instruments and evaluate students' numeracy skills in 10 schools in Pekanbaru. The research employed a combination of Research and Development (R&D) and survey methods. The population consisted of all junior high school students in Pekanbaru. A total of 430 students were selected as samples using multiple random sampling techniques. Data were collected through a survey approach, with the main instrument being a multiple-choice test. Seven types of data analysis were employed: content validity, Cronbach's Alpha reliability, Confirmatory Factor Analysis (CFA), construct reliability, and descriptive statistics. The results indicated that four items required revision due to ambiguous wording based on the content validity analysis. The Cronbach's Alpha reliability was 0.872, which is considered acceptable. CFA analysis confirmed that 19 indicators were valid as constructs. Overall, the students' numeracy skills from twelve regencies/cities in Riau Province are still in the low category.

Keywords: Diagnostics test; Numeracy Skill; Diagnostics test; Malay Culture; Junior High School; Pekanbaru.

Article Information:

Accepted Article: 14 January 2025, Revised: 18 February 2025, Published: 30 March 2025

How to Cite:

Nofriyandi, Andrian, D., Abdurrahman, Setiawan, A., Loska, F., & Nurhalimah, S. (2025). Improving Students' Numeracy Skills Through the Development of Diagnostic and Evaluation Instruments for Numeracy Proficiency. *Plusminus: Jurnal Pendidikan Matematika, 5*(1), 155-172.

DOI: https://doi.org/10.31980/plusminus.v5i1.2552

1. INTRODUCTION

Numeracy skills are essential competencies in the 21st century (Nofriyandi, Abdurraman, & Andrian, 2023; Andrian et al., 2024). In line with the goals of the Indonesian Ministry of Education and Culture (Kemdikbud RI), the necessary skills for the 21st century include literacy and numeracy skills, knowledge, skills, affective abilities, and technological proficiency." (Prendergast et al., 2023). Numeracy skills are the most important part of the educational process (Aunio et al., 2019; Tito, Muhtadi, & Sukirwan, 2024). Numeracy skills are the foundation for learning more complex mathematics subjects in school (Gashaj et al., 2023). Numeracy is the ability to formulate, use, and interpret mathematics in various contexts. It encompasses mathematical reasoning, the application of concepts, procedures, facts, and mathematical tools to describe, explain, and predict phenomena. Numeracy skills enable individuals to make constructive and reflective decisions (Chung & Yoo, 2015; OECD, 2018). This understanding implies that numeracy skills are not limited to mastering mathematical content but also involve the application of reasoning, concepts, facts, and mathematical tools to solve real-life problems (Segers et al., 2015; Safitra, Mulyono, & Susanti, 2023).

Students who demonstrate sensitivity in applying appropriate mathematical concepts to problem-solving can be considered to have better numeracy skills. Numeracy requires individuals to understand, analyze, interpret, evaluate, and synthesize information obtained from problems, which is then modeled mathematically and solved effectively using mathematical concepts (Persson et al., 2021). Through numeracy skills, students also learn to think critically, analyze information, and develop well-founded arguments (Bitterly et al., 2022). These skills can be applied across various aspects of life, both personal and professional (Wannenburg & Curlewis, 2023). With strong numeracy skills, individuals can be more independent in managing personal finances, make smarter decisions, and enhance their overall quality of life (Val é rio & Ferrara, 2022).

The importance of numeracy extends beyond personal life, as it also plays a significant role in students' academic success. (Peters et al., 2021). Numeracy serves as the foundation for various subjects, including mathematics, science, economics, and technology. Students with strong numeracy skills tend to grasp complex concepts more easily and are better able to connect theory with practical applications (Mimeau et al., 2016; Jayanti & Cesaria, 2024). Furthermore, numeracy also supports the development of analytical thinking skills, which are essential for interdisciplinary learning.

However, in Indonesia, students' numeracy skills remain a significant challenge that requires serious attention. Various studies indicate that many Indonesian students struggle to understand fundamental concepts and apply them in real-world contexts (Qolbi & Afriansyah, 2024). This

issue not only affects students' academic performance but also impacts their preparedness to face the increasingly competitive job market. One of the main factors contributing to the low numeracy skills among Indonesian students is the unequal quality of education. In remote areas, many schools lack essential resources, including facilities, textbooks, and qualified teachers. Additionally, inadequately trained teachers in effective teaching methods pose another barrier to the development of students' numeracy skills (Zaki et al., 2024). In addition to educational factors, the socio-economic conditions of families also play a role in influencing students' numeracy skills. Children from low-income families often face limited access to supplementary learning resources, such as private tutoring or additional coaching. They also tend to have less study time due to the need to help their parents earn a living, which ultimately affects their academic development. Cultural factors should also not be overlooked when discussing numeracy issues in Indonesia. There is a prevailing societal belief that certain subjects are inherently difficult and can only be mastered by a select few with exceptional talent. This stereotype can reduce students' motivation to learn, create anxiety toward specific subjects, and hinder the development of their numeracy skills.

The Indonesian education curriculum, which still emphasizes rote memorization and procedural mastery without deep understanding (Oktaviana, Effendi, & Rosyadi, 2023), is also a contributing factor to the low numeracy skills. Students are often taught to memorize facts and problem-solving steps without grasping the underlying concepts (Krisma, Muqtada, & Khasanah, 2024). As a result, they struggle to apply their knowledge in different contexts or real-life situations.

To address this issue, an integrated effort from various stakeholders—including the government, educators, parents, and the community—is necessary. Improving teacher training quality, developing a curriculum based on conceptual understanding, and ensuring equal access to education are crucial steps that must be taken (Effendi, Ummah, & Cahyono, 2023). By doing so, it is expected that Indonesian students can develop strong numeracy skills, which will not only support their academic success but also prepare them to face future challenges.

The novelty of this study lies in its comprehensive and context-sensitive approach to evaluating numeracy skills among Indonesian students, particularly in the Riau Province. While previous research has underscored the importance of numeracy and highlighted the generally low performance of Indonesian students, this study offers a new perspective by developing and validating a localized assessment instrument through rigorous psychometric procedures, including content validity analysis, Cronbach's Alpha reliability testing, and Confirmatory Factor Analysis (CFA). Furthermore, the research explores how socio-economic disparities, cultural beliefs, curriculum structure, and teacher competencies affect students' ability to acquire and apply numeracy skills in real-world contexts. Unlike many prior studies, which tend to focus solely

on cognitive outcomes, this research integrates educational, social, and cultural dimensions to explain the root causes of numeracy gaps. Thus, it provides both theoretical insight and practical strategies for improving numeracy education in regions with diverse and unequal educational conditions.

2. METHOD

This study is development research, commonly referred to as Research and Development (R&D), which employs the ADDIE model. According to Borg and Gall, development research is a process used to develop and validate existing products or create new ones. Additionally, development research can be utilized to generate new knowledge or address existing problems (Bray et al., 2020).

The ADDIE model consists of several stages: (1) analysis, (2) design, (3) development, (4) implementation, and (5) evaluation. The strength of the ADDIE development model lies in its systematic procedure, where each step is interconnected with the previous steps that have been refined to create an effective new product.

The population of this study comprised all junior high schools (SMP) in the city of Pekanbaru, Indonesia. To ensure representativeness and account for the diversity of educational contexts across the region, the sample was selected using a multiple stage random sampling technique. In the first stage, several schools were randomly chosen from different districts within Pekanbaru. In the subsequent stage, a random selection of students was taken from each selected school to participate in the study. This sampling approach was employed to minimize selection bias and to provide a more accurate reflection of the overall student population in Pekanbaru.

The research instrument used in this study is a numeracy test. Data collection instruments include test validation sheets and questionnaires, which are used to validate the test and the questionnaire being developed. The validity test is conducted with the assistance of several experts who are well-versed in student literacy and numeracy. The validated instrument is then tested in schools to determine its validity and reliability according to the expected standards.

The data analysis technique employed in this study is descriptive analysis, which describes the validity and reliability of the developed test in the form of percentage data. The assessment categories provided by the validators follow a Likert scale. This research is conducted through five main stages or procedures.

a. Analysis

The research procedure begins with an analysis aimed at determining the necessary development of the test instrument. Based on the needs analysis, a diagnostic instrument for students' numeracy skills is required.

b. Design

In the design phase, multiple-choice test questions are developed to measure students' numeracy skills. A set of 40 multiple-choice questions is created. In addition to the test instrument, a non-test instrument in the form of a questionnaire is used to collect data on variables that can enhance students' numeracy skills at the school level.

c. Development

The development stage focuses on creating the numeracy skill instrument. This phase involves several activities such as content and construct validation, reliability testing of the questions, revision of the questions, and finalizing the product.

Content validation is analyzed using Aiken' s formula to measure the consistency of expert assessments. Meanwhile, construct validation is analyzed using Confirmatory Factor Analysis (CFA) to ensure that the instrument accurately measures the intended construct. The instrument validation process involves six experts.

d. Implementation

This stage involves the trial testing of the numeracy skills test. A small-scale testing phase is conducted to assess the difficulty level of the questions and their discriminatory power. The results from the small-scale test are used to refine the questions before conducting a large-scale test.

3. RESULT AND DISCUSSION

Six experts evaluated the content quality of the developed instrument. The evaluation was conducted to assess the instrument's alignment with the indicators and the coverage of all aspects of the measured topics. The content of the instrument must accurately represent the concepts being assessed. Additionally, the instrument should be usable within the available resources, including time, effort, and budget. It must also be easy to use by both researchers and respondents without requiring complex interpretation. The language used should be appropriate for the target respondents, easy to understand, and unambiguous.

Out of the 40 developed questions, four required revisions. The experts noted that these four questions could create difficulties for students due to ambiguous wording and overly long sentences. Once the content validation process was completed, the next step was to revise these questions and conduct a trial test with 100 students. The evaluation results are presented in Table 1.

Table 1. Expert Assessment Result toward Instrument development

Items	sum s	Aiken'Index	Conclusion	Items	sum s	Aiken'Index	Conclusion
1	8	0,667	middle	21	6	0,500	middle

Items	sum s	Aiken'Index	Conclusion	Items	sum s	Aiken'Index	Conclusion
2	6	0,500	middle	22	6	0,500	middle
3	5	0,417	middle	23	9	0,750	middle
4	4	0,333	middle	24	8	0,667	middle
5	7	0,583	middle	25	6	0,500	middle
6	9	0,750	middle	26	7	0,583	middle
7	10	0,833	High	27	9	0,750	middle
8	4	0,333	Low	28	9	0,750	middle
9	10	0,833	High	29	4	0,333	middle
10	9	0,750	middle	30	6	0,500	middle
11	7	0,583	middle	31	5	0,417	middle
12	8	0,667	middle	32	8	0,667	middle
13	7	0,583	middle	33	8	0,667	middle
14	8	0,667	middle	34	10	0,833	High
15	7	0,583	middle	35	9	0,750	middle
16	6	0,500	middle	36	9	0,750	middle
17	10	0,833	High	37	4	0,333	Low
18	3	0,250	Low	38	7	0,583	middle
19	11	0,917	High	39	7	0,583	middle
20	11	0,917	High	40	8	0,667	middle

The Table 1 presents an analysis of the validity of test items based on Aiken’ s Index, a statistical measure used to assess content validity. The index ranges from 0 to 1, with higher values indicating stronger agreement among experts regarding the relevance of the test items. The results indicate that six items (7, 9, 17, 19, 20, and 34) have high validity with an Aiken’ s Index of 0.833 or higher, meaning these items are strongly aligned with the intended numeracy constructs and can be considered reliable for assessment.

The majority of the items (28 out of 40) fall into the moderate validity category, with Aiken’ s Index values ranging from 0.417 to 0.750. While these items are somewhat valid, they may require slight revisions or further validation to enhance their effectiveness in measuring numeracy skills. However, three items (8, 18, and 37) have low validity, with Aiken’ s Index values of 0.333 or lower. These items may not adequately measure the intended numeracy constructs and should be revised or removed from the instrument.

Overall, the test contains a mix of high, moderate, and low-validity items, suggesting that some items need refinement to ensure the overall validity of the instrument. Items with low validity should be carefully examined and modified, while moderate validity items may benefit from improvements in clarity and alignment with learning objectives. Meanwhile, high-validity items can be retained as they effectively measure the intended constructs. Strengthening the weaker items will enhance the instrument’ s overall reliability and effectiveness in evaluating students’ numeracy skills.

a. Small-Scale Trial

The trial was conducted to test the empirical validity of the developed instrument. This trial was carried out in three schools, involving 100 students as samples. The analysis of the trial results includes question difficulty levels, discrimination index, question validity, and instrument reliability. The analysis results can be seen in Tables 2, 3, and 4.

Table 2. Analysis Results of Item Difficulty Levels and Discrimination Index

Items	Difficulty Level	Discrimination Index	Conclusion	Items	Difficulty Level	Discrimination Index	Conclusion
1	0,07	0,12	discarded	21	0,11	-0,59	discarded
2	0,59	0,27	Accepted	22	0,79	0,43	Accepted
3	0,43	0,8	Accepted	23	0,77	0,52	Accepted
4	0,52	0,8	Accepted	24	0,5	0,55	Accepted
5	0,41	0,87	Accepted	25	0,56	0,13	Accepted
6	0,72	0,41	Accepted	26	0,57	0,87	Accepted
7	0,41	0,33	Accepted	27	0,59	0,8	Accepted
8	0,51	0,78	Accepted	28	0,6	0,27	Accepted
9	0,34	0,87	Accepted	29	0,61	0,8	Accepted
10	0,73	0,67	Revision	30	0,31	-0,19	discarded
11	0,59	0,83	Accepted	31	0,49	1	Accepted
12	0,81	0,87	Accepted	32	0,51	0,41	Accepted
13	0,72	0,87	Accepted	33	0,52	0,38	Accepted
14	0,55	0,93	Accepted	34	0,53	0,36	Accepted
15	0,58	0,33	Accepted	35	0,55	0,33	Accepted
16	0,61	0,27	Accepted	36	0,61	0,3	Accepted
17	0,44	0,25	Accepted	37	0,79	0,3	Accepted
18	0,67	0,22	Accepted	38	0,77	0,27	Accepted
19	0,49	0,19	Accepted	39	0,5	0,25	Accepted
20	0,64	0,17	Accepted	40	0,61	0,22	Accepted

Based on the analysis results regarding difficulty levels and discrimination indices, it was found that four items in the instrument needed to be revised or removed. This was due to the fact that the difficulty levels and discrimination indices of these four questions did not meet the expected standards. Furthermore, to test the empirical validity of the instrument, additional analysis was conducted using First Order Confirmatory Factor Analysis (CFA). CFA was used to ensure whether the factors within the instrument aligned with the collected data. The results of this analysis provide a clearer picture of the quality and reliability of the developed instrument. The analysis results can be seen in Table 3.

Table 3. Empiric Validity by CFA

Items	Loading Factor	Conclusion	Items	Loading Factor	Conclusion
1	0,35	Good	21	0,58	Good

Items	Loading Factor	Conclusion	Items	Loading Factor	Conclusion
2	0,46	Good	22	0,63	Good
3	0,51	Good	23	0,43	Good
4	0,55	Good	24	0,41	Good
5	0,67	Good	25	0,51	Good
6	0,49	Good	26	0,54	Good
7	0,24	Discarded	27	0,41	Good
8	0,52	Good	28	0,4	Good
9	0,43	Good	29	0,49	Good
10	0,47	Good	30	0,43	Good
11	0,44	Good	31	0,49	Good
12	0,26	Good	32	0,62	Good
13	0,72	Good	33	0,52	Good
14	0,31	Good	34	0,29	Discarded
15	0,44	Good	35	0,46	Good
16	0,51	Good	36	0,44	Good
17	0,61	Good	37	0,29	Discarded
18	0,41	Good	38	0,51	Good
19	0,22	Discarded	39	0,41	Good
20	0,55	Good	40	0,51	Good

Based on the analysis presented in Table 3, four items need to be replaced to enhance the quality of the ethnomathematics-based instrument. The four items that do not meet the criteria will be revised to ensure that the instrument is more accurate and effective. Subsequently, the revised items will be tested on a larger scale to assess the construct validity and construct reliability of the obtained data. The next step in the analysis process is to determine the reliability of all 40 test items. This process aims to ensure that the developed instrument provides consistent and reliable results. The reliability analysis results can be seen in Table 4.

Table 4. Reliability

Cronbach' s Alpha	N of Items
0.819	40

Table 4 presents the results of the reliability analysis, showing a Cronbach's Alpha value of 0.819. This value indicates that the instrument, consisting of 40 items, has a high level of internal consistency. An instrument is considered reliable if the Cronbach's Alpha value exceeds 0.7; therefore, the 0.819 value suggests that the instrument meets good reliability criteria. Consequently, this instrument can be deemed dependable for further research, as it provides consistent and stable measurement results.

b. Large-Scale Trial

The large-scale trial was conducted to obtain deeper insights into the construct validity and reliability of the collected data. This trial involved 150 students from four schools. By including a

substantial number of participants, the trial aimed to ensure that the instrument used could validly and consistently measure the intended aspects. Construct validity refers to the extent to which the instrument accurately measures the intended concept, while reliability assesses the consistency of the obtained results. The results of this trial analysis are presented in Tables 5 and 6, providing an overview of the quality of the tested instrument.

Table 5. Construct Validity Results Use the Confirmatory Factor Analysis

No	Construct	Loading Factor	Conclusion
1	Whole Numbers	0,51	Valid
2	Plane Shapes	0,43	Valid
3	Statistics	0,59	Valid
4	Number Patterns	0,58	Valid
5	Ratios	0,61	Valid
6	Lines and Angles	0,46	Valid
7	Social Arithmetic	0,63	Valid
8	Probability	0,69	Valid
9	Linear Equations	0,47	Valid
10	Data Presentation	0,45	Valid
11	Algebra	0,68	Valid
12	Sets	0,39	Valid
13	Pythagoras	0,51	Valid
14	Relations and Functions	0,52	Valid
15	Circles	0,67	Valid
16	System of Two-Variable Equations	0,49	Valid
17	Flat-Sided Solid Shapes	0,47	Valid
18	Curved-Sided Solid Shapes	0,43	Valid
19	Transformation	0,55	Valid
20	Similarity and Congruence	0,51	Valid

Based on the data presented in Table 5, all constructs tested during the development of the ethnomathematics-based numeracy instrument have been proven valid. This finding indicates that the instrument is reliable for assessing and evaluating students' numeracy skills in Pekanbaru. Construct validity ensures that each aspect of the instrument is relevant and aligned with its intended measurement objectives. Furthermore, to ensure the instrument's reliability, a construct reliability analysis was conducted. This analysis aims to evaluate the consistency of the instrument in generating stable and trustworthy data. The results of the construct reliability analysis are presented in Table 6.

Table 6. Construct Reliability Analysis

No	Indicators	Loading Factor	Errors	CR	Conclusion
1	Whole Numbers	0,51	0,44	0,83	Reliable
2	Plane Shapes	0,43	0,81		
3	Statistics	0,59	0,55		

No	Indicators	Loading Factor	Errors	CR	Conclusion
4	Number Patterns	0,58	0,54		
5	Ratios	0,61	0,63		
6	Lines and Angles	0,46	0,79		
7	Social Arithmetic	0,63	0,54		
8	Probability	0,69	0,52		
9	Linear Equations	0,47	0,71		
10	Data Presentation	0,45	0,66		
11	Algebra	0,68	0,25		
12	Sets	0,39	0,66		
13	Pythagoras	0,51	0,75		
14	Relations and Functions	0,52	0,45		
15	Circles	0,67	0,56		
16	System of Two-Variable Equations	0,49	0.61		
17	Flat-Sided Solid Shapes	0,47	0,63		
18	Curved-Sided Solid Shapes	0,43	0,62		
19	Transformation	0,55	0,51		
20	Similarity and Congruence	0,51	0,57		

Based on the construct reliability analysis results presented in Table 6, a reliability coefficient of 0.83 was obtained. This value indicates that the instrument, consisting of 20 constructs or indicators, has an acceptable construct reliability coefficient. In other words, the instrument demonstrates good consistency in measuring students' numeracy skills. With this high reliability coefficient, the ethnomathematics-based instrument can be reliably used to assess numeracy skills among junior high school students in Pekanbaru. These results confirm that the instrument is not only valid but also stable and capable of providing consistent measurement outcomes.

c. Evaluation Results

Students' numeracy skills were evaluated in eight schools in Pekanbaru City. The objective of this evaluation was to measure the percentage of students capable of solving numeracy problems across different cognitive levels: application, analysis, evaluation, and creation. This assessment aimed to examine students' ability to apply numeracy concepts in more complex contexts, analyze information, evaluate situations, and develop solutions to numerical problems. A total of 430 students participated in this evaluation, completing all the assigned questions. The results of the analysis of students' numeracy skills in Pekanbaru are presented in Table 7, which illustrates the distribution of students' abilities in solving numeracy problems based on these levels.

Table 7. Evaluation Results by School Based on Cognitive Levels

no	Schools	%C3	%C4	%C5	%C6	%Truth answer	Conclusion
1	School 1	17	8	5	0	30	Not Good
2	School 2	16	13	2	0	26	Not Good
3	School 3	16	8	3	0	31	Not Good

no	Schools	%C3	%C4	%C5	%C6	%Truth answer	Conclusion
4	School 4	18	12	3	0	32	Not Good
5	School 5	13	10	5	1	31	Not Good
6	School 6	21	9	4	1	30	Not Good
7	School 7	15	11	5	2	35	Not Good
8	School 8	23	13	6	2	39	Not Good
Mean		17,37	10,5	4,1	0,75	31,75	

From the analysis results, the average percentage of students who could answer questions at the C3 level (application) was 17.375%. The average percentage of students who could answer questions at the C4 level (analysis) was 10.5%. The average percentage of students who could answer questions at the C5 level (evaluation) was 4.1%. The average percentage of students who could answer questions at the C6 level (creation) was 0.75%. Overall, the percentage of students in Pekanbaru who could answer the questions correctly was 31.75%. These results indicate that students' numeracy skills still require significant improvement. The majority of students struggled with answering questions that required higher cognitive levels, such as analysis, evaluation, and creation. Therefore, greater efforts are needed to enhance students' understanding and numeracy skills to reach higher competency levels.

The analysis shows that the content validity of the numeracy instrument has been proven valid and reliable. The results indicate that this instrument is effective for data collection in the field. Content validity is one of the crucial aspects that must be considered in developing an instrument to assess students' numeracy skills (Shrotryia & Dhanda, 2019). Content validity refers to the extent to which the items in an instrument representatively cover all aspects intended to be measured (Kovacic, 2018). In the context of students' numeracy skills, content validity ensures that the instrument effectively measures the key components of numeracy, such as understanding basic mathematical concepts, problem-solving ability, logical thinking, and the skill to apply mathematical concepts in real-life situations (Wright & Craig, 2011). Content validity is crucial to ensuring the relevance and representation of the assessed material. When developing an instrument for numeracy skills, developers must ensure that each test item reflects the competencies being measured, such as arithmetic operations, geometry, algebra, or statistics (Masuwai et al., 2024). An instrument that lacks good content validity can lead to inaccurate measurements, where important aspects of numeracy are either overlooked or overly emphasized. As a result, the instrument will not provide a valid overall picture of students' numeracy skills.

Furthermore, the importance of content validity in developing numeracy skill instruments is also related to the learning objectives that need to be achieved (Andrian et al., 2018; Sawada et al., 2021). Every educational curriculum has specific competency standards that students must meet, including numeracy skills (Dumas et al., 2021). An instrument with good content validity

will help teachers and educators ensure that the instrument aligns with the learning objectives outlined in the curriculum (Manchaiah et al., 2019). Thus, the test results can accurately reflect the extent to which students have achieved the expected competencies. (Ayre & Scally, 2014). The process of ensuring content validity generally involves a panel of experts, consisting of educators, mathematicians, and psychometric experts, who assess the appropriateness of test items in relation to the intended measurement objectives (Juengst et al., 2019). These experts evaluate whether the test items adequately cover various aspects of numeracy and how relevant each item is in the context of mathematics education. Additionally, item analysis can be conducted to ensure that each question meaningfully contributes to the overall measurement of numeracy skills (Melkamu Asaye et al., 2021). Strong content validity ensures that the instrument can provide accurate and relevant information about students' numeracy skills.

The analysis results also indicate that construct validity falls into the good category. This finding suggests that the accuracy of the instrument in collecting data is accountable. Construct validity refers to the extent to which an instrument truly measures the construct or concept it is intended to assess (Hadi et al., 2019; Lund, 2024; Setiawan et al., 2019). In the context of students' numeracy skills, construct validity ensures that the instrument accurately measures key aspects of numeracy, such as arithmetic skills, understanding of mathematical concepts, and the application of these concepts in real-life situations (Burnett, 2016). Good construct validity indicates that the items in the instrument theoretically and empirically reflect students' numeracy skills (Rezeki et al., 2020). Therefore, ensuring that the instrument has strong construct validity is essential so that measurement results can be used to make accurate decisions in the educational context, such as interventions or learning evaluations. (Istiyono et al., 2020; Wahyuni et al., 2020). With strong construct validity, the instrument will provide a clearer and more accurate representation of students' numeracy skills, ultimately supporting efforts to improve the quality of education.

The reliability analysis results also indicate a good outcome, with a reliability score exceeding 0.8. High reliability ensures that a valid and dependable instrument, both in terms of content and construct, will yield optimal results in data collection (Berger & Karabenick, 2016). A reliable instrument will produce consistent results when administered to the same group of students at different times or under varying conditions (Dawadi & Shrestha, 2018; Newell et al., 2021). This is crucial because reliability indicates that the instrument is stable and not influenced by irrelevant factors, such as students' moods during the test or the environmental conditions of the assessment setting. High reliability is essential to ensure that numeracy test results can be trusted (Nicolella et al., 2013). If the instrument is unreliable, its results will vary and cannot be used to accurately assess students' abilities. (Clifton, 2020). For example, an instrument that produces significantly different results when administered twice to the same students under

nearly identical conditions indicates low reliability (Ismail et al., 2022; Kullan et al., 2022). This can lead to errors in interpreting results and inaccurate conclusions about students' numeracy skills.

The evaluation results indicate that students' numeracy skills in eight schools in Pekanbaru are significantly low. The percentage of students who can correctly answer numeracy questions remains below 40%. These findings suggest that multiple factors need to be addressed to improve students' numeracy skills (Cook, 2018; Støren et al., 2018). One of the main factors contributing to low numeracy skills in Riau Province is the quality of teaching and teacher competency (Oluyemiadetunji & Nel, 2015). Teachers play a crucial role in mathematics education, particularly in developing numeracy skills. However, many teachers in remote areas of Riau may not have adequate access to training and professional development (Chan & Scalise, 2022; Purpura et al., 2015). Teachers who lack mastery of the subject matter or are unable to teach numeracy concepts in an engaging and easily understandable manner will directly impact students' ability to grasp numeracy material. Additionally, conventional and monotonous teaching methods pose an additional challenge (Trickett et al., 2022). Many teachers still rely on traditional lecture-based approaches without actively engaging students in the learning process. This often leads to boredom and a lack of interest in learning mathematics. In reality, to improve numeracy skills, students need to be more involved in activities that encourage exploration, problem-solving, and the application of concepts in real-life situations.

4. CONCLUSION

The instrument used in this study is a diagnostic test designed to evaluate students' numeracy skills. Based on the analysis conducted, the instrument falls into the valid category in terms of both content and construct validity. This indicates that the test accurately measures the aspects intended for evaluation, namely students' numeracy skills. Additionally, the instrument is also reliable in terms of both content and construct, meaning it is consistent and dependable in providing accurate results. Although the difficulty level and discriminatory power of this diagnostic test are generally acceptable, some test items still require revision to improve clarity and effectiveness. The evaluation of students' numeracy skills in Pekanbaru revealed inadequate results, with the majority of students struggling to answer questions related to numeracy skills. This finding highlights the need for significant attention and improvement in students' numeracy proficiency.

ACKNOWLEDGMENT

Thank you to Universitas Islam Riau for funding this research, which has been conducted over the past few months. Hopefully, these findings will be beneficial for improving education in the future, particularly in enhancing students' numeracy skills.

BIBLIOGRAPHY

Andrian, D., Kartowagiran, B., & Hadi, S. (2018). The Instrument Development to Evaluate Local Curriculum in Indonesia. *International Journal of Instruction*, *11*(4), 921 – 934. https://doi.org/10.12973/iji.2018.11458a

Andrian, D., Nofriyandi, N., Indriati, M., & Abdurrahman, A. (2024). Comprehensive Numeracy Evaluation Model of Junior High Schools in Indonesia. *Mosharafa: Jurnal Pendidikan Matematika*, *13*(3), 727-740. https://doi.org/10.31980/mosharafa.v13i3.2035

Aunio, P., Korhonen, J., Ragpot, L., Törmänen, M., Mononen, R., & Henning, E. (2019). Multi-factorial approach to early numeracy—The effects of cognitive skills, language factors and kindergarten attendance on early numeracy performance of South African first graders. *International Journal of Educational Research*, *97*, 65 – 76. https://doi.org/10.1016/j.ijer.2019.06.011

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe' s content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79 – 86. https://doi.org/10.1177/0748175613513808

Berger, J. L., & Karabenick, S. A. (2016). Construct Validity of Self-Reported Metacognitive Learning Strategies. *Educational Assessment*, *21*(1), 19 – 33. https://doi.org/10.1080/10627197.2015.1127751

Bitterly, T. B., VanEpps, E. M., & Schweitzer, M. E. (2022). The predictive power of exponential numeracy. *Journal of Experimental Social Psychology*, *101*. https://doi.org/10.1016/j.jesp.2022.104347

Bray, A., Byrne, P., & O' Kelly, M. (2020). A Short Instrument for Measuring Students' Confidence with 'Key Skills' ' (SICKS): Development, Validation and Initial Results.' *Thinking Skills and Creativity*, *37*(June), 100700. https://doi.org/10.1016/j.tsc.2020.100700

Burnett, C. M. (2016). Reconsidering the Construct Validity of "Political Knowledge." *Critical Review*, *28*(3 – 4), 265 – 286. https://doi.org/10.1080/08913811.2016.1239805

Chan, J. Y. C., & Scalise, N. R. (2022). Numeracy skills mediate the relation between executive function and mathematics achievement in early childhood. *Cognitive Development*, *62*. https://doi.org/10.1016/j.cogdev.2022.101154

Chung, J. H., & Yoo, T. H. (2015). Government policy, network externalities and mobile telecommunication services: evidence from OECD countries. *Applied Economics*, *47*(21), 2173 – 2183. https://doi.org/10.1080/00036846.2015.1005816

Clifton, J. D. W. (2020). Managing validity versus reliability trade-offs in scale-building decisions. *Psychological Methods*, *25*(3), 259 – 270. https://doi.org/10.1037/met0000236

Cook, R. (2018). Gender differences in adult numeracy skills: what is the role of education? *Educational Research and Evaluation*, *24*(6 – 7), 370 – 393. https://doi.org/10.1080/13803611.2018.1540992

Dawadi, S., & Shrestha, P. N. (2018). Construct validity of the Nepalese school leaving english reading test. *Educational Assessment*, *23*(2), 102 – 120. https://doi.org/10.1080/10627197.2018.1430511

Dumas, H. M., Fragala-Pinkham, M. A., Rosen, E. L., & Ni, P. (2021). A content validity evaluation of the PEDI-CAT Speedy Mobility domain. *Physiotherapy Theory and Practice*, *37*(4), 517 – 526. https://doi.org/10.1080/09593985.2019.1633716

Effendi, M., Khoiruli Ummah, S., & Cahyono, H. (2023). Teacher Perspective and Performance in Curriculum Prototype Implementation through the Development of Innovative Project-Based Learning Modules. *Mosharafa: Jurnal Pendidikan Matematika*, *12*(1), 47-58. https://doi.org/10.31980/mosharafa.v12i1.751

Gashaj, V., Thaqi, Q., Mast, F. W., & Roebers, C. M. (2023). Foundations for future math achievement: Early numeracy, home learning environment, and the absence of math anxiety. *Trends in Neuroscience and Education*, *33*. https://doi.org/10.1016/j.tine.2023.100217

Hadi, S., Andrian, D., & Kartowagiran, B. (2019). Evaluation model for evaluating vocational skills programs on local content curriculum in Indonesia: Impact of educational system in Indonesia. *Eurasian Journal of Educational Research*, *2019*(82), 45 – 62. https://doi.org/10.14689/ejer.2019.82.3

Ismail, S. N., Ab Rahman, F., Zain, F. M., & Yaacob, A. (2022). The validity and reliability of quality improvement and accreditation system instrument in managing childcare center. *International Journal of Evaluation and Research in Education*, *11*(3), 1258 – 1267. https://doi.org/10.11591/ijere.v11i3.22522

Istiyono, E., Dwandaru, W. S. B., Setiawan, R., & Megawati, I. (2020). Developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use. *European Journal of Educational Research*, *9*(1), 91 – 101. https://doi.org/10.12973/eu-jer.9.1.91

Jayanti, R., & Cesaria, A. (2024). Pengaruh kemampuan literasi numerasi dan dukungan orang tua terhadap hasil belajar matematika soal cerita di sekolah dasar. *Jurnal Inovasi Pembelajaran Matematika: PowerMathEdu*, *3*(2), 137-148. https://doi.org/10.31980/pme.v3i2.1441

Juengst, S. B., Terhorst, L., Dicianno, B. E., Niemeier, J. P., & Wagner, A. K. (2019). Development and content validity of the behavioral assessment screening tool (BAST β). *Disability and Rehabilitation*, *41*(10), 1200 – 1206. https://doi.org/10.1080/09638288.2017.1423403

Kovacic, D. (2018). Using the Content Validity Index to Determine Content Validity of an Instrument Assessing Health Care Providers' General Knowledge of Human Trafficking. *Journal of Human Trafficking*, *4*(4), 327 – 335. https://doi.org/10.1080/23322705.2017.1364905

Krisma, D. A., Muqtada, M. R., & Khasanah, F. U. (2024). Merdeka Curriculum: How the Profile of Planning for Mathematics Learning by Teachers?. *Mosharafa: Jurnal Pendidikan Matematika*, *13*(1), 175-186. https://doi.org/10.31980/mosharafa.v13i1.1985

Kullan, S., Mansor, M., & Ishak, R. (2022). The validity and reliability of an instrument to evaluate the practices of learning organization. *International Journal of Evaluation and Research in Education*, *11*(4), 1725 – 1733. https://doi.org/10.11591/ijere.v11i4.22974

Lund, T. (2024). The Campbellian construct validity: some comments. *Scandinavian Journal of Educational Research*. https://doi.org/10.1080/00313831.2024.2350642

Manchaiah, V., Granberg, S., Grover, V., Saunders, G. H., & Ann Hall, D. (2019). Content validity and readability of patient-reported questionnaire instruments of hearing disability. *International Journal of Audiology*, *58*(9), 565 – 575. https://doi.org/10.1080/14992027.2019.1602738

Masuwai, A., Zulkifli, H., & Hamzah, M. I. (2024). Evaluation of content validity and face validity of secondary school Islamic education teacher self-assessment instrument. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2308410

Melkamu Asaye, M., Gelaye, K. A., Matebe, Y. H., Lindgren, H., & Erlandsson, K. (2021). Assessment of content validity for a Neonatal Near miss Scale in the context of Ethiopia. *Global Health Action*, *14*(1). https://doi.org/10.1080/16549716.2021.1983121

Mimeau, C., Coleman, M., & Donlan, C. (2016). The role of procedural memory in grammar and numeracy skills. *Journal of Cognitive Psychology*, *28*(8), 899 – 908. https://doi.org/10.1080/20445911.2016.1223082

Newell, S. J., Chur-Hansen, A., & Strelan, P. (2021). A construct validity analysis of the concept of psychological literacy. *Australian Journal of Psychology*, *73*(4), 462 – 474. https://doi.org/10.1080/00049530.2021.1922069

Nicolella, D. P., Torres-Ronda, L., Saylor, K. J., & Schelling, X. (2013). Validity and reliability of an accelerometer- based player tracking device. *Plos*, *62*(4), 198 – 200. https://doi.org/10.7910/DVN/4WHUZH

Nofriyandi, Abdurrahman, & Andrian, D. (2023). Digital Learning Media Integrated with Malay Culture to Improve Students' Numeration Ability and Motivation. *Mosharafa: Jurnal Pendidikan Matematika*, *12*(2), 301-314. https://doi.org/10.31980/mosharafa.v12i2.785

OECD. (2018). *PISA 2018 Worldwide Ranking – average score of mathematics, science and reading*.

Oktaviana, R., Effendi, M., & Athma Putri Rosyadi, A. (2023). Analysis of Mathematics Learning Based on Minimum Competency Assessment in the 2013 Curriculum at SMP Muhammadiyah 1 Malang. *Mosharafa: Jurnal Pendidikan Matematika*, *12*(4), 833-844. https://doi.org/10.31980/mosharafa.v12i4.1195

Oluyemiadetunji, S., & Nel, N. M. (2015). Challenges of Improving Literacy and Numeracy Skills of Prison Inmates in Nigeria. *Journal of Psychology*, *6*(1), 66 – 75. https://doi.org/10.1080/09764224.2015.11885525

Persson, E., Andersson, D., Koppel, L., Västfjäll, D., & Tinghög, G. (2021). A preregistered replication of motivated numeracy. *Cognition*, *214*. https://doi.org/10.1016/j.cognition.2021.104768

Peters, G. W., Kelly, J. R., Beckta, J. M., White, M., Marks, L. B., Ford, E., & Evans, S. B. (2021). An Evaluation of Health Numeracy among Radiation Therapists and Dosimetrists. *Advances in Radiation Oncology*, *6*(3). https://doi.org/10.1016/j.adro.2020.10.022

Prendergast, M., Forster, A., O' Meara, N., O' Sullivan, K., & Faulkner, F. (2023). Numeracy-Meets: an innovative professional development model for adult numeracy practitioners in Ireland. *Irish Educational Studies*. https://doi.org/10.1080/03323315.2023.2209854

Purpura, D. J., Reid, E. E., Eiland, M. D., & Baroody, A. J. (2015). Using a Brief Preschool Early Numeracy Skills Screener to Identify Young Children With Mathematics Difficulties. *School Psychology Review*, *44*(1), 41 – 59. https://doi.org/10.17105/SPR44-1.41-59

Qolbi, A. N., & Afriansyah, E. A. (2024). Capacity for mathematical literacy reviewing the learning style. *Journal of Authentic Research on Mathematics Education (JARME)*, *6*(1), 94-113.

Rezeki, S., Andrian, D., Wahyuni, A., & Nurkholisah, H. (2020). The sustainability concept of Riau cultures through development of mathematics learning devices based on Riau folklore at elementary schools. *Journal of Physics: Conference Series*, *1538*(1). https://doi.org/10.1088/1742-6596/1538/1/012066

Safitra, M. D., Mulyono, B., & Susanti, E. (2023). Numeration-Based Teaching Materials on Algebra Shape Materials for Blended Learning. *Mosharafa: Jurnal Pendidikan Matematika*, *12*(2), 375-388. https://doi.org/10.31980/mosharafa.v12i2.793

Sawada, T., Tomori, K., Sakaue, K., Higashikawa, Y., Ohno, K., Okita, Y., Seike, Y., Fujita, Y., & Umeda, M. (2021). Evaluating the Content Validity of a New On-Road Driving Test. *Physical and Occupational Therapy in Geriatrics*, *39*(3), 295 – 315. https://doi.org/10.1080/02703181.2021.1873476

Segers, E., Kleemans, T., & Verhoeven, L. (2015). Role of Parent Literacy and Numeracy Expectations and Activities in Predicting Early Numeracy Skills. *Mathematical Thinking and Learning*, *17*(2 – 3), 219 – 236. https://doi.org/10.1080/10986065.2015.1016819

Setiawan, A., Mardapi, D., & Andrian, D. (2019). The Development of Instrument for Assessing Students' Affective Domain Using Self-and Peer-Assessment Models. In *International Journal of Instruction* (Vol. 12, Issue 3). https://doi.org/10.29333/iji.2019.12326a

Shrotryia, V. K., & Dhanda, U. (2019). Content Validity of Assessment Instrument for Employee Engagement. *SAGE Open*, *9*(1). https://doi.org/10.1177/2158244018821751

Støren, L. A., Lundetræ, K., & Børing, P. (2018). Country differences in numeracy skills: how do they vary by job characteristics and education levels? *International Journal of Lifelong Education*, *37*(5), 578 – 597. https://doi.org/10.1080/02601370.2018.1554718

Tito, A. S. D. A. P., & Muhtadi, D. (2024). Kemampuan literasi numerasi siswa dalam menyelesaikan masalah perbandingan senilai. *Jurnal Inovasi Pembelajaran Matematika: PowerMathEdu*, *3*(1), 127-136. https://doi.org/10.31980/pme.v3i1.1721

Trickett, J., Batchelor, S., Brittle, B., Foulkes, M., Pickering, J., Slocombe, F., & Gilmore, C. (2022). The role of parent-led and child-led home numeracy activities in early mathematical skills. *Cognitive Development*, *63*. https://doi.org/10.1016/j.cogdev.2022.101189

Val é rio, M., & Ferrara, S. (2022). Numeracy at the dawn of writing: Mesopotamia and beyond. *Historia Mathematica*, *59*, 35 – 53. https://doi.org/10.1016/j.hm.2020.08.002

Wahyuni, A., Effendi, L. A., Angraini, L. M., & Andrian, D. (2020). Developing instrument to increase students' geometry ability based on Van Hiele level integrated with Riau Malay culture.

Jurnal Penelitian dan Evaluasi Pendidikan, *24*(2), 208 – 217. https://doi.org/10.21831/pep.v24i2.33811

Wannenburg, E., & Curlewis, L. (2023). Exploring the need for numeracy skills in legal practice. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186X.2023.2190307

Wright, P. M., & Craig, M. W. (2011). Tool for assessing responsibility-based education (TARE): Instrument development, content validity, and inter-rater reliability. *Measurement in Physical Education and Exercise Science*, *15*(3), 204 – 219. https://doi.org/10.1080/1091367X.2011.590084

Zaki, A., Ramdhani, N., Arwadi, F., & Talib, A. (2024). Numerical Skills Analysis of Mathematics Department Students in the Kampus Mengajar Batch 7 Program. *Mosharafa: Jurnal Pendidikan Matematika*, *13*(2), 551-562. https://doi.org/10.31980/mosharafa.v13i2.2017