Development of Numerical Literacy Question Instrument based on Computational Thinking for Mathematics Learning

by Pustaka UIR

Al-Ishlah: Jurnal Pendidikan
Vol.16, 4 (December, 2024), pp. 4489-4502
ISSN: 2087-9490 EISSN: 2597-940X, DOI: 10.35445/alishlah.v16i4.5985

Development of Numerical Literacy Question Instrument based on Computational Thinking for Mathematics Learning

Zafrullah[1], Syukrul Hamdi[2], Astri Wahyuni[3], Rina Safitri[4], Resky Nuralisa Gunawan[5], Yoel Istiawanto[6]

[1] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; zafrullah.2022@student.uny.ac.id
[2] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; syukrulhamdi@uny.ac.id
[3] Universitas Islam Riau, Pekanbaru, Indonesia; astriwahyuni@edu.uir.ac.id
[4] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; rina0045pasca.2021@student.uny.ac.id
[5] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; reskynuralisa.2022@student.uny.ac.id
[6] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; yoelistiawanto.2023@student.uny.ac.id

ARTICLE INFO

Keywords:

Numeracy Literacy;
Computational Thinking;
Mathematics Learning

Article history:

Received 2024-09-11
Revised 2024-11-06
Accepted 2024-11-07

ABSTRACT

This study is developmental research that aims to develop a numeracy literacy instrument based on computational thinking in mathematics learning, using the Plomp model to ensure the quality and relevance of the instrument. Selection of respondents using Cluster Random Sampling involving 220 grade 8 students at 5 State Junior High Schools in Bantul Yogyakarta, the instrument was tested for content validity using the Aiken method, resulting in an average validity score of 0.81, indicating that the instrument is valid. Confirmatory Factor Analysis (CFA) results showed that all latent variables had loading factor values above 0.5. These items more accurately represent numerical literacy abilities based on the computational thinking being measured. Construct Reliability (CR) and Average Variance Extracted (AVE) analysis show that all dimensions meet the criteria, although the Problem Decomposition dimension requires more attention to its indicators because its AVE value is only slightly above the minimum limit. This study is limited to a relatively small number of respondents, so it is necessary to expand the sample coverage to increase the generalizability of the results. This research develops an effective instrument to improve mathematics learning and prepare students for final exams, such as the Asesmen Standardisasi Pendidikan Daerah (ASPD), with better numeracy literacy skills.

Corresponding Author:
Zafrullah
Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; zafrullah.2022@student.uny.ac.id

1. INTRODUCTION

Education is an important factor in shaping the quality of human resources (Habib et al., 2024; Piwowar-Sulej, 2021). Along with the times, education has also undergone many changes to suit the needs of modern society (Biesta, 2021; Williamson et al., 2020). Curricula and learning methods are constantly updated to be relevant to technological developments and global challenges (González-Salamanca et al., 2020; Schmidt & Tang, 2020). Learning in schools no longer only focuses on theory, but also encourages students to have practical skills and critical thinking (Alsaleh, 2020; Dekker, 2020). Schools serve as the main platform for implementing these educational innovations. One of the innovations implemented to support more effective and interactive learning is the use of collaborative learning spaces, namely classrooms.

The classroom is the primary environment for students to acquire knowledge and skills (Khasawneh et al., 2024; Quadir et al., 2024). In the classroom, attention is focused on the interaction between teachers and students as well as the learning materials presented (Aspbury-Miyanishi, 2024; Wang et al., 2024). Students are invited to be actively involved in the learning process through various methods, ranging from discussions, experiments, to group work (Hanapi & Kamal, 2024; Reilly & Reeves, 2024). The classroom is also a place where students can develop critical thinking and problem-solving skills independently (Islamiati et al., 2024; Kusumawardani & Aminatun, 2024). Many lessons are designed to suit the needs of students, with a variety of approaches that are more creative and collaborative. In addition, regular evaluations are conducted to assess student progress and the effectiveness of the teaching strategies applied. With this approach, students are expected to develop a variety of important skills, including critical thinking and innovative computational thinking.

Computational thinking is a cognitive approach that teaches students how to solve problems in a systematic and structured way, similar to how computers work (Angeli & Giannakos, 2020; Lyon & J. Magana, 2020; Zafrullah et al., 2023, 2024). Computational thinking involves skills such as problem-solving, data analysis, and algorithms, all of which are essential to understanding and solving complex challenges (Palts & Pedaste, 2020; Rodríguez del Rey et al., 2021). In the classroom, computational thinking can be applied through various activities, such as simple programming, technology-based projects, and logic exercises. Thus, students not only learn the subject matter in depth but also develop critical and analytical thinking skills that are useful in a variety of contexts (Polat et al., 2021; Saritepeci, 2020). To ensure that this approach is effective, it is important to periodically evaluate students' understanding and progress in mastering computational thinking skills.

Evaluation is an important part of the educational process to measure the extent to which learning objectives are achieved (H. Lee, 2024; Oermann et al., 2024). Evaluation provides an overview of students' development and understanding of the material that has been taught (Carpenter et al., 2020; Saroyan & Amundsen, 2023; Tomlinson, 2023). Through evaluation, teachers can assess the success of the learning method and identify areas that need improvement (Arustamyan et al., 2020; Suchyadi et al., 2020). In addition, evaluation also serves as feedback for students so that they can understand their strengths and weaknesses in learning (Kasani et al., 2020; Nicol, 2021). Therefore, the planning and implementation of evaluation must be done carefully so that the results are accurate and objective. One important aspect of effective evaluation is the use of appropriate and valid instruments.

An instrument is a measuring tool designed to collect data and information needed in research or evaluation (Ng et al., 2024). The instrument used must be able to measure variables or attributes accurately and consistently (Pandey & Pandey, 2021). A good instrument is one that is proven to have high validity and reliability, so that the measurement results can be trusted and scientifically accepted (Cheung et al., 2024; Sürücü & Maslakci, 2020). To ensure the quality of the instrument, it is important to conduct in-depth testing to verify that the instrument is truly fit for the purpose of the research. To ensure the quality of the instrument, it is important to carry out in-depth verification to verify that the instrument really meets the research objectives. This verification includes the evaluation of various technical and theoretical aspects to assess the suitability of the instrument to the variables being measured. Therefore, one of the most common and important methods in assessing this suitability is proving construct validity.

Construct validity is a measure of how well an instrument measures the theoretical concepts it is intended to measure (Sürücü & Maslakci, 2020; Wilson, 2023). Construct validity involves assessing the extent to which the measuring instrument is in accordance with the underlying theory and how the measured variables are interconnected within the theoretical framework (Compeau et al., 2022; Hilkenmeier et al., 2020). Proof of construct validity is done through various analytical methods, such as

Confirmatory Factor Analysis (CFA), which tests the extent to which the data obtained supports the hypothesized factor structure. Construct validity ensures that the instrument not only measures what it is supposed to measure but is also relevant to the research context (Alavi et al., 2024). Thus, it is important to ensure that the instrument is not only valid but also consistent in providing results namely reliability.

Reliability is a measure of the consistency of an instrument in producing the same results on various measurement occasions (Kennedy, 2022; Loewenthal & Lewis, 2020). Reliability has an important role in ensuring that the instrument can be relied upon to measure variables repeatedly without producing results that vary significantly (Sürücü & Maslakci, 2020). An instrument with good reliability will provide stable and consistent results, regardless of when or who takes the measurement (Fenn et al., 2020; Kalkbrenner, 2023). With a high level of reliability, researchers can be more confident that changes in measurement results truly reflect changes in the variables being measured, not due to instrument inaccuracies. Thus, good reliability is an important requirement in ensuring the quality of research instruments.

The results of previous research show that the development of instruments for assessing student collaboration in Vocational High Schools (SMK) and learning independence instruments for high school students have produced valid and reliable measuring tools (Firdausa & Istiyono, 2019; Mayola et al., 2023). In the first study, a student collaboration assessment instrument was developed through seven steps and involved nine experts, with the results showing that all items were valid and the instrument was tested on 211 student respondents from various expertise programs in Indonesia. Instrument analysis produced a KMO value of 0.859, a Cronbach Alpha reliability value of 0.870, and a total variance explained of 65.553%, and produced seven new indicators. Meanwhile, in the second study, the learning independence instrument for high school students showed good content and construct validity, with reliability exceeding 0.7. However, this study was only conducted once and required retesting to ensure that the instrument was truly standardized. The weakness of these two studies is that the focus of the instruments is more on the affective aspect, not the cognitive aspect, even though the cognitive aspect is very important in assessing students' understanding and thinking abilities in depth. Therefore, further research needs to be carried out to develop instruments that also cover cognitive aspects comprehensively.

From the results of an interview with one of the mathematics teachers at a junior high school in Bantul, Yogyakarta, Indonesia, it is known that the use of numeracy literacy-based questions has not yet become a habit in the classroom, and teachers often only use ordinary exercises taken from the internet or Lembar Kerja Peserta Didik (LKPD) which are not yet integrated with numeracy literacy. In fact, the question of numeracy literacy is part of the Regional Asesmen Standardisasi Pendidikan Daerah (ASPD) which requires an assessment of this skill. This shows the importance of integrating numeracy literacy and computational thinking in mathematics learning. Computational thinking has the advantage of equipping students with systematic and structured thinking skills that are important in solving mathematical problems (Isharyadi & Juandi, 2023; H.-Y. Lee et al., 2024). Numeracy literacy allows students to better understand and interpret data and numbers that are relevant to everyday life (Rakhmawati & Mustadi, 2022). By integrating computational thinking and numeracy literacy in mathematics learning, students are not only better prepared to face assessments such as ASPD, but are also able to develop the critical, analytical, and logical thinking skills that are so needed in the modern world.

Based on the results of initial research, a review of relevant literature, and interviews, researchers developed a cognitive question instrument for mathematics learning. The novelty of this research lies in the application of numeracy literacy questions based on computational thinking, which proved to be very useful for respondents in Bantul, Yogyakarta. This instrument helps students prepare for the final exam the name ASPD, by emphasizing numeracy literacy questions which are the focus of mathematics learning. This research makes a significant contribution by providing measurement tools that suit local needs and supporting more effective exam preparation.

2. METHODS

2.1 Research Model

This research is a development-research that aims to develop numeracy literacy instruments based on computational thinking in mathematics learning. This research uses the Plomp model because this model provides a systematic and structured approach to the process of developing educational instruments. By following this model, this research can ensure that the instrument developed not only fits the learning needs but also meets the quality standards required to improve students' numeracy literacy through the computational thinking approach.

Figure 1. Plomp Development Model by Plomp (2013). Inspiration Flowchart by Thalhah et al. (2022)

At the Initial Investigation stage, the researcher conducted interviews with mathematics teachers to explore the problems in the school to be studied. The results of the interview showed that the teacher only used questions from the internet and Lembar Kerja Peserta Didik (LKPD) that had not integrated numeracy literacy, even though mathematics questions for the Asesmen Standardisasi Pendidikan Daerah (ASPD) should focus on numeracy literacy. In addition, teachers have also never analyzed students' computational thinking skills. Therefore, it is necessary to develop an instrument that integrates numeracy literacy and computational thinking so that students are better prepared for the ASPD and improve the quality of mathematics learning.

In the Design stage, the researcher adjusted the instrument based on the needs identified from the interviews. The researcher developed a numeracy literacy instrument based on computational thinking skills for grade 8 students, with the material "Bentuk Aljabar" as the focus because it is a fundamental topic in computational thinking. The instrument consists of 5 questions divided into 9 sections, complete with scoring guidelines, material details, and computational thinking skills. After the instrument was developed, validation was conducted by 5 experts to ensure the suitability of the content, relevance, and quality of measurement.

In the Realization stage, the validated and revised instruments were distributed to 5 schools in Bantul, Yogyakarta. The schools were selected using the Cluster Random Sampling method to represent relevant population characteristics, such as academic level and geographical conditions. The distribution of this instrument took place from August 10 to September 1, 2024, according to the student learning schedule on algebra material. The selection of this material aims to enable data collection to be carried out effectively and in accordance with the curriculum.

The Testing, Evaluation, and Revision stage is carried out after all student test results are collected. Researchers continued with construct validity testing through Confirmatory Factor Analysis (CFA) using LISREL 8.80. CFA is conducted to ensure that each indicator in the instrument can significantly measure the desired construct. The CFA results show that the instrument has good validity, with a loading factor value above 0.5, and a model that matches the empirical data. After CFA was proven, the researcher analyzed construct reliability and Average Variance Extracted (AVE) to evaluate the internal consistency and quality of the instrument in measuring computational thinking skills and numeracy literacy in students.

2.2 Sample and Data Collection

This research used 220 samples obtained through the Cluster Random Sampling method because this method is effective in representing populations with diverse characteristics, such as academic level and geographical conditions in Bantul, Yogyakarta. By randomly selecting schools within relevant groups, this research can ensure that the data obtained is more representative and can be generalized to a wider population. A sample size of 220 students was also considered appropriate because it was large enough to provide adequate statistical power, allowing for more accurate analysis and reducing potential bias, so that the research results could be relied upon to evaluate the effect of computational thinking-based numeracy literacy instruments in mathematics learning.

Table 1. Respondent Demographics

Demographics aspect	***n* (%)**
Gender:	
Female	118 (53.64%)
Male	102 (46.36%)
City:	
Banguntapan	31 (14.09%)
Kota Bantul	63 (28.64%)
Bambanglipuro	126 (57.27%)
Age:	
12	3 (1.36%)
13	100 (45.45%)
14	67 (30.45%)
15	4 (1.82%)
Don't want to tell	46 (20.91%)

Source: Data by Research (2024)

2.3 Analyzing Data

In content validity, researchers involved five experts in their fields who were analyzed using Aiken. A good Aiken score is one that is above 0.8 (high) (Istiyono, 2020). In this study, Aiken's formula was applied to prove the content validity of the computational thinking-based numeracy literacy instrument.

$$V = \frac{\sum s}{n(c-1)}$$

Description:

$\sum s$: Score from rater

n : Number of raters

c : Highest validity number (3)

The data analysis technique used in this research is confirmatory factor analysis, with LISREL 8.80 software for data analysis. In confirmatory factor analysis, there are latent variables that cannot be measured directly and indicator variables that can be observed and measured directly. The analytical model is built based on the relationship between indicator variables and latent variables using factor loading as the main parameter. For interpretation, an indicator is said to be meaningful in measuring latent variables if it has a factor loading coefficient of not less than 0.5. In addition, reliability analysis was carried out using construct reliability analysis. After conducting construct reliability, researchers conducted an Average Variance Extracted (AVE) analysis to measure how well the construct explains the variance of the indicators used. A goodness of fit model test was carried out using certain model suitability indicators to test the suitability between empirical data and the theoretical model (Tungkunanan, 2020).

Table 2. Goodness of Fit Model Testing Indicators

Goodness of Fit Index	Value
p-value	> 0.05
RMSEA	< 0.08
GFI	≥ 0.9
AGFI	≥ 0.9
Chi-square (x^2)	Expected low ($x^2 < 2$df)

3. FINDINGS AND DISCUSSION

3.1 Initial Investigation Phase

At this stage, the researcher conducted interviews with mathematics teachers to explore the problems that existed in one of the schools studied. The results of the interview showed that the teacher only used problems from the LKPD that did not include numeracy literacy, whereas mathematics problems for the ASPD should focus on numeracy literacy. In addition, teachers also identified that they had never analyzed students' computational thinking skills and felt that students often had difficulty in solving problems that required a critical thinking approach.

From the interview, the teacher suggested that the instrument should include more contextualized problem types that are relevant to students' daily lives and focus more on developing computational thinking skills. This indicates the need to develop an instrument that not only integrates numeracy literacy but also includes problems that can help students overcome their weaknesses in computational thinking. Thus, the instrument developed is expected to ensure students' readiness for the ASPD and improve the overall quality of mathematics learning.

3.2 Design Phase

In the Design Phase, the researcher adjusted the instrument to the needs identified from the interviews. The researcher then developed a numeracy literacy instrument based on computational thinking skills for grade 8 students. Grade 8 was chosen because it is a critical stage where students begin to face more complex mathematical concepts and require a deeper understanding. The material used was "Bentuk Aljabar", because this topic is a basic material that is often introduced at the beginning of a new school year and is very relevant for the development of computational thinking skills. The purpose of selecting this material is to ensure that the instrument can effectively measure students' abilities in a curriculum-appropriate context.

The material on "Bentuk Aljabar" is crucial in the development of computational thinking skills as this topic requires students to solve complex problems through pattern identification, use of algorithms, and generalization of concepts. For Grade 8 students, "Bentuk Aljabar" introduces them to more in-depth abstract math concepts, such as factoring, equations, and polynomial operations. This material demands systematic, analytical, and logical thinking skills, which are at the core of computational thinking. Through algebra-based problems, students are trained to deconstruct problems into simpler parts (decomposition), recognize patterns, develop algorithm-based solution strategies, and draw general conclusions. These skills are essential to prepare students for the ASPD exam and more advanced mathematical challenges.

To adjust to the "Learning Outcomes" or learning outcomes, researchers compiled 5 questions that were broken down into 9 sections of numeracy literacy based on computational thinking skills. Each question is equipped with assessment guidelines, material details, as well as details of computational thinking skills, as listed in Table 3.

Table 3. List of Materials for "Bentuk Aljabar" Problems based on Computational Thinking Skills

Number		Score	Total	Dimensions of CT	Indicators	Material Details
1	a	8.5	20	Problem Decomposition (PD)	Identification	Understand and know the Structure of Algebraic Forms
	b	11.5			Solving problems	
2	a	7	20	Pattern Recognition (PR)	Make a conjecture	Simplifying Polynomial Forms.
	b	13			Determining patterns	
3		20	20	Algorithmic Thinking (AT)	Mention logical steps	Understanding Explanations Using Algebraic Forms
5		20	20		Found a solution	Understanding Multiplication and Division of Mono-Syllable Forms
4	a	3	20	Abstraction and Generalization (AG)	Formulate complex problems into mathematical sentences	Understand how to Transform Equations
	b	3			Using properties	
	c	14			Summing up the solution	

After the instrument is created, the next step is to validate it against 5 experts. This validation aims to ensure that the instrument developed is suitable in terms of content, relevance and measurement quality. Experts will provide evaluation and input regarding each aspect of the instrument, so that revisions or adjustments can be made if necessary. This process is important to ensure that the instrument is suitable for use in further research and can provide accurate results. The biography of each of these experts will be explained in Table 4.

Table 4. Biographies of All Expert

Expert	Position	Area of Expertise	Affiliation	Years of Experience
V1	Associate Professor	Mathematics Education	Department of Mathematics Education, Universitas Islam Riau	12 Years
V2	Lecture	Mathematics Education	Institut Agama Islam Negeri Ponorogo	12 Years
V3	Professor	Mathematics Education	Department of Mathematics Education, Universitas Islam Riau	30 Years
V4	Associate Professor	Educational Research and Evaluation	Department of Mathematics Education, Universitas Islam Riau	8 Years
V5	Lecture	Mathematics Education	Department of Mathematics Education, Universitas Islam Riau	11 Years

The selection of validators was based on experience in areas of expertise and years of expertise.The areas chosen were mathematics education as well as educational research and evaluation, as these two areas are highly relevant to assessing the quality and suitability of the instrument with mathematics learning objectives, as well as ensuring that the instrument is able to measure numeracy literacy based on computational thinking appropriately. Validators with more than 5 years of experience were selected because they have a deep understanding and sufficient practical experience in instrument evaluation and development to provide comprehensive and accurate feedback. The results of the content validity can be seen in Table 5.

Table 5. Results of Content Validity with the Aiken Method

Question	V1	V2	V3	V4	V5	$\sum s$	v	Category
1a	3	2	3	3	2	8	0,8	Valid
1b	2	2	3	3	2	7	0,7	Valid
2a	2	3	2	3	2	7	0,7	Valid
2b	3	2	3	3	2	8	0,8	Valid
3	3	3	2	3	2	8	0,8	Valid
4a	3	2	3	3	2	8	0,8	Valid
4b	3	3	3	3	2	9	0,9	Valid
4c	3	3	3	3	2	9	0,9	Valid
5	3	3	3	3	2	9	0,9	Valid
Mean							**0,81**	**Valid**

The results of the instrument validity analysis showed that all questions had good validity values, with an average value of 0.81. Each question was scored by five validators (V1 to V5) using a 1-3 scale, where a score of 3 indicates a high level of agreement between the question and the indicator being measured. For example, questions 1a and 1b received scores of 0.8 and 0.7, respectively, meaning they were rated as valid by the experts. The validity value is calculated by Aiken's V formula, where a value of 0.7 or more is considered as valid, indicating that these questions are appropriate in measuring the aspects to be tested.

In the table, the questions with the highest validity are questions 4b, 4c, and 5 which have a validity value of 0.9. This indicates that these questions fit the construct being measured, showing a very high level of relevance and accuracy. The mean score of 0.81 indicates that the instrument as a whole is adequate and valid for use in further research, especially in measuring numeracy literacy based on computational thinking skills.

Suggestions from all validators were dominated by improving the sentences in some questions to make them more suitable for the language and level of understanding of grade 8 students. This is important to ensure that the instrument can be easily understood by the target users, so that students can answer the questions without confusion. These adjustments involve the use of simpler and more familiar words for students, as well as the arrangement of sentences that are clearer and less confusing, so that the objectives of measuring numeracy literacy and computational thinking skills can be achieved more effectively.

3.3 Realization Phase

After the expert validation was carried out and the instrument revision was adjusted to the suggestions from the experts, the next step was to enter the Realization Phase. At this stage, the author distributed the instrument to 5 schools in Bantul, Special Region of Yogyakarta, Indonesia. The schools were selected using the Cluster Random Sampling method, which ensures that the randomly selected schools can represent the population with relevant characteristics, such as academic level and geographical conditions. The process of distributing this instrument took place from August 10, 2024 to September 1, 2024 following the students' learning schedule on algebraic form material. This material was chosen because it is relevant to the research objectives and is part of the curriculum at the beginning of the school year. Thus, data collection can be done effectively while measuring students' numeracy literacy ability based on computational thinking.

3.4 Test, Evaluation, and Revision Phase

After collecting all student test results, researchers continued to prove construct validity through Confirmatory Factor Analysis (CFA) analysis using LISREL 8.80. CFA analysis is conducted to ensure that each indicator in the instrument can significantly measure the desired construct. This process is important so that the instruments used are truly appropriate and valid in measuring numeracy literacy based on computational thinking in students.

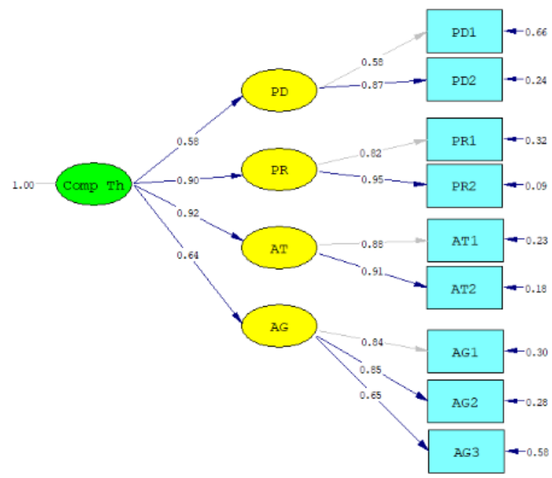

Figure 3. The Results of Confirmatory Factor Analysis about Literacy Numeracy

The results of Confirmatory Factor Analysis (CFA) show that all latent variables, namely PD (Problem Decomposition), PR (Pattern Recognition), AT (Algorithmic Thinking), and AG (Abstraction and Generalization), have significant loading factor values, in which all values are above 0.5. For example, PD1 and PD2 have values of 0.58 and 0.87, indicating that these two indicators greatly contribute to measuring the PD latent variable. Similarly, other latent variables such as PR, which have values of 0.82 and 0.95 on indicators PR1 and PR2. This implies that the instrument used has a good ability to measure each construct individually, thus strengthening the construct validity of the numeracy literacy instrument based on computational thinking skills.

In addition, the strong relationship between the latent variable Computational Thinking and other variables such as PD (0.58), PR (0.90), AT (0.92), and AG (0.64) indicates that these variables can comprehensively explain computational thinking ability. The high loading factor values indicate that each latent construct has a significant role in this theoretical model. Since all values are above the minimum threshold of 0.5, the implication is that the developed model has a good fit with the empirical data, so it can be concluded that this instrument is valid to measure students' computational thinking ability in the context of mathematics learning.

Table 6. Results of Goodness of Fit Model

Criteria	Value	Standard	Explanation
p-value	0.082	> 0.05	Meets the standard (fit)
RMSEA	0.044	< 0.08	Meets the standard (fit)
GFI	0.97	≥ 0.9	Meets the standard (fit)
AGFI	0.94	≥ 0.9	Meets the standard (fit)
Chi-Square	34.18	Expected low ($\chi^2 < 2df$)	Meets the standard ($\chi^2 = 34.18$, df = 23)

Source: LISREL 8.80

The results of the model fit test show that the model used fits well with the empirical data, which can be interpreted as the instrument accurately reflecting the relationship between the variables being measured. With a p-value of 0.082, which is greater than 0.05, as well as a low RMSEA value (0.044), this indicates that the model has a low error rate in estimating parameters, which means that the instrument's measurement results are reliable. The GFI and AGFI values of more than 0.9 each further reinforce the belief that the model fits the empirical data well. The low Chi-Square value also supports the conclusion that the instrument can accurately describe the data.

Thus, these results ensure that the instrument developed can accurately measure students' computational thinking skills and numeracy literacy. The instrument has been shown to capture the interrelationships between variables such as problem decomposition, pattern recognition, algorithmic thinking, and generalization, thus providing a comprehensive picture of students' abilities in a mathematical context. Thus, this instrument can be reliably used in further research or practical applications in the field to evaluate and improve numeracy literacy based on computational thinking skills.

After proving CFA and Goodness of Fit Model, researchers then analyzed Construct Reliability (CR) to ensure the internal consistency of the measured constructs. Furthermore, researchers also evaluate Average Variance Extracted (AVE) to measure how well the construct explains the variance of the indicators used.

Table 7. The Results of Construct Reliability and Average Variance Extracted

Dimensions	Indicators	Loading Factor	Construct Reliability (> 0.6)	Average Variance Extracted (> 0.5)
Problem Decomposition	PD1	0.58	0.69	0.54
	PD2	0.87		
Pattern Recognition	PR1	0.82	0.83	0.78
	PR2	0.95		
Algorithmic Thinking	AT1	0.88	0.86	0.80
	AT2	0.91		
Abstraction and Generalization	AG1	0.84	0.82	0.61
	AG2	0.85		
	AG3	0.65		

Source: Data by Researcher

In the Construct Reliability analysis, it is seen that all dimensions tested show values above the minimum threshold of 0.6. The Problem Decomposition dimension has a Construct Reliability of 0.69, indicating adequate internal consistency even though one of its indicators (PD1) has a slightly low loading factor. The Pattern Recognition dimension has the highest Construct Reliability of 0.83, indicating that its indicators consistently measure the construct. The Algorithmic Thinking dimension also showed good Construct Reliability at 0.86, indicating strong consistency of the measured construct. Abstraction and Generalization also have good Construct Reliability at 0.82, although one indicator (AG3) shows a lower loading factor.

For Average Variance Extracted (AVE), the Pattern Recognition and Algorithmic Thinking dimensions have AVE values that exceed the 0.5 threshold, at 0.78 and 0.80 respectively, indicating that the indicators in both dimensions explain the construct variance significantly. The Abstraction and Generalization dimension also has a qualified AVE value of 0.61, indicating that the construct explains the indicator variance quite well. However, the Problem Decomposition dimension only achieved an AVE value of 0.54, which is slightly above the minimum limit and suggests that some indicators may be sub-optimal in explaining construct variance.

The Problem Decomposition dimension shows a slightly lower Average Variance Extracted (AVE) value of 0.54, which although above the minimum threshold of 0.5, still indicates that some indicators may be less than optimal in explaining construct variance. One reason why this could be the case is because the PD1 indicator has a lower loading factor than the other indicators, which means that PD1's contribution to the overall measurement of the Problem Decomposition dimension is weaker. This could be due to students' difficulty in understanding the concept measured by the indicator.

For previous improvements, this instrument can be repeated by reviewing the indicators in the Problem Decomposition dimension, especially PD1. Considerations that can be made include clarifying or enlightening the question language, adjusting the question context to the student's level of understanding, or adding new indicators that are more representative. This will improve the quality of dimension measurements and is expected to increase the AVE value, so that this instrument is more accurate in measuring students' ability to describe problems. In addition, it is important to involve testing the instrument with a larger sample in order to get more comprehensive feedback regarding students' understanding of the questions given. The use of a more detailed and specific assessment rubric can also help in more objectively assessing each stage in the problem decomposition process, so that the construct validity results in PD1 are better.

The results of the analysis show that most of the dimensions in this study meet the criteria of Construct Reliability and Average Variance Extracted, although the Problem Decomposition dimension needs more attention because its AVE value is the lowest dimension. These findings have important implications for educational research, particularly in the development of computational thinking and its integration in mathematics teaching. By developing a numeracy literacy instrument that focuses on computational thinking skills, this study not only improves students' understanding of mathematical concepts, but also prepares them for ASPD. The integration of computational thinking gives students the tools to deconstruct complex problems and formulate solutions systematically, which in turn can improve the quality of learning and student evaluation results.

The challenge in creating this instrument is to ensure that the questions designed can be well understood by students at different levels of understanding, as well as ensuring the validity and reliability of the instrument in measuring the desired skills. In order to get maximum results in the future, it is necessary to expand the research population coverage and test instruments with different educational contexts or different levels of education. This can help strengthen the generalizability of research results and provide new insights into the effectiveness of instruments in various educational fields, as well as make it possible to adapt instruments to the needs and challenges that exist in each context. So, with deeper development and refinement based on the results of wider trials, it is hoped that this instrument can become a more effective tool in measuring student skills, especially in a learning context that continues to develop.

4. CONCLUSION

This study aimed to develop a numeracy literacy instrument integrated with computational thinking, which successfully achieved the main objectives of improving numeracy literacy and assessing students' computational thinking. The findings suggest that the instrument developed is not only effective in preparing students for the Asesmen Standardisasi Pendidikan Daerah (ASPD), but also enriches the field of educational research with innovative approaches in assessment. However, the limitations of this study include challenges in accurately measuring computational thinking and potential biases that may arise from the sampling method and instrument design. For further improvement, it is recommended that the problem decomposition dimension be refined, the instrument be tested in different educational contexts, and its applicability in other subjects be explored to improve the validity and reliability of the instrument more broadly.

REFERENCES

Alavi, M., Biros, E., & Cleary, M. (2024). Notes to factor analysis techniques for construct validity. *Canadian Journal of Nursing Research*, *56*(2), 164–170.

Alsaleh, N. J. (2020). Teaching Critical Thinking Skills: Literature Review. *Turkish Online Journal of Educational Technology-TOJET*, *19*(1), 21–39.

Angeli, C., & Giannakos, M. (2020). Computational thinking education: Issues and challenges. In *Computers in human behavior* (Vol. 105, p. 106185). Elsevier.

Arustamyan, Y., Siddikova, Y., Sadullaeva, N., Solieva, M., & Khasanova, N. (2020). Assessment of Educational Process and its Organization. *International Journal of Psychosocial Rehabilitation*, *24*(S1), 573–578.

Aspbury-Miyanishi, E. (2024). Guided attention strategies for developing teacher sight. *Teacher Development*, 1–16.

Biesta, G. (2021). *World-centred education: A view for the present*. Routledge.

Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students'(mis) judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition*, *9*(2), 137–151.

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, *41*(2), 745–783.

Compeau, D., Correia, J., & Bennett Thatcher, J. (2022). When Constructs Become Obsolete: A Systematic Approach to Evaluating and Updating Constructs for Information Systems Research. *MIS Quarterly*, *46*(2).

Dekker, T. J. (2020). Teaching critical thinking through engagement with multiplicity. *Thinking Skills and Creativity*, *37*, 100701.

Fenn, J., Tan, C.-S., & George, S. (2020). Development, validation and translation of psychological tests. *BJPsych Advances*, *26*(5), 306–315.

Firdausa, A. R., & Istiyono, E. (2019). Developing instrument for assessing student collaboration in vocational high schools. *Journal of Physics: Conference Series*, *1273*(1), 12043.

González-Salamanca, J. C., Agudelo, O. L., & Salinas, J. (2020). Key competences, education for sustainable development and strategies for the development of 21st century skills. A systematic literature review. *Sustainability*, *12*(24), 10366.

Habib, S., Rahayu, M., Dinta, S. A., Fariq, W. M., Maula, I., & Riadi, H. (2024). Human Resource Management in Improving The Quality of A Productive Educational Environment From Al-Qur'an Perspective; Analysis of Jalalain's Tafsir. *IJIRME: International Journal Of Innovative Research In Multidiciplinary Education*, *3*, 1072–1080.

Hanapi, N. M., & Kamal, M. M. M. (2024). Jigsaw classroom: A process of cooperative learning and discussion. *Semarak International Journal of Innovation in Learning and Education*, *1*(1), 24–31.

Hilkenmeier, F., Bohndick, C., Bohndick, T., & Hilkenmeier, J. (2020). Assessing distinctiveness in multidimensional instruments without access to raw data–a manifest Fornell-Larcker criterion. *Frontiers in Psychology*, *11*, 223.

Isharyadi, R., & Juandi, D. (2023). A systematics literature review of computational thinking in mathematics education: Benefits and challenges. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, *13*(1).

Islamiati, A., Fitria, Y., Sukma, E., Fitria, E., & Oktari, S. T. (2024). The Influence of The Problem Based Learning (PBL) Model and Learning Style on the Thinking Abilities. *Jurnal Penelitian Pendidikan IPA*, *10*(4), 1934–1940.

Istiyono, E. (2020). Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern. *Yogyakarta: UNY Press. L, I.(2019). Evaluasi Dalam Proses Pembelajaran. Jurnal Manajemen Pendidikan Islam*, *9*, 478–492.

Kalkbrenner, M. T. (2023). Alpha, omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*, *14*(1), 77–88.

Kasani, H. A., Mourkani, G. S., Seraji, F., & Abedi, H. (2020). Identifying the weaknesses of formative assessment in the e-learning management system. *Journal of Medical Education*, *19*(2).

Kennedy, I. (2022). Sample size determination in test-retest and Cronbach alpha reliability estimates. *British Journal of Contemporary Education*, *2*(1), 17–29.

Khasawneh, Y., Khasawneh, N., & Khasawneh, M. (2024). Exploring the long-term effects: Retention and transfer of skills in gamified learning environment. *International Journal of Data and Network Science*, *8*(1), 195–200.

Kusumawardani, W., & Aminatun, T. (2024). PBL in Blended Learning Design to Increase Critical Thinking and Problem-Solving Skills. *Jurnal Penelitian Pendidikan IPA*, *10*(6), 3303–3308.

Lee, H.-Y., Wu, T.-T., Lin, C.-J., Wang, W.-S., & Huang, Y.-M. (2024). Integrating Computational thinking into scaffolding learning: An innovative approach to enhance Science, Technology, Engineering, and Mathematics hands-on learning. *Journal of Educational Computing Research*, *62*(2), 431–467.

Lee, H. (2024). The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*, *17*(5), 926–931.

Loewenthal, K. M., & Lewis, C. A. (2020). *An introduction to psychological tests and scales*. Routledge.

Lyon, J. A., & J. Magana, A. (2020). Computational thinking in higher education: A review of the literature. *Computer Applications in Engineering Education*, *28*(5), 1174–1189.

Mayola, E., Ayuni, R. T., & Adhelia, C. (2023). Development of Instruments for Learning Independence for High School Students: Construct Validity and Reliability. *Beginner: Journal of Teaching and Education Management*, *1*(2), 91–103.

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, *55*(3), 1082–1104.

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, *46*(5), 756–778.

Oermann, M. H., Gaberson, K. B., & De Gagne, J. C. (2024). *Evaluation and testing in nursing education*. Springer Publishing Company.

Palts, T., & Pedaste, M. (2020). A model for developing computational thinking skills. *Informatics in Education*, *19*(1), 113–128.

Pandey, P., & Pandey, M. M. (2021). *Research methodology tools and techniques*. Bridge Center.

Piwowar-Sulej, K. (2021). Human resources development as an element of sustainable HRM–with the focus on production engineers. *Journal of Cleaner Production*, *278*, 124008.

Plomp, T. (2013). Educational design research: An introduction. *Educational Design Research*, 11–50.

Polat, E., Hopcan, S., Kucuk, S., & Sisman, B. (2021). A comprehensive assessment of secondary school students' computational thinking skills. *British Journal of Educational Technology*, *52*(5), 1965–1980.

Quadir, B., Yang, J. C., & Wang, W. (2024). Factors influencing the acquisition of English skills in an English learning environment using Rain Classroom. *Interactive Learning Environments*, *32*(1), 1–19.

Rakhmawati, Y., & Mustadi, A. (2022). The circumstances of literacy numeracy skill: Between notion and fact from elementary school students. *Jurnal Prima Edukasia*, *10*(1), 9–18.

Reilly, C., & Reeves, T. C. (2024). Refining active learning design principles through design-based research. *Active Learning in Higher Education*, *25*(1), 81–100.

Rodríguez del Rey, Y. A., Cawanga Cambinda, I. N., Deco, C., Bender, C., Avello-Martínez, R., & Villalba-Condori, K. O. (2021). Developing computational thinking with a module of solved problems. *Computer Applications in Engineering Education*, *29*(3), 506–516.

Saritepeci, M. (2020). Developing computational thinking skills of high school students: Design-based learning activities and programming tasks. *The Asia-Pacific Education Researcher*, *29*(1), 35–54.

Saroyan, A., & Amundsen, C. (2023). *Rethinking teaching in higher education: From a course design workshop to a faculty development framework*. Taylor & Francis.

Schmidt, J. T., & Tang, M. (2020). Digitalization in education: challenges, trends and transformative potential. *Führen Und Managen in Der Digitalen Transformation: Trends, Best Practices Und Herausforderungen*, 287–312.

Suchyadi, Y., Sundari, F. S., Sutisna, E., Sunardi, O., Budiana, S., Sukmanasa, E., & Windiyani, T. (2020). Improving The Ability Of Elementary School Teachers Through The Development Of Competency Based Assessment Instruments In Teacher Working Group, North Bogor City. *Journal Of Community Engagement (JCE)*, *2*(1), 1–5.

Sürücü, L., & Maslakci, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726.

Thalhah, S. Z., Tahrim, T., & Sari, R. E. (2022). Development of lecturer performance assessment instruments through the plomp model in higher education. *AL-ISHLAH: Jurnal Pendidikan*, *14*(2), 1527–1540.

Tomlinson, B. (2023). *Developing materials for language teaching*. Bloomsbury Publishing.

Tungkunanan, P. (2020). Learning Model of Undergraduate Students: Confirmatory Factor Analysis. *International Journal of Instruction*, *13*(3), 665–678.

Wang, Z., Chen, Z., Gong, B., & Feng, Z. (2024). The interactive effects of instructors' guidance frequency and type on Chinese secondary school students' learning. *Education and Information Technologies*, *29*(9), 11257–11280.

Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency. In *Learning, media and technology* (Vol. 45, Issue 2, pp. 107–114). Taylor & Francis.

Wilson, M. (2023). *Constructing measures: An item response modeling approach*. Routledge.

Afrullah, Z., Gunawan, R. N., Haidir, H., & Ramadhani, A. M. (2024). Implementasi penggunaan kemampuan computational thinking pada sekolah anak-anak: Analisis bibliometrik (2014–2024). *Jurnal Binagogik*, 11(2), 201–225.

Zafrullah, Z., Suyanto, S., Wahyuni, A., Ayuni, R. T., & Novilanti, F. R. E. (2023). Development of Android-based Learning to Improve Computational Thinking Skills in Junior High School. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *7*(2), 1309–1320. https://doi.org/10.31004/cendekia.v7i2.2239

Development of Numerical Literacy Question Instrument based on Computational Thinking for Mathematics Learning

ORIGINALITY REPORT

5%	4%	6%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.journal.staihubbulwathan.id Internet Source	2%
2	eprints.uad.ac.id Internet Source	1%
3	files.eric.ed.gov Internet Source	1%
4	Ade Gafar Abdullah, Vina Adriany, Cep Ubad Abdullah. "Borderless Education as a Challenge in the 5.0 Society", CRC Press, 2020 Publication	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		