EDUKASIA: Jurnal Pendidikan dan Pembelajaran
Vol. 7, 1 (January-June, 2026), pp. 83-98
ISSN: 2721-1150; EISSN: 2721-1169; DOI: 10.62775/edukasia.v7i1.1789

Development and Validation of Local Context-Based Literacy and Numeracy Instruments for the Sekolah Sepanjang Hari (SSH) Program

Quin Donspri Tulalessy[1], Achmad Rante Suparman[1], Helena Trivona Turirday[1], Kaleb Amaghi Yenusi[1]

[1] Universitas Papua, Indonesia

ARTICLE INFO

Keywords:

Confirmatory Factor Analysis;
construct validity;
literacy;
numeracy;
sekolah sepanjang hari (SSH)

Article history:

Received 2025-10-09
Revised 2025-12-21
Accepted 2026-02-01

ABSTRACT

This This study aims to develop and validate a basic literacy and numeracy assessment instrument for participants in the Sekolah Sepanjang Hari Program (SSH) in South Sorong Regency, for both regular students and dropouts. The instrument is designed to map Reading, writing, and numeracy skills as the foundation for lifelong learning. Development was carried out through four stages: theoretical construct analysis, expert content validation, field trials, and construct validity analysis using Confirmatory Factor Analysis (CFA). The content validation results showed an average Content Validity Index (CVI) of 0.92, indicating a high fit between the test items and the constructs. A pre-test on 30 students was used to refine the language and local context, while the main pre-test on 145 students ensured proportional levels of readability and difficulty. The CFA results confirmed a three-dimensional structure with perfect model fit (CFI = 1.000; TLI = 1.005; RMSEA = 0.000; SRMR = 0.043; GFI = 0.939). The factor loading values of 0.67–0.92, composite reliability of 0.899–0.918, and AVE of 0.598–0.653 indicate strong validity and reliability. This instrument has been proven feasible and contextually relevant, making it suitable for diagnostic assessment of students' basic abilities in non-formal education, especially in the 3T regions

Corresponding Author:
Quin Donspri Tulalessy
Universitas Papua, Indonesia; q.tulalessy@unipa.ac.id

1. INTRODUCTION

Literacy and numeracy skills are essential for quality education and 21st-century human resource development. They provide students with the foundation for critical thinking, problem-solving, and participation in a knowledge-driven society. Low skills in these areas harm individual learning and prevent national progress on global challenges that require advanced thinking and technology (Geiger et al., 2015).

21st-century education requires students to possess higher-order thinking skills, including critical, creative, communicative, and collaborative thinking (Novia et al., 2024). These skills cannot be achieved without mastering two basic competencies that underpin the entire learning process: literacy and numeracy. Literacy is not merely the ability to read and write, but also the capacity to comprehend, analyze, assess, and apply information in real-life contexts (Cui et al., 2023). Similarly, numeracy is not only related to calculation skills but also encompasses the ability to think logically, analyze patterns, solve problems, and apply mathematical concepts to make informed decisions in everyday life (Geiger et al., 2015). These two competencies form the primary foundation for active participation in a knowledge- and technology-based society, serving as drivers of economic and social development.

Globally, the results of the Programme for International Student Assessment (PISA) conducted by the Organisation for Economic Co-operation and Development (OECD) confirm that literacy and numeracy are key benchmarks of educational quality (McComas, 2014). Countries with high literacy and numeracy achievements, such as Singapore, Finland, and South Korea, demonstrate a positive relationship between these basic competencies and economic progress, innovation, and the quality of life of their citizens. Conversely, low literacy and numeracy skills in many developing countries pose a serious challenge to improving educational quality and global competitiveness (Grotlüschen et al., 2020).

Indonesia still faces disparities in literacy and numeracy achievement between regions. This situation is reinforced by the results of the National Assessment, which show that many students have not achieved the minimum competencies in reading and mathematics (Kementerian Pendidikan, Kebudayaan Riset, 2023). These low achievements not only reflect challenges in basic learning but also imply limited opportunities for developing the quality of human resources in the future. Mastery of literacy and numeracy from an early age has been proven to be a crucial foundation for academic success, productivity, and individual economic well-being later in life. One of the main factors contributing to this disparity in achievement is the disparity in education quality between developed regions and frontier, outermost, and underdeveloped regions (Moon et al., 2015; Thi et al., 2023). This disparity is characterized by limited educational infrastructure, a shortage of professional educators, and limited access to quality teaching materials and learning resources. Efforts to equalize the quality of education and strengthen basic competencies are a strategic agenda to increase competitiveness and educational equity in Indonesia.

The Papua/West Papua Province and its surrounding areas face significant challenges in achieving universal access to basic education. For example, an analysis of Merauke Regency found that poor access to education, including long school distances, limited teacher numbers, and inadequate facilities, are key factors in the low School Participation Rate (APS) in the region (Maharani et al., 2020). Other research indicates that poverty has a significant impact on academic achievement and can even lead students in Papua to drop out of school (Saifullah & Yawan, 2022). Furthermore, a study of two decades of education development in the "Land of Papua" found that the implementation of education policies is often unresponsive to local geographic, demographic, social, and cultural conditions (Afriansyah, 2022). Thus, even though formal basic education is available, local geographic, economic, and socio-cultural barriers continue to create a high risk for children in the region not completing primary school.

As a strategic step to address this issue, the South Sorong Regency government initiated the Sekolah Sepanjang Hari Program (SSH). This program aims to provide broader learning opportunities, especially for children who have not previously completed primary education. SSH is designed with a morning-to-afternoon learning system, integrating academic activities, character building, and basic skills training. This model aligns with the second-chance education approach, recognized by UNESCO as an effective strategy for increasing access to education for children who are out of school. The

primary focus of the SSH program is to rebuild the foundation of literacy and numeracy skills, because these two competencies are prerequisites for successful learning at the next level.

The success of this program depends heavily on the availability of valid and reliable measurement tools that are appropriate for the characteristics of the SSH participants. Instruments used in regular schools are often irrelevant to SSH participants, most of whom have discontinuous educational backgrounds, varying levels of learning motivation, and a heterogeneous age range. To date, basic skills assessments in the 3T (Uninhabited and Remote Areas) regions generally use instruments developed based on the national curriculum without considering the social, cultural, and learning experiences of the students. This has the potential to produce biased assessment results that do not reflect the actual abilities of children in these areas (Suparman et al., 2022). Therefore, the development of literacy and numeracy assessment instruments that are contextual, standardized, and based on construct validity theory is necessary.

This research aims to address this gap by developing literacy and numeracy assessment instruments tailored to the characteristics of All-Day School (SSH) participants in South Sorong Regency. The instrument was developed based on literacy and numeracy theory (Yasukawa et al., 2013), subjected to content validation by experts, and empirically tested using Confirmatory Factor Analysis (CFA) to ensure that the instrument's factor structure aligns with the theoretical model and is capable of measuring constructs with adequate validity and reliability (Suparman et al., 2024).

Education in the 3T (frontier, remote, and underdeveloped) regions, particularly in eastern Indonesia, such as Papua and West Papua, still faces significant challenges in implementing educational assessments relevant to the local context. Difficult geographical access, limited teaching staff, and significant socio-cultural differences require the development of educational measurement tools that can fairly and contextually reflect student abilities (Citrawan et al., 2024). Several studies have demonstrated that assessments based on local wisdom can enhance the validity of measurement results, as they are more tailored to students' learning environments (Dini & Kuswanto, 2025). Furthermore, this approach also supports the principles of inclusive education, which prioritize diversity as the foundation for developing assessment policies and practices (Sumandya et al., 2025). Developing contextual instruments is crucial for promoting equitable education quality, particularly in structurally disadvantaged regions. Therefore, this research aims to design and develop educational measurement tools tailored to the characteristics of the 3T (frontier, remote, and underdeveloped) regions, thereby strengthening the implementation of inclusive and socially just education programs in eastern Indonesia.

Existing efforts to measure literacy and numeracy target formal education, but no tools exist for non-formal learners in the 3T regions. Past research focused on regular school students with similar backgrounds and sufficient resources (Azizah, 2022). This study fills that gap by creating a local assessment tool for Papua, tailored to the context of participants in the SSH Program. The goal is an equitable, accurate instrument for underserved students.

This research aims to develop and validate a context-based literacy and numeracy assessment for participants in the SSH Program in South Sorong Regency. The instrument will offer an accurate picture of students' basic abilities in the 3T region and inform more inclusive, socially just education policies and practices.

2. METHODS

2.1. Type of Research

This research is a research and development (R&D) study aimed at producing a valid and reliable literacy and numeracy assessment instrument for participants in the Sekolah Sepanjang Hari Program

(SSH) in South Sorong Regency. The instrument development model refers to the modified development stages according to Retnawati (2017), Azwar (2013) dan Mardapi (2018), resulting in four main stages of development: theoretical construct analysis, expert content validation, field trials, and construct validity analysis using Confirmatory Factor Analysis (CFA).

Specifically, this research adapted a development model with the following flow:

a. Theoretical construct analysis through a literature review and mapping of literacy and numeracy ability indicators,
b. Content validation by experts to ensure construct representativeness and item suitability,
c. Preliminary and main trials to obtain empirical data on item characteristics, and
d. Measurement model analysis using CFA to ensure construct validity and instrument reliability.

2.2. Research Subjects and Sampling Method

The research subjects were participants in the All-Day School Program (SSH) in South Sorong Regency, most of whom were previously dropouts and now participating in all-day learning activities. The primary trial phase involved 145 students from various educational units that implemented the SSH program.

The sampling technique used was purposive sampling, with the following criteria: (1) students actively participating in the SSH Program for at least one semester, (2) basic reading and writing skills, and (3) willingness to participate in the research. This selection was carried out to ensure that respondents truly represented the characteristics of the SSH program's target audience.

2.3. Research Instrument

The instrument developed consisted of 18 items covering three main dimensions of basic abilities:

a. Reading Ability, consisting of 6 items, measures comprehension of simple texts, identification of information, and interpretation of meaning.
b. Writing Ability, consisting of 6 items, measures the ability to construct simple sentences, complete paragraphs, and express ideas in writing.
c. Calculation Ability, consisting of 6 items, measures basic arithmetic skills, number operations, and the application of numeracy in everyday contexts.

The question format was developed according to the characteristics of the SSH participants, taking into account readability, local context, and a moderate level of difficulty to be suitable for participants who were not previously involved in formal education.

2.4. Instrument Development Procedure

The development process involved the following stages:

a. Theoretical Construct Analysis

   Establishing the dimensions of literacy and numeracy through theoretical studies, national assessment documents, and previous research results. Each dimension was broken down into behavioral indicators, which were then translated into item grids.
b. Content Validation by Experts

   The initial draft of the instrument was consulted with three experts (in basic education, educational assessment, and SSH). Each expert assessed the suitability, clarity, and relevance of the items to the theoretical construct. A Content Validity Index (CVI) was calculated to evaluate the level of agreement among the experts.

c. Preliminary and Main Pilot Tests
   A preliminary pilot test was conducted with 30 students to assess the clarity of the language, the context, and the difficulty of the items. Revisions were made based on feedback. Next, a primary pilot test was conducted with 145 respondents, and the resulting data were used for further psychometric analysis.
d. Construct Validity Using Confirmatory Factor Analysis (CFA)
   A CFA analysis was conducted in RStudio to verify the fit of the theoretical model to the empirical data. The final measurement model was determined based on fit index indicators and factor loadings that met the eligibility criteria.

2.5. Data Analysis Techniques

Data analysis is carried out in two main stages:

a. Content Validity Analysis
   Calculated using the Content Validity Index (CVI) formula based on expert assessments of item suitability and clarity.
b. Construct Validity Analysis
   Construct validity testing was conducted using Confirmatory Factor Analysis (CFA) with the help of R Studio software.

Data analysis was conducted in RStudio using a Confirmatory Factor Analysis (CFA) to test the fit between the theoretical model and the empirical data. The initial stage included checking the CFA assumptions, namely the multivariate normality test and identifying outliers. Next, the measurement model was estimated using the Maximum Likelihood Estimation (MLE) method.

Each item was analyzed to obtain its factor loading, and items with loadings below 0.5 were eliminated because they were deemed not to represent the construct adequately. The model's goodness of fit was evaluated using several key indices, while the interpretation of the results focused on the model's overall empirical suitability.

The next stage included calculating Composite Reliability (CR) and Average Variance Extracted (AVE) to assess internal consistency and convergent validity between items. The analysis results were used to determine the final instrument model and to provide a basis for revising items with insufficient validity, ensuring the instrument was ready for implementation with SSH Program participants.

3. FINDINGS AND DISCUSSION

3.1. Findings

a. Theoretical Construct Analysis

The literacy and numeracy instrument developed in this study comprises 18 items, representing three dimensions of basic skills: reading (6 items), writing (6 items), and arithmetic (6 items). Each item was designed based on indicators adapted from credible international research and assessment results, such as the OECD–PISA, the Ministry of Education and Culture's Center for Education and Culture, and various empirical studies in the field of basic literacy and numeracy. To clarify the conceptual basis and development sources for each item, Table 1 summarizes the 18 indicators used in the three main dimensions, along with their theoretical references and local adaptations in the Sekolah Sepanjang Hari Program (SSH) participants in South Sorong.

Table 1. Basic literacy and numeracy indicators and development reference sources

Number	Capacity Indicator	Reference Source	Adaptation Form SSH Context
1	Recognizing capital and lowercase letters in simple sentences	(Piasta et al., 2022)	Using short texts containing local names and activities
2	Reading simple words frequently used in everyday life	(Strong et al., 2018)	Using vocabulary from the home and school environment
3	Understanding text content and drawing simple inferences	(Pusmendik Kemdikbud, 2022)	Texts about the human body and the natural environment
4	Interpreting information from simple pictures, tables, or graphs	(Pusmendik Kemdikbud, 2022)	Tables containing data on local students' healthy lifestyle habits
5	Determining the main idea and relevant information in informative texts	(Christianti et al., 2022)	Paragraphs about the daily lives of SSH students
6	Answering factual questions based on short reading texts	(Ajmal & Hussain, 2022)	Contextual reading with simple vocabulary
7	Writing words with correct spelling	(Harris et al., 2017; Tavşanlı & Kara, 2021)	Object words are taken from surrounding objects (fruit, animals, classroom objects).
8	Writing simple sentences with complete structure	(Datchuk & Dembek, 2018; Lembke et al., 2003)	Using pictures of children's daily activities.
9	Using punctuation correctly in sentences	(Sülükçü & Kırboga, 2020)	Contextual sentences based on household activities.
10	Writing paragraphs with one clear main idea	(Dockrell et al., 2019)	Paragraph themes about hobbies or personal experiences.
11	Constructing sentences based on specific pictures or situations	(Ritchey, 2008)	Visualizing activities at school and the surrounding environment.
12	Correcting spelling or punctuation errors in sentences	(Tavşanlı & Kara, 2021)	Sentences adapted from local activity examples
13	Determining the number of objects from a set of concrete objects (whole numbers ≤ 99)	(Blanton & Kaput, 2004; Mahpop & Sivasubramaniam, 2010; Starkey & McCandliss, 2014)	Pictures of real objects (fruit, school supplies).
14	Adding 1–3 digit numbers	(Mahpop & Sivasubramaniam, 2010; Wildmon et al., 1999)	Context of the child's activity (sharing candy, counting items).

15	Subtracting 1–3 digit numbers from a set of concrete objects	(Ellemor-Collins et al., 2007; Mauhibah & Karso, 2020)	Everyday situations, such as buying and selling, or playing games.
16	Predicting number patterns from a set of objects or sequences of numbers	(Liljedahl, 2004; Wilkins et al., 2021)	Number patterns are associated with children's routine activities.
17	Dividing 1-3 digit numbers	(Huber et al., 2013; Mulligan & Mitchelmore, 1997)	Contexts involving dividing objects or food.
18	Multiplying 1-3 digit numbers	(Ma et al., 2020)	Contextual situations, such as grouping crops or merchandise, are relevant.

b. Content Validity

Content validity was determined through an expert judgment process aimed at ensuring that each item accurately represented the theoretical constructs of basic literacy and numeracy that were intended to be measured. The validation process was conducted by experts with expertise in basic education, learning assessment, and psychometrics, with between 10 and 20 years of professional experience in educational instrument development.

Each expert assessed the 18 test items based on four main criteria:
(1) Indicator alignment with the competency construct,
(2) Clarity of language and local context,
(3) Appropriateness of difficulty level, and
(4) Content representation of the core literacy and numeracy domains.

The Content Validity Index (CVI) values in Table 2 show that all aspects scored above 0.90, indicating high agreement among experts that the instrument's items have adequate content representativeness and relevance to their theoretical constructs (Istiyono, 2018).

Table 2. Results of the content validity assessment by experts

Number	Aspects Assessed	CVI	Category
1	Conformity to competency indicators	0.94	Very Good
2	Clarity of language and local context	0.90	Very Good
3	Appropriateness of difficulty level	0.91	Very Good
4	Average overall CVI	0.92	Very Good

c. Initial and Main Trials

A preliminary trial was conducted with 30 elementary school students participating in the Sekolah Sepanjang Hari Program (SSH) in a regency in Papua. This trial aimed to identify the clarity of language, suitability for the local context, and the difficulty level of the items in the basic literacy and numeracy instruments developed.

Generally, the pilot test results indicated that most students effectively understood the content and instructions. However, several items needed revision because the terminology and context were not entirely appropriate for the socio-cultural environment of Papuan students.

In the literacy component, several words and phrases were deemed unfamiliar, such as terms that were too common in urban contexts (e.g., "bus terminal," "shopping center"), so they were replaced with contexts more relevant to students' daily lives, such as "traditional market," "beach," or "boat." Meanwhile, in the numeracy component, several items that used specific units of measurement (e.g., meters, liters) were revised to be more contextualized to children's daily activities, such as "measuring the length of wood to build a small boat" or "dividing the fish catch".

In terms of language, several sentences were also found to be too long and complex for elementary school students, especially those who speak local languages as their first language. Therefore, the question wording was simplified and visual illustrations were added to facilitate understanding.

Overall, the results of the initial pilot test indicated that the instrument adequately measured basic literacy and numeracy. However, revisions were needed to strengthen readability, local cultural context, and the appropriateness of the difficulty level to the characteristics of students in Papua. Revisions were made based on the results of the limited pilot test and input from accompanying teachers and experts in basic education and local culture.

Following the revisions, a primary pilot test was conducted with 145 elementary school students from several schools that implemented the SSH Program in Papua. The primary pilot test aimed to obtain broader empirical data on understandability, readability, the distribution of difficulty levels, and the representation of local context in the instrument items.

d. Confirmatory Factor Analysis (CFA) Results

Confirmatory Factor Analysis (CFA) was conducted to test the fit of the three-dimensional construct model, reading, writing, and arithmetic, with empirical data obtained from the instrument pilot test on 145 participants in the All-Day School Program (SSH) in South Sorong Regency. This analysis aimed to verify whether the theoretical structure developed based on the literature review and previous content validity was empirically confirmed.

The estimation process was conducted using the Maximum Likelihood Estimation (MLE) method using R Studio software. The theoretical model tested consisted of three main latent factors: Reading, Writing, and Arithmetic, each measured by six observed indicators. The results of the fit index are presented in Table 3. Each indicator is assumed to represent a specific dimension of basic literacy and numeracy skills.

Table 3. Model fit index results

Number	Fitness Index	Result Value	Criteria	Description
1	SRMR	0.043	< 0.10 (Dagnall et al., 2018)	Fit
2	CFI	1.000	≥ 0.87 (Dagnall et al., 2018)	Fit
3	TLI	1.005	≥ 0.90 (Marsh et al., 1988)	Fit
4	RMSEA	0.000	≤ 0.08 (Cornick, 2015)	Fit
5	GFI	0.939	≥ 0.90 (Kwahk & Lee, 2008)	Fit

From a local perspective, the model's fit is significant. Given that SSH Program participants are primarily students returning to formal education after a period of non-schooling, the successful fit of the three-factor model indicates that the instrument successfully captures the underlying cognitive structure relevant to the non-formal population. The CFA results in Figure 1 show that all measurement paths (path loadings) are significant at the $p < 0.001$ level, indicating that each indicator makes a substantial contribution to the latent construct it represents.

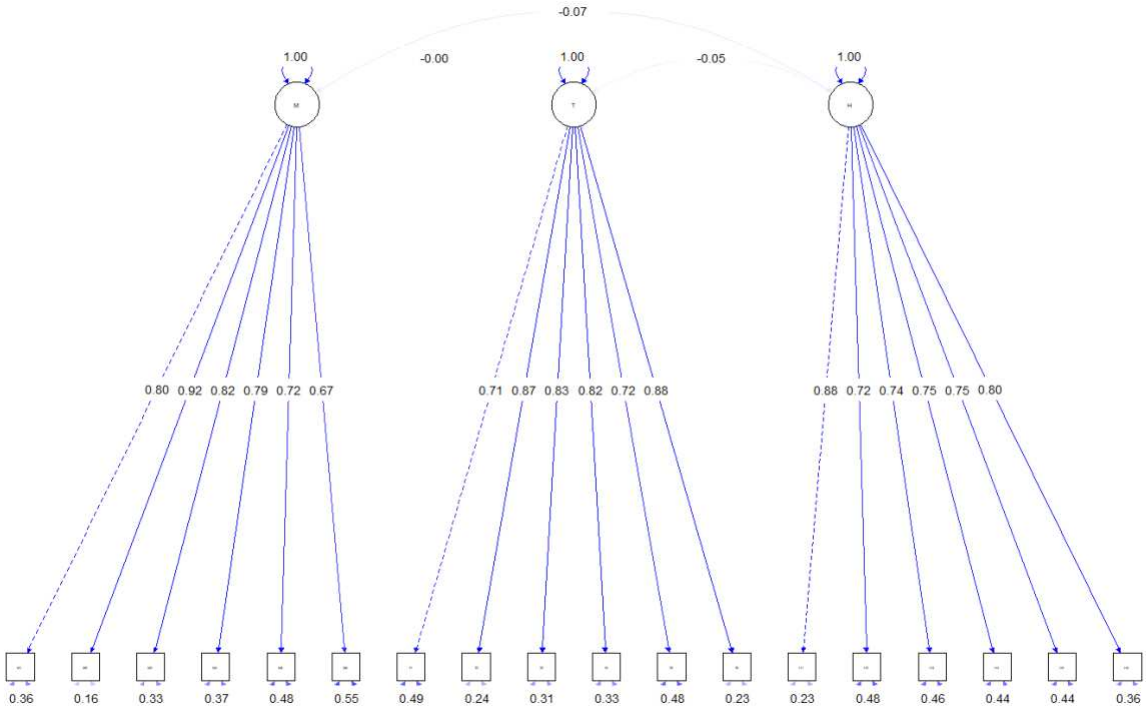

Figure 1. Final CFA model of the SSH program literacy and numeracy instrument (source: RStudio)

The standardized loading factor estimation results in Table 4 indicate that all indicators make significant contributions to their respective constructs. Loading values range from 0.67 to 0.92, suggesting that each item has a strong representation of the dimension it measures.

Table 4. Standardized loading factor values for each dimension

Dimension	Item Code	Loading Factor	Description
Reading	M1–M6	0.67–0.92	Valid
Writing	T1–T6	0.71–0.88	Valid
Calculating	H1–H6	0.72–0.88	Valid

All indicators were declared construct valid because they had loading values above 0.60 (Suparman et al., 2024). High loading values indicate that each item consistently explains the variability of its construct, thereby strengthening the previous CFA results, which indicated good model fit.

To test the internal consistency and reliability of the constructs, Composite Reliability (CR) and Average Variance Extracted (AVE) values were calculated for each dimension, as shown in Table 5.

Table 5. Composite Reliability (CR) and Average Variance Extracted (AVE) Results

Dimension	CR	AVE	Description
Reading	0.908	0.624	Reliable
Writing	0.918	0.653	Reliable
Calculating	0.899	0.598	Reliable

A CR value above 0.70 indicates that each dimension has high internal consistency, where the items within a construct provide relatively stable and homogeneous measurement results. An AVE value above 0.50 indicates that most of the indicator's variance can be explained by the latent construct

it represents (Ghozali & Fuad, 2008). Thus, both convergent validity and composite reliability of this instrument have been met.

3.1. Discussion

Each indicator in Table 1 above represents a theoretical construct of basic literacy and numeracy that can be directly measured through context-based performative tasks. These three dimensions are conceptually complementary and reflect domains of ability oriented toward understanding and application (applied literacy and numeracy).

All indicator items are then structured into test items tailored to the socio-cultural context of participants in the All-Day School Program (SSH) in South Sorong Regency. This program is designed for children who have previously dropped out of school and are now re-engaging in the learning process through adaptive, non-formal channels.

A pilot test of 145 SSH participants showed that over 90% of respondents were able to understand and answer each item effectively, indicating a high level of readability and contextual relevance. This finding suggests that adapting to the local context successfully enhanced stimulus clarity and participant engagement during the assessment. This aligns with the validity principle, which states that the congruence between item content and participants' actual experiences is a crucial element of measurement fairness (Suparman et al., 2022).

Field observations also confirmed that the locally context-based instrument design encouraged participants to be more active, confident, and motivated in answering questions. Therefore, this instrument met face validity and initial content validity (content representativeness) and was suitable for further testing in the Confirmatory Factor Analysis (CFA) stage to ensure the construct validity of the developed three-dimensional model.

The results in Table 2 reinforce the findings from the previous readability test phase, indicating that the test items were structured contextually and communicatively, in accordance with the characteristics of the SSH Program participants in South Sorong Regency. Scores for clarity of language and local context indicate that the experts considered the language used to be simple, familiar, and unambiguous. This is crucial because the majority of SSH participants are students with limited formal learning experience, making the meaningfulness of the context a key determinant of assessment success.

The indicator suitability aspect (0.94) also indicates that all items represent the targeted competencies in each dimension (reading, writing, and arithmetic). This confirms that the development of theory-based indicators, as summarized in Table 1, was successfully translated operationally into measurable items. Thus, there is strong consistency between the theoretical framework and the empirical implementation of the instrument.

Furthermore, the appropriateness score for difficulty (0.91) indicates that experts assessed the cognitive and linguistic level of each item as commensurate with the abilities of SSH participants. This indicates that the instrument has an appropriate challenge level, where items are neither too easy nor too difficult, allowing for optimal ability discrimination.

Overall, the CVI value of 0.92 indicates that this instrument has high content validity, is theoretically grounded, and is sensitive to local context. These results support the importance of assessment systems that generate diverse forms of data, which must be combined in different constellations for specific purposes (Gordon & Campbell, 2014). Assessments rooted in the socio-cultural realities of participants can provide a more authentic and fair picture of ability.

Thus, the instrument developed is considered worthy of proceeding to the construct validity testing stage using Confirmatory Factor Analysis (CFA), to ensure the three-dimensional latent structure underlying the basic literacy and numeracy skills of the SSH Program participants.

The pilot test was conducted successfully, and all students completed the instrument within the allotted time. Based on observations, students appeared more enthusiastic about working on questions related to daily activities in their environment, such as reading a story about a "child helping his parents in the garden" or calculating simple harvest yields. This suggests that integrating elements of local wisdom helped increase student engagement and motivation in the test.

The descriptive analysis results showed that the average item score fell within the moderate category, indicating that the level of question difficulty was proportionate to the overall test difficulty. No more confusing or overly complex items were found. The accompanying teachers also responded positively to the format and content of the instrument, considering it relevant to the Papuan cultural context and supporting the learning objectives of the SSH Program.

Thus, the primary pilot test confirmed that the revisions from the initial stage were effective in improving item quality, both in terms of language clarity, cultural contextual appropriateness, and distribution of difficulty levels. The instrument was then deemed suitable for use in the construct validity analysis stage, utilizing Confirmatory Factor Analysis (CFA).

Statistically, the results in Table 3 indicate that the developed three-dimensional conceptual structure of literacy, reading, writing, and arithmetic is empirically acceptable. The indicators within each dimension are highly correlated and form a latent construct consistent with the underlying theory. These CFA findings reinforce the content validity results, which previously demonstrated the items' fit with the competency construct. The empirical results indicate that the developed indicators are not only deemed relevant by experts but are also statistically confirmed as part of a consistent latent structure.

Pedagogically, these results suggest that enhancing performance in one area, such as reading, can improve performance in other areas, including writing or arithmetic. This pattern aligns with research by Hanemann (2015), which emphasized the link between literacy, numeracy, and critical thinking skills as a foundation for lifelong learning.

Figure 1 reinforces previous findings that the developed three-factor structure is not only conceptually appropriate but also empirically confirmed in the participant population of the Sekolah Sepanjang Hari Program (SSH). Furthermore, the relationships between the latent constructs indicate that the correlations between the latent factors are relatively low, at –0.00 between Factor 1 and Factor 2, –0.07 between Factor 1 and Factor 3, and –0.05 between Factor 2 and Factor 3, respectively. These correlation values approaching zero indicate that the three constructs are nearly independent of each other, so that each factor represents a distinct and unique dimension of ability or characteristic.

Conceptually, each construct developed in the model measures a distinct aspect of the ability being assessed, with minimal overlap. Thus, the measurement model demonstrated good discriminant validity, as each latent construct could be empirically distinguished from the others.

Overall, the CFA results indicate that the SSH Program literacy and numeracy instrument has strong construct validity, high theoretical consistency, and good empirical fit, making it suitable for use as a diagnostic measurement tool in the context of non-formal education in Indonesia.

Statistically, high Composite Reliability (CR) and Average Variance Extracted (AVE) values strengthen evidence that the relationship between the indicators and their constructs is not coincidental but consistent with the conceptual framework of basic literacy-numeracy. This finding aligns with the results of the Confirmatory Factor Analysis (CFA), which demonstrated good model fit and strong interfactor correlations.

Theoretically, these results support the view that literacy and numeracy are two integrated yet distinctive core competency domains. Within the framework of numeracy literacy theory, literacy is understood as the ability to understand, use, and reflect on written information to participate effectively in society. In contrast, numeracy encompasses the ability to think and reason using

mathematical concepts in real-life contexts. High factor loadings, CR, and AVE values demonstrate that the reading, writing, and numeracy indicators adequately represent the construct of literacy-numeracy as defined in the theory.

When compared with previous studies, these results are consistent with those of the OECD in the 2018 PISA (Programme for International Student Assessment) test, which found that social and cultural contexts influence students' understanding of literacy and numeracy (OECD, 2019). Naumann's study on ecological validity showed that measurements that take participants' actual contexts into account are more valid (Naumann et al., 2022). This similarity suggests that a contextual approach based on local culture, as implemented in the development of the SSH Program instrument, is highly relevant and practical.

The difference is that this study emphasizes ecological validity the extent to which the measurement tool reflects students' actual conditions by adapting to the local Papuan context, including language, family activities, and social environment. Empirical results indicate that participants were more enthusiastic and better able to understand questions about their lives, such as activities in the garden, market, or sea. This finding strengthens the theory that assessments rooted in participants' sociocultural realities can produce more fair and meaningful measurements.

In practical terms, the success of this instrument demonstrates that contextualized literacy and numeracy assessments can serve two purposes: (1) as an initial diagnostic tool to map the basic abilities of SSH participants, and (2) as a means of monitoring learning progress and evaluating program effectiveness. Thus, this instrument is not only psychometrically valid and reliable, but also pedagogically and socially relevant, as it supports adaptive and differentiated learning practices in non-formal education contexts.

Overall, the results of this study confirm that literacy and numeracy assessments developed on the basis of local theory and context produce scientifically sound, relevant, and equitable measurements. These findings enrich the literature on measuring basic abilities in regions with unique sociocultural characteristics and open the door to implementing similar models in other areas of Indonesia.

4. CONCLUSION

This study aimed to develop and validate a contextualized basic literacy and numeracy assessment instrument for participants in the Sekolah Sepanjang Hari Program (SSH) in South Sorong Regency. This objective was achieved through testing the instrument's content validity, construct validity, and reliability, which demonstrated excellent results. The resulting three-dimensional model of reading, writing, and numeracy proved theoretically appropriate and empirically confirmed through Confirmatory Factor Analysis (CFA), with model fit indices meeting ideal criteria. These results demonstrate that the literacy and numeracy skills of SSH participants can be measured consistently, despite their diverse social and educational backgrounds and learning experiences.

This study found that a context-based assessment approach improved validity, measurement fairness, and student engagement in the assessment process. The developed instrument serves not only as an academic measurement tool but also as a diagnostic tool that helps educators map participants' basic abilities and design adaptive learning interventions tailored to their needs. Conceptually, this study reinforces the view that literacy and numeracy are interconnected, multidimensional competencies that serve as the foundation for lifelong learning.

Acknowledgments: The authors would like to thank the Regent of South Sorong Regency, the facilitators of the Sekolah Sepanjang Hari Program (SSH), and the experts who contributed to this research process. Administrative and technical support from the school and local community significantly contributed to the success of this research.

Conflicts of Interest: The authors declare that they have no conflicts of interest in the conduct or reporting of this research.

REFERENCES

Afriansyah, A. (2022). Refleksi dua puluh tahun pembangunan pendidikan di tanah Papua (2001-2021). *Masyarakat Indonesia*, *48*(1), 63–77. https://doi.org/10.14203/jmi.v48i1.1170

Ajmal, M., & Hussain, J. (2022). Comparative Analysis of Literacy Skills, Writing and Numeracy Attained by the Students of Formal and Non-Formal Schools at Primary Level in Islamabad Capital Territory. *Bulletin of Education & Research*, *44*(1), 23–38.

Azizah, S. (2022). The 21st century skillsthrough literacy and numeracyat public junior high schools in pamekasan madura. *Proceedings of the 6thInternational Conference on Islamic Studies (ICONIS)*, 13–18. https://doi.org/10.19105/iconis.v6i.590

Azwar, S. (2013). *Reliabilitas dan validitas* (Keenam). Pustaka Pelajar.

Blanton, M. L., & Kaput, J. J. (2004). Elementary Grades Students' Capacity for Functional Thinking. *Proceedings of the 28th Conference of the International Group for the Pyschology of Mathematics Education*, *2*, 135–142.

Christianti, M., Retnowati, T. H., Wening, S., Hasan, A., & Ratnawati, H. (2022). Early Literacy Assessment among Kindergarten Teachers in Indonesia: A Phenomenological Study. *European Journal of Educational Research*, *11*(4), 2401–2411. https://doi.org/10.12973/eu-jer.11.4.2401

Citrawan, I. W., Widana, I. W., Sumandya, I. W., Widana, I. N. S., Mukminin, A., Arief, H., Razak, R. A., Hadiana, D., & Meter, W. (2024). Special education teachers' ability in literacy and numeracy assessments based on local wisdom. *Mathematics and Science Education*, *8*(1). https://doi.org/10.22437/jiituj.v8i1.32608

Cornick, J. E. (2015). Factor structure of the exercise self-efficacy scale. *Measurement in Physical Education and Exercise Science*, *19*(4), 208–218. https://doi.org/10.1080/1091367X.2015.1074579

Cui, Y., Chen, F., Lutsyk, A., Leighton, J. P., & Cutumisu, M. (2023). Data literacy assessments: a systematic literature review. *Assessment in Education: Principles, Policy \& Practice*, *30*(1), 76–96. https://doi.org/10.1080/0969594X.2023.2182737

Dagnall, N., Denovan, A., Parker, A., Drinkwater, K., & Stephen Walsh, R. (2018). Confirmatory factor analysis of the inventory of personality organization-reality testing subscale. *Frontiers in Psychology*, *9*(Article 1116), 1–12. https://doi.org/10.3389/fpsyg.2018.01116

Datchuk, S. M., & Dembek, G. A. (2018). Adapting a Sentence Intervention with Spelling and Handwriting Support for Elementary Students with Writing Difficulties: A Preliminary Investigation. *Insights into Learning Disabilities*, *15*(1), 7–27.

Dini, N. A. I., & Kuswanto, H. (2025). Integrating local wisdom: Innovative assessment instrument of critical thinking skills in science learning. *Jurnal Eduscience (JES)*, *12*(3), 621–630. https://doi.org/10.36987/jes.v12i3.6849

Dockrell, J. E., Connelly, V., & Arfè, B. (2019). Struggling writers in elementary school: Capturing drivers of performance. *Learning and Instruction*, *60*, 75–84. https://doi.org/10.1016/j.learninstruc.2018.11.009

Ellemor-Collins, D., Wright, R., & Lewis, G. (2007). Documenting the Knowledge of Low-Attaining Third- and Fourth-Graders: Robyn's and Bel's Sequential Structure and Multidigit Addition and Subtraction. *Proceedings of the 30th Annual Conference of the Mathematics Education Research Group of Australasia, Hobart*, *1*(1), 265–274. https://doi.org/10.7326/M18-0336

Geiger, V., Goos, M., & Forgasz, H. (2015). A rich interpretation of numeracy for the 21st century: a survey of the state of the field. *ZDM Mathematics Education*, *47*(4), 531–548. https://doi.org/10.1007/s11858-015-0708-1

Ghozali, I., & Fuad. (2008). *Structural equation modeling: teori, konsep, dan aplikasi dengan program lisrel 8.80*. Badan Penerbit Universitas Diponegoro.

Gordon, E. W., & Campbell, E. B. (2014). Context and perspective: Implications for assessment in

education. *Teachers College Record*, *116*(11), 1–18. https://doi.org/10.1177/016146811411601103

Grotlüschen, A., Desjardins, R., & Liu, H. (2020). Literacy and numeracy: Global and comparative perspectives. *International Review of Education*, *66*(2), 127–137. https://doi.org/10.1007/s11159-020-09854-x

Hanemann, U. (2015). Lifelong literacy: Some trends and issues in conceptualising and operationalising literacy from a lifelong learning perspective. *International Review of Education*, *61*(3), 295–326. https://doi.org/10.1007/s11159-015-9490-0

Harris, K. R., Graham, S., Aitken, A. A., Barkel, A., Houston, J., & Ray, A. (2017). Teaching Spelling, Writing, and Reading for Writing: Powerful Evidence-Based Practices. *Teaching Exceptional Children*, *49*(4), 262–272. https://doi.org/10.1177/0040059917697250

Huber, S., Fischer, U., Moeller, K., & Nuerk, H.-C. (2013). On the interrelation of multiplication and division in secondary school children. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00740

Istiyono, E. (2018). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika* (Pertama). UNY Press.

Kementerian Pendidikan, Kebudayaan Riset, dan T. (2023). *Laporan hasil asesmen nasional 2023: Literasi membaca dan numerasi*. Badan Standar, Kurikulum, dan Asesmen Pendidikan (BSKAP), Kemendikbudristek. https://pusmendik.kemdikbud.go.id

Kwahk, K. Y., & Lee, J. N. (2008). The role of readiness for change in ERP implementation: Theoretical bases and empirical validation. *Information and Management*, *45*(7), 474–481. https://doi.org/10.1016/j.im.2008.07.002

Lembke, E., Deno, S. L., & Hall, K. (2003). Identifying an Indicator of Growth in Early Writing Proficiency for Elementary School Students. *Assessment for Effective Intervention*, *28*(4), 23–35. https://doi.org/10.1177/073724770302800304

Liljedahl, P. (2004). Repeating pattern or number pattern: The distinction is blurred. *Focus on Learning Problems in Mathematics*, *26*(3), 24–42.

Ma, Y., Xie, S., Wang, Y., & McDougall, D. (2020). Analysis of Students' Systematic Errors and Teaching Strategies for 3-Digit Multiplication. In S. C. Zhu, S. Xie, Y. Ma, & D. McDougall (Eds.), *Reciprocal Learning for Cross-Cultural Mathematics Education : A Partnership Project Between Canada and China* (pp. 159–182). Springer International Publishing. https://doi.org/10.1007/978-3-030-56838-2_9

Maharani, K., Effendi, M. M., & Nurjaman, A. (2020). Analisis tingkat ketercapaian angka partisipasi sekolah di papua (studi kasus di kabupaten merauke). *Jurnal Kebijakan Dan Pengembangan Pendidikan*, *7*(2). https://doi.org/10.22219/jkpp.v7i2.12046

Mahpop, H. bin, & Sivasubramaniam, P. (2010). Addition of Whole Numbers with Regrouping using the "Soroban." *Procedia - Social and Behavioral Sciences*, *8*, 50–56. https://doi.org/10.1016/j.sbspro.2010.12.007

Mardapi, D. (2018). *Teknik penyusunan instrumen tes dan nontes*. Parama Publishing.

Marsh, H. W., Balla, J. R., & Mcdonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391–410. https://doi.org/10.1037/0033-2909.103.3.391

Mauhibah, R., & Karso, K. (2020). Student Difficulties in Addition and Subtraction of Two Digit Numbers. *International Conference on Elementary Education*, *2*(1), 618–623.

McComas, W. F. (2014). Programme for International Student Assessment (PISA). In W. F. McComas (Ed.), *The Language of Science Education: An Expanded Glossary of Key Terms and Concepts in Science Teaching and Learning* (p. 79). SensePublishers. https://doi.org/10.1007/978-94-6209-497-0_69

Moon, G., Aitken, G., Roderick, P., Fraser, S., & Rowlands, G. (2015). Towards an understanding of the relationship of functional literacy and numeracy to geographical health inequalities. *Social Science & Medicine*, *143*, 185–193. https://doi.org/10.1016/j.socscimed.2015.08.045

Mulligan, J. T., & Mitchelmore, M. C. (1997). Young Children's Intuitive Models of Multiplication and Division. *Journal for Research in Mathematics Education JRME*, *28*(3), 309–330. https://doi.org/10.5951/jresematheduc.28.3.0309

Naumann, S., Byrne, M. L., de la Fuente, A., Harrewijn, A., Nugiel, T., Rosen, M., van Atteveldt, N., & Matusz, P. J. (2022). Assessing the degree of ecological validity of your study: Introducing the multidimensional assessment of research in context (MARC) tool. *Mind, Brain, and Education*, *16*(3), 228–238. https://doi.org/10.1111/mbe.12318

Novia, F., Nurdianti, D., & Purwanto, M. B. (2024). English learning and innovation skills in 21st: Implementation of critical thinking, creativity, communication, and collaboration. *Asian Journal of Applied Education (AJAE)*, *3*(2), 113–124. https://doi.org/10.55927/ajae.v3i2.8318

OECD. (2019). *PISA 2018 results (volume I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

Piasta, S. B., Park, S., Fitzgerald, L. R., & Libnoch, H. A. (2022). Young children's alphabet learning as a function of instruction and letter difficulty. *Learning and Individual Differences*, *93*, 102113. https://doi.org/10.1016/j.lindif.2021.102113

Pusmendik Kemdikbud. (2022). *Komponen, Level, Domain, Subdomain, Kompetensi dan Contoh Soal*. https://pusmendik.kemdikbud.go.id/an/asesmen_kompetensi_minimum/view/literasi-membaca

Retnawati, H. (2017). *Validitas reliabilitas & karakteristik butir*. Parama Publishing.

Ritchey, K. D. (2008). The building blocks of writing: Learning to write letters and spell words. *Reading and Writing*, *21*(1), 27–47. https://doi.org/10.1007/s11145-007-9063-0

Saifullah, & Yawan, H. (2022). Poverty and students' academic achievement in Papua, Indonesia. *International Journal of Indonesian Education and Teaching*, *6*(2), 341–350. https://doi.org/10.24071/ijiet.v6i2.4457

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology*, *126*, 120–137. https://doi.org/10.1016/j.jecp.2014.03.006

Strong, J. Z., Amendum, S. J., & Conradi Smith, K. (2018). Supporting Elementary Students' Reading of Difficult Texts. *The Reading Teacher*, *72*(2), 201–212. https://doi.org/10.1002/trtr.1702

Sülükçü, Y., & Kırboga, A. (2020). A study on the punctuation knowledge levels of freshmen students. *Journal of Language and Linguistic Studies*, *16*(2), 684–710. https://doi.org/10.17263/jlls.759273

Sumandya, I. W., Mukminin, A., Widana, I. W., Suryawan, I. P. P., Dewi, N. W. D. P., Hendra, R., & Yaakob, M. F. M. (2025). Development of an instrument to measure students' and teachers' perceptions of understanding by design-based mathematics learning evaluation in inclusive schools. *Discover Sustainability*, *6*(1), 797. https://doi.org/10.1007/s43621-025-01514-0

Suparman, A. R., Rohaeti, E., & Wening, S. (2022). Development of attitude assessment instruments towards socioscientific issues in chemistry learning. *European Journal of Educational Research*, *11*(4), 1947–1958. https://doi.org/10.12973/eujer.11.4.1947

Suparman, A. R., Rohaeti, E., & Wening, S. (2024). Development of computer-based chemical five-tier diagnostic test instruments: A generalized partial credit model. *Journal on Efficiency and Responsibility in Education and Science*, *17*(1), 91–105. https://doi.org/10.7160/eriesj.2024.170108

Tavşanlı, Ö. F., & Kara, Ü. E. (2021). The Effect of a Peer and Self-Assessment-Based Editorial Study on Students' Ability to Follow Spelling Rules and Use Punctuation Marks Correctly. *Participatory Educational Research*, *8*(3), 268–284. https://doi.org/10.17275/per.21.65.8.3

Thi, H. C., Le, T. A., Ngoc, B. T., & Phuong, T. P. T. (2023). Factors affecting the numeracy skills of students from mountainous ethnic minority regions in Vietnam: Learners' perspectives. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186X.2023.2202121

Wildmon, M. E., Skinner, C. H., McCurdy, M., & Sims, S. (1999). Improving secondary students' perception of the "dreaded mathematics word problem assignment" by giving them more word problems. *Psychology in the Schools*, *36*(4), 319–325. https://doi.org/10.1002/(SICI)1520-6807(199907)36:4<319::AID-PITS5>3.0.CO;2-T

Wilkins, J. L. M., Woodward, D., & Norton, A. (2021). Children's number sequences as predictors of later mathematical development. *Mathematics Education Research Journal*, *33*(3), 513–540. https://doi.org/10.1007/s13394-020-00317-y

Yasukawa, K., Brown, T., & Black, S. (2013). Production workers' literacy and numeracy practices: using cultural-historical activity theory (CHAT) as an analytical tool. *Journal of Vocational Education & Training*, *65*(3), 369–384. https://doi.org/10.1080/13636820.2013.820214