EDUKASIA: Jurnal Pendidikan dan Pembelajaran
Vol. 6, 2 (December, 2025), pp. 891-902
ISSN: 2721-1150; EISSN: 2721-1169; DOI: 10.62775/edukasia.v6i2.1641

Psychometric Evaluation of a Contextualized Literacy Instrument for Primary Students in Papua

Apri Damai Sagita Krissandi

Universitas Sanata Dharma, Indonesia; apridamai@usd.ac.id

ARTICLE INFO

Keywords:

literacy;
psychometrics;
instrument;
contextual;
Papua

Article history:

Received 2025-07-22
Revised 2025-09-20
Accepted 2025-10-12

ABSTRACT

National literacy tests often fail in Indonesia's 3T regions (Frontier, Remote, and Disadvantaged) because of cultural bias and limited contextual relevance, resulting in distorted measurements of children's actual abilities. This study addresses that gap by psychometrically evaluating a contextual literacy instrument composed of images and texts, designed specifically for primary school children in Papua. The instrument assesses functional literacy skills—literal, inferential, and evaluative—using familiar local contexts to enhance readability and engagement. This is among the first literacy assessments in Papua to integrate both visual and textual elements with systematic psychometric validation, thereby directly challenging the urban bias embedded in national literacy assessments. A descriptive quantitative methodology was employed with sixth-grade pupils in Papua, examining item validity, reliability, difficulty levels, and discriminative power. Results show that 18 of 20 items demonstrated strong validity ($r > 0.30$), with a KR-20 reliability coefficient of 0.79, indicating satisfactory internal consistency. The difficulty distribution was 65% moderate, 20% easy, and 15% difficult, while 70% of items effectively distinguished between high- and low-performing pupils. Incorporating local visual and contextual features enhanced item relevance without compromising psychometric quality. These findings suggest that the instrument is effective for assessing literacy in low-performing regions such as Papua, although results remain preliminary due to the small-scale pilot with only 32 students.

Corresponding Author:
Apri Damai Sagita Krissandi
Universitas Sanata Dharma, Indonesia; apridamai@usd.ac.id

1. INTRODUCTION

Fundamental literacy skills are central to educational achievement and active participation in society (Paris, 2012). In Indonesia, literacy has been prioritized through national policies and assessments; however, large disparities persist across regions (Swargiary, n.d.). In frontier, remote, and

disadvantaged (3T) areas such as Papua, challenges are particularly acute (Harsono et al., 2024) (Sidauruk, Susilowati, & Akbar, 2025). Only 47.5% of students in the province meet minimum literacy benchmarks, far below the national average (Wijaya et al., 2025). In Asmat, for example, many sixth graders struggle to comprehend basic texts and to construct accurate sentences, underscoring the urgent need for culturally responsive literacy assessment (Othman, 2025) (Foster, 2025).

National literacy tests have often been criticized for their limited cultural relevance (Shorrocks-Taylor & Hargreaves, 1999). Instruments developed in urban contexts frequently employ language, examples, and references unfamiliar to children in remote areas (Miller, Webster, Knight, & Comino, 2014), which risks producing distorted outcomes (Cormack & Comber, 2013). Nugraheni and Budiman (2021) describe this as an "urban paradigm bias," where the lived experiences of 3T pupils are systematically excluded from test design. Consequently, low scores in Papua may reflect contextual misalignment rather than genuine literacy deficits (Brown, 1980).

Several prior studies have highlighted the importance of cultural and visual contextualization in literacy assessment. Mufidah, Harsiati, and Nurchasanah (2023) developed a PIRLS-based instrument using culturally educative texts, which proved effective for measuring reading competence in Indonesian primary schools (Mufidah & Harsiati, 2023). Similarly, Kurniawati, Wahyuriningsih, and Stotlikova (2024) designed and validated a visual literacy instrument for elementary students, demonstrating that integrating graphics can enhance both engagement and measurement accuracy. In the domain of contextualized assessment (Kurniawati, Wahyuriningsih, & Stotlikova, 2025), Widiyanti and Susilayati (2023) confirmed the validity and reliability of a science literacy tool rooted in Islamic and Indonesian values, highlighting the feasibility of culturally responsive approaches (Widiyanti & Susilayati, 2023). The use of local wisdom as a basis for literacy materials was also supported by Hanifah et al. (2022), who found that contextualizing reading content with local traditions in Kuningan increased student motivation (Hanifah, Hamidah, Suntini, Trang, & Zulfitriyani, 2025). Visual elements have been shown to play a crucial role as well: Alam, Suryatna, and Kusumadinata (2020) demonstrated that visual communication media improved students' literacy skills (Alam, Suryatna, & Kusumadinata, 2020), while Fikri, Lestari, and Zulaikah (2024) reported that visual literacy significantly enhanced reading comprehension among secondary school pupils (Fikri & Lestari, 2024). Finally, Syafruddin et al. (2023) emphasized the need for culturally responsive digital learning materials, reflecting students' demand for resources that align with their sociocultural backgrounds (Syafruddin, Wahyuni, Ananda, & Rachmaningsih, 2025).

Several initiatives have attempted to address this issue by contextualizing assessment tools through simplified language and local themes (Mery, Newby, & Peng, 2011). However, most of these remain predominantly text-based and provide little integration of visual elements that could enhance comprehension and engagement (Keane, 2025). Prior research demonstrates that culturally relevant graphics can significantly strengthen assessment validity in diverse contexts (Hajaroh, Purwastuti, & Nurhayati, 2021). Nonetheless, few instruments in Indonesia systematically incorporate visual alongside textual content, and when they do, they often lack psychometric validation (Nurussaniah, Setyosari, Kuswandi, & Ulfa, 2025). For instance, Hanif and Lustyantie (2021) emphasized the potential of visuals but did not conduct rigorous validation, while Nugraheni and Budiman (2021) highlighted bias in national assessments without proposing concrete design solutions (Au & Raphael, 2000).

This study seeks to fill that gap by developing and psychometrically evaluating a contextualized literacy instrument for sixth-grade pupils in Asmat, Papua. Unlike prior tools that emphasize textual content only, the instrument integrates both visual and textual stimuli and is validated through item analysis, reliability testing, and discrimination indices. The aim is to produce an assessment that is not only psychometrically sound but also culturally inclusive, thereby ensuring fairer evaluation of literacy competencies in marginalized regions.

2. METHOD

This study employed a descriptive quantitative methodology to examine the viability of a contextual literacy assessment tool composed of images, texts, and tables (Sugiyono, 2018)(Prof.Dr.Sugiyono, 2022). The instrument was specifically designed for sixth-grade primary school pupils in 3T regions (frontier, remote, and disadvantaged), with a focus on Asmat Regency, Papua. The objective was to statistically assess the quality of each test item through analyses of validity, reliability, item difficulty, and discrimination index.

Instrument Development. Test items were created to reflect the lived experiences of Papuan pupils, incorporating short texts, form excerpts, narrative visuals, and activity tables. The instrument consisted of 20 items measuring literal comprehension, inferential reasoning, and sentence construction. Graphic and linguistic components were carefully chosen to ensure readability, cultural familiarity, and clarity of instructions. The instrument utilized in this study can be accessed at the following link: [instrument link].

Sample and Setting. The instrument was administered to 32 sixth-grade students from several purposively selected schools in Asmat. The schools were chosen not only for their accessibility to researchers but also for their representativeness of local cultural and linguistic contexts, ensuring that pupils' daily experiences were adequately reflected. Although the sample size was modest, it provided important preliminary insights into the applicability of the tool. In psychometric research, a commonly cited rule of thumb recommends 5–10 respondents per item (Teeluckdharry, Teeroovengadum, & Seebaluck, 2021). With 20 items, this would imply a minimum of 100–200 respondents. The present sample of 32 therefore falls short of this threshold, positioning the study as a pilot investigation and highlighting the need for larger-scale validation.

Data Collection. Students completed the test individually using printed sheets during regular school hours, supervised by classroom teachers. Responses were scored with a rubric developed by the researchers. Ethical protocols were followed, with formal authorization obtained from participating schools.

Data Analysis. Data were analyzed using Microsoft Excel. Item validity was determined using Pearson Product-Moment correlations, with coefficients ≥0.30 considered acceptable (Arikunto, n.d.). Reliability was assessed via Kuder-Richardson Formula 20 (KR-20), where ≥0.70 indicated satisfactory consistency (Pearson & Lumpkin, 2011). Item difficulty was classified as easy (>0.70), moderate (0.30–0.70), or difficult (<0.30). Discrimination indices were calculated by contrasting responses of high- and low-performing pupils; values ≥0.40 were deemed excellent. While Excel provided the basic functions required for these analyses, its capacity is limited compared to dedicated psychometric software. For instance, advanced approaches such as factor analysis or Item Response Theory (IRT) modeling could not be conducted within Excel, restricting the depth of psychometric evidence obtained in this study.

Summary. Overall, this methodology allowed for a preliminary yet systematic evaluation of the instrument's psychometric properties. At the same time, limitations related to sample size, school selection, and analytic tools emphasize that findings should be interpreted cautiously and verified in subsequent studies with larger, more diverse samples and more robust statistical techniques.

3. FINDINGS AND DISCUSSION

3.1. Quality of the Literacy Instrument

The contextual literacy measure created for the specific setting of Papua shown robust psychometric features. The pilot test with 20 questions demonstrated substantial internal validity, with 18 items attaining a Pearson correlation coefficient of ≥ 0.30, so satisfying the minimum criterion for valid items in educational assessment (Arikunto, 2021). Two items did not satisfy this criteria and necessitate change due to inadequate item-total correlations. The Kuder-Richardson 20 (KR-20) coefficient for reliability was 0.79, categorizing it as high. This number surpasses the conventional threshold of 0.70, commonly employed as a measure of acceptable internal consistency (Nunnally,

1975). The test is both dependable and uniform in assessing the literacy competencies of the sampled kids.

The bulk of items were classified as moderate in difficulty, with around 65% yielding accurate response rates between 0.30 and 0.70. Merely 20% of items were deemed easy (> 0.70), while 15% were classified as challenging (< 0.30). This distribution reflects an equitable range of difficulty levels, suitable for evaluating various student competencies. No valid items were so challenging that they went unanswered by the majority of students, however a couple were exceedingly simple (e.g., items answered correctly by 100% of respondents). This indicates that the measure may somewhat advantage higher-performing pupils, several of whom attained perfect scores.

Concerning item discrimination, 70% of the items exhibited discrimination indices classified as fair to good (≥ 0.30), with no valid items demonstrating very low discrimination. Significantly, 30% of items attained a "good" discrimination index (> 0.40), demonstrating their efficacy in distinguishing between high- and low-performing pupils. The statistical analysis verifies that the instrument is of satisfactory quality: the items are valid, reliable, demonstrate an appropriate difficulty range, and show sufficient discriminatory power.

These findings are summarized in Table 1, which presents the statistical characteristics of each item based on validity, difficulty, and discrimination analyses.

Table 1. Item-Level Statistical Summary of the Literacy Instrument

No	Validity (r)	Validity Category	Item Difficulty	Difficulty Category	Discrimination Index	Discrimination Category
1	0.51	Valid	0.65	Moderate	0.47	Good
2	0.49	Valid	0.56	Moderate	0.42	Good
3	0.46	Valid	0.71	Easy	0.38	Fair
4	—	Invalid	—	—	—	—
5	0.43	Valid	0.68	Moderate	0.31	Fair
6	0.47	Valid	0.34	Moderate	0.35	Fair
7	0.53	Valid	0.39	Moderate	0.44	Good
8	0.50	Valid	0.43	Moderate	0.40	Good
9	0.48	Valid	0.77	Easy	0.36	Fair
10	0.46	Valid	0.71	Easy	0.30	Fair
11	0.44	Valid	0.29	Difficult	0.25	Low
12	0.51	Valid	0.28	Difficult	0.28	Low
13	0.42	Valid	0.64	Moderate	0.38	Fair
14	0.43	Valid	0.59	Moderate	0.33	Fair
15	0.45	Valid	0.32	Moderate	0.29	Low
16	0.49	Valid	0.70	Moderate	0.39	Fair
17	0.50	Valid	0.57	Moderate	0.41	Good
18	0.39	Valid	0.30	Moderate	0.26	Low
19	—	Invalid	—	—	—	—
20	0.51	Valid	0.35	Moderate	0.37	Fair

To provide a more comprehensive summary of the overall quality of the instrument, Table 2 presents the distribution of items based on validity, difficulty, discrimination index, and reliability results.

Table 2. Summary of Instrument Feasibility Categories

Aspect	Number of Items	Notes
Valid items	18 items	r ≥ 0.30
Invalid items	2 items	No significant correlation (Items 4, 19)
Items with moderate difficulty	13 items	Difficulty index between 0.30–0.70
Easy items	4 items	Difficulty index > 0.70
Difficult items	3 items	Difficulty index < 0.30
Items with good discrimination	6 items	Discrimination index > 0.40
Items with fair discrimination	8 items	Discrimination index between 0.30–0.39
Items with low discrimination	4 items	Discrimination index between 0.20–0.29
Instrument reliability (KR-20)	—	0.79 (High category)

The efficacy of this instrument is intricately linked to the endeavor of creating contextually relevant and unequivocal test items. Every question included clear instructions and a precise solution key, reducing the likelihood of evaluation bias. During the item evaluation, just a few small flaws were identified; for example, Item 1 utilized the phrase “one object” although the accompanying image depicted multiple items, perhaps causing confusion for inattentive students. The term "role" in Item 10 was quite ambiguous for several pupils. Overall, the instructions were straightforward, the answer keys precise, and the content consistent with the specified literacy competencies.

The significant proportion of valid and reliable items indicates both the readability and suitability of the items concerning the target population's capabilities. This instrument has empirically demonstrated its feasibility for assessing the literacy of sixth-grade children in Papua. This feature indicates that incorporating local context enhances the instrument: modifications in language, visuals, and item structure to resonate with Papuan students' experiences boosted its clarity and accessibility. The research indicates that linguistic simplification, such as use concise and direct sentences, along with recognizable visual signals, improves student understanding. These findings align with the perspective that culturally responsive examinations are generally more equitable and precise in reflecting true student capability. Conversely, decontextualized national instruments are frequently regarded as prejudiced and inequitable towards pupils from diverse cultural backgrounds, highlighting the necessity for the creation of culturally adapted literacy assessments.

A robust literacy assessment must demonstrate content and construct validity, high reliability, and statistically sufficient item attributes, including balanced difficulty and discrimination indices. The aforementioned evaluation findings demonstrate that this instrument satisfies the specified criteria. The KR-20 value nearing 0.8 indicates that the instrument meets the reliability criteria in educational assessment (Nunnally, 1975). Elevated item validity signifies that each item accurately assesses the designated components of fundamental literacy. The fair discrimination index indicates that the instrument can accurately rank pupils by ability, which is crucial for formative assessment.

The contextual literacy instrument created for Papua has shown empirical strength and conformity with recognized psychometric standards, rendering it a promising and inclusive tool for literacy assessment in under-resourced and culturally diverse regions like Indonesia's 3T areas.

3.2. Student Performance on the Instrument

The initial assessment of the literacy tool demonstrated promising outcomes about the performance of Papuan pupils, underscoring specific patterns of strengths and weaknesses. Students exhibited significant accomplishment, averaging 17 accurate responses out of 20 questions (about 85%), with a median score of 18. Significantly, 25% of the pupils (6 out of 24) attained perfect scores, and over half of the class secured a minimum of 18 right answers. Only one student achieved a score below 50% (9/20), while no students received exceptionally low results. This indicates that all things fell within the cognitive capacity of the pupils; even the lowest-performing students could correctly answer several

items. The results demonstrate that sixth-grade Papuan students have robust foundational literacy skills when the assessment is conducted inside a familiar context. This discovery is noteworthy, considering literacy inequalities in Papua are frequently ascribed to poor performance on standardized tests. Nonetheless, with contextually suitable instruments, pupils were able to more properly exhibit their genuine literacy capacity.

At the item level, pupils excelled on literal questions with clearly articulated answers. Items 5, 6, 7, and 11 were answered correctly by all pupils. The inquiries concentrated on basic factual information—such as specifics from photographs or texts (e.g., poem titles, submission dates)—and presented no challenges. Supplementary literal/inferential items, including Items 3, 9, 12, and 13, were answered correctly by ≥92% of students, demonstrating robust literal comprehension, especially in identifying clear information (who, what, when, where). This corresponds with national statistics indicating that Indonesian primary school kids typically succeed in recognizing explicit information. Provided that instructions are unambiguous and contextually pertinent, pupils can effectively extract explicit and straightforward inferential knowledge.

Nonetheless, student performance deteriorated on tasks necessitating advanced reasoning or producing abilities. Item 15, an evaluative question requiring students to compare text and artwork, proved to be the most challenging, yielding just 62% accurate replies. Approximately 40% of students encountered challenges with this critical evaluation test, presumably due to difficulty in identifying discrepancies between narrative material and visual components. Likewise, Item 1, although straightforward, was answered correctly by merely 62% of pupils. This may have sprung from inattention or confusion due to the directive to "name one object" while numerous were illustrated, indicating problems with precision or interpretation of instructions rather than understanding.

Additional difficult items comprised Items 16 and 20 (about 71% accuracy). Item 16 necessitated the interpretation of a timetable, leading to confusion among certain students, whereas Item 20 needed students to formulate a sentence pertaining to a certain activity—29% of students were unable to generate grammatically proper words. Errors frequently resulted from incorrect sentence structure or spelling rather than a misinterpretation of the substance.

Certain open-ended inferential questions presented modest difficulty yet were generally answered satisfactorily. For example, 79% of students responded correctly to Item 19, which suggested an alternative activity for inclement weather. Items 17–18 (table reading) and Items 8, 10, and 14 exhibited success rates ranging from 75% to 79%. The principal issue seemed to arise from the unconventional format (e.g., tables or forms), which may have impeded pupils' performance. Furthermore, abstract terminology (e.g., "role" in Item 10) appeared to be unfamiliar, underscoring students' restricted academic vocabulary, which likely hindered their capacity to comprehend the question.

The overall trend suggests that higher-order thinking questions, especially inferential and evaluative types, posed more challenges for students, yet remained attainable for many. The findings align with established educational literature (Richland & Simms, 2015), indicating that literal questions are typically less challenging for primary pupils, but inferential and evaluative activities present greater difficulty. In this study, Papuan students excelled in literal tests, fared satisfactorily in simple inferential skills, and encountered difficulties with critical evaluation. This indicates not a deficiency in capability but rather inadequate training, as early elementary curricula typically prioritize fundamental understanding, with critical literacy skills introduced subsequently.

Moreover, student-level data indicated individual variability. Notwithstanding elevated average scores, few students achieved markedly lower results (as low as 45% accuracy), especially on open-ended or writing-centric assignments. These students may necessitate specialized assistance and corrective education. In contrast, the 25% who achieved perfect scores indicate that the test may have been excessively facile for high achievers, hence constraining its capacity to distinguish among top performers. Consequently, subsequent versions may integrate more challenging questions with enhanced discriminative capability while maintaining accessible for lower-performing students.

The finding that one-quarter of pupils attained perfect scores suggests a potential ceiling effect: the instrument may not have been sufficiently challenging for high-achieving students. While the contextual design successfully improved fairness for the majority, it also compressed score variance at the upper end. For diagnostic and policy purposes, this limits the instrument's ability to distinguish among top performers. Future iterations should therefore include more complex evaluative and open-ended tasks—e.g., requiring students to justify an opinion or critique discrepancies between multiple sources—while still maintaining accessibility for lower-performing pupils. Doing so would prevent the instrument from plateauing in its discriminative power and would ensure its utility across the full spectrum of student ability.

The presence of a ceiling effect, with approximately one-quarter of students achieving perfect scores, suggests that the instrument lacked sufficient challenge for higher-achieving pupils. Future versions should therefore include a wider range of item difficulties to capture variation across the full ability spectrum.

In conclusion, Papuan pupils excelled in this literacy assessment, particularly in literal comprehension. They demonstrated developing inferential abilities but continued to encounter difficulties with evaluative and expressive tasks. These findings indicate that instructional emphasis should persist on enhancing fundamental comprehension while augmenting engagement with critical thinking and writing articulation. This performance demonstrates that culturally and contextually relevant exams allow pupils to exhibit their genuine capabilities, countering deficit-oriented preconceptions sometimes derived from urban-centric standardized tests.

3.3. Cultural Context Appropriateness in Test Item Development

This literacy instrument's primary strength is its compatibility with the cultural environment of Papuan students. All objects were created utilizing a contextual methodology grounded in students' daily experiences and local cultural values. Items 1–5 depicted images of spear fishing activities—recognizable scenarios for coastal Papuan children. Imagery illustrating youngsters diving, fishing, and transporting their catch enabled pupils to grasp the tale seamlessly, devoid of cultural impediments. Such activities remained creatively accessible even for pupils from mountainous regions. Items 6–10 included a form for a poem reading competition for Papua Children's Day. Names such as Yuli and Markus, along with the term "lembah" (valley), exemplified by "Lembah Baliem," further solidified contextual familiarity. This not only enhanced emotional involvement but also familiarized students with practical texts in a pertinent and accessible way.

Items 11–15 had a tale entitled Tuesday Market in Wamena Village, which graphically illustrated Papuan cultural aspects, including mama-mama (female sellers) selling agricultural products, the utilization of noken (traditional bags), and native goods such as taro, pineapples, and cassava leaves. The story successfully engaged students' cognitive frameworks on village marketplaces, as demonstrated by the elevated success rate on Items 11–13 (≥96%). Although disparities were there between the pictures and the text, they were intentionally employed as the foundation for inferential and evaluative inquiries. Items 16–20 included a schedule from "Papua Creative Children's Elementary School," comprising contextually rich activities such as Alphabet Adventures in the Story Forest, Painting the Tree of Papuan Life, Mini Garden, and Traditional Dance and Exercise. While 20–30% of students experienced challenges in reading the table, the difficulties seemed to be technical rather than stemming from a lack of familiarity with the topic.

All texts utilized standard Indonesian, simplified for accessibility, while effectively integrating local language such as noken, mama-mama, keladi (taro), and yel-yel (chants), so enhancing students' comfort and interest. The content aligned with local cultural norms, reinforcing traditional identity and principles such as reverence for elders, the significance of folklore, and communal cooperation. This method is consistent with the tenets of culturally sustaining pedagogy (Paris, 2012), which acknowledges students' cultural backgrounds and actively preserves them throughout the educational

experience. The trial results indicate that this contextual method enhanced reading comprehension and student motivation.

These findings underscore a mismatch between the design of national assessments and the lived experiences of pupils in marginalized areas. If evaluative items are presented without scaffolding or contextual alignment, they risk being misinterpreted as evidence of student deficiency rather than as signals of curricular and instructional gaps. National assessments therefore need to integrate culturally relevant stimuli and provide graduated levels of cognitive demand. Embedding local cultural references in critical-thinking items may help pupils practice evaluative reasoning while still feeling grounded in familiar contexts. This approach would align national testing with the *Merdeka Curriculum*'s emphasis on both equity and higher-order skills.

Numerous research show the efficacy of culturally embedded assessment design. Kholid, Zubaidah, and Sari (2022) observed that inadequate literacy levels in 3T regions frequently result from a discordance between assessment tools and the cultural contexts of students (Bansilal & Debba, 2012). Tam (2015) similarly condemned the urban bias in national assessments, which disadvantages students in rural regions (Tam & Jiang, 2015). Trumbull and Nelson-Barber (2019) asserted that assessments devoid of Indigenous cultural context are often unhelpful and invalid at the international level (Preston & Claypool, 2021). In contrast, culturally sensitive exams are more egalitarian and precise in demonstrating students' abilities. A recent study by Kristiono, Harsono, and Minsih (2025) confirmed the significance of contextual methodologies in literacy teaching in Papua. This tool indicated that integrating local cultural content into exams enhances student recognition, engagement, and performance efficacy. This demonstrates that a locality-based strategy does not impede standardization; instead, it enhances educational fairness and instructional efficacy (Robertson, Curtis, & Dann, 2018).

3.4. Literacy Challenges among Papuan Students: Writing Production and Critical Evaluation

The pilot application of the contextual literacy instrument in Papua identified two significant difficulties requiring urgent attention: pupils' writing proficiency and their critical assessment capabilities. The two competences, situated at the advanced tiers of the literacy continuum, seem to be inadequately developed owing to a lack of effective facilitation in routine classroom practices. Although the majority of students exhibited robust literal and fundamental inferential comprehension, deficiencies emerged when tasked with composing original written replies or critically assessing reading passages and visual materials.

Initially, regarding writing creation, numerous items (e.g., Items 4, 8, 14, and 20) necessitated that students compose complete sentences in their responses. Despite a reasonably high percentage of right responses (between 70% and 88%), qualitative examination indicated that numerous answers did not adhere to the criteria of proper Indonesian writing conventions (Ejaan yang Disempurnakan, EYD) (Kurniawan, 2023). Frequent errors encompassed the absence of capital letters at the commencement of sentences, inadequate punctuation, and structurally deficient phrases (lacking subjects or predicates). For example, in Item 8, several students just replicated excerpts from the poetry competition form—such as "Hari Anak Papua" (Papua Children's Day) or "Bapak Markus"—without formulating coherent lines. This indicates that while students comprehended the material, they encountered difficulties in articulating it in grammatically accurate and coherent written expression.

This study aligns with the results of Oktavia, Sa'odah, and Ginanjar (2023), who indicated that elementary pupils typically encounter difficulties in implementing EYD norms, especially with capitalization and punctuation (Oktavia, Sa'odah, & Ginanjar, 2023). In Papua, the impact of indigenous languages and creole usage in everyday communication also influences pupils' cognitive processes in writing. The shift from vernacular languages to formal Indonesian, in the absence of enough practice, renders such writing errors comprehensible. Consequently, educators are urged to include specific daily writing assignments, such as crafting one or two phrases post-reading, with an emphasis on proper sentence structure. This incremental strategy is essential to guarantee that sixth-

grade kids establish a robust foundation prior to transitioning to secondary school, when writing proficiency becomes progressively significant. While evaluators in this study shown leniency towards flawed sentence structures if the content was valid, forthcoming examinations with more stringent criteria will necessitate significant enhancements in writing quality.

Secondly, a notable problem exists in students' restricted capacity for critical appraisal. This was particularly apparent in Item 15, which evaluated students' ability to compare two sources of information: a narrative text and a market image. This item exhibited the lowest accuracy rate, with merely approximately 62% of responses being right. A considerable number of students were unable to recognize anomalies (e.g., the omission of "pineapple" in the image despite its reference in the text), while others found it challenging to articulate their assessments coherently in writing. This question evaluated critical reading and visual literacy—two competencies that are seldom explicitly instructed in early education. Students were required to read the text meticulously, examine the image diligently, and subsequently assess both sources concurrently—a cognitively demanding task for young learners.

Moreover, challenges in inferential comprehension were noted in Item 10, which necessitated students to deduce Mr. Markus's role from a form. Certain pupils lacked comprehension of the term "role" or were unacquainted with deriving inferences from implicit information. This signifies an inadequate academic lexicon and a lack of proficiency in inferential reading. Critical literacy beyond the comprehension of apparent content; it involves assessing implicit meanings and contextual subtleties. According to Basaraba et al. (2013), evaluative comprehension represents the most challenging tier in reading taxonomies, necessitating the synthesis of knowledge and metacognitive judgment (Basaraba, Yovanoff, Alonzo, & Tindal, 2013). This corresponds with the performance trends noted in Papuan students—proficiency in literal and fundamental inferential comprehension, yet deficiencies in evaluative skills.

The ramifications are evident: educators must incorporate higher-order thinking inquiries into daily classroom practices. Activities like juxtaposing text and images or conducting class discussions with evaluative inquiries such as "What is your opinion on this character?" can foster critical thinking from an early age. It is essential to acknowledge the interconnection between writing and critical thinking—students proficient in articulating viewpoints are more adept at text evaluation, while critical reasoning subsequently improves writing quality. Consequently, literacy programs must integrate writing instruction with critical thinking exercises, such as requiring students to compose straightforward opinion sentences regarding the narratives they read.

These findings further corroborate national recommendations for focused literacy initiatives in 3T regions. Wijaya et al. (2025) revealed that children who obtain early literacy foundations through mother tongue-based preschool programs, like the "Wahana Literasi" initiative in Papua, attain superior literacy outcomes (Wijaya et al., 2025). For present sixth-grade children, pertinent interventions may encompass basic writing clinics or critical thinking courses tailored to local circumstances. This literacy evaluation verifies that contextualized methods not only evaluate student capability more fairly but also diagnostically identify particular areas requiring enhancement. The 2013 Curriculum and Merdeka Curriculum prioritize Higher Order Thinking Skills (HOTS), necessitating corresponding actions include teacher training, the creation of writing modules, and the provision of enrichment reading resources that foster critical thinking (Nasional, 2013) (Mulyasa, 2023).

Evaluative questions (e.g., Item 15) proved more challenging largely because critical literacy has not been a major focus in earlier grades. Both the 2013 Curriculum and the current *Kurikulum Merdeka* emphasize Higher Order Thinking Skills (HOTS), yet in practice, classroom teaching in 3T regions such as Papua often prioritizes literal understanding and basic decoding. This gap reflects limited exposure to tasks requiring synthesis, comparison, or judgment. Furthermore, the sociolinguistic context of Papua—where students navigate between local vernaculars and formal Indonesian—reduces familiarity with abstract academic vocabulary such as "role," further constraining performance. Social learning habits also tend to be oral and communal, whereas evaluative items require individual, text-based judgment, creating additional cognitive demand.

In summary, this literacy instrument serves both as an evaluative tool and as a stimulus for more impactful educational policies. The synthesis of empirical evidence with current literature underscores two primary objectives for enhancing literacy in Papua: proficiency in fundamental sentence construction and the development of critical thinking skills. By enhancing these skills while preserving the local cultural environment, the literacy of Papuan pupils can improve significantly and sustainably.

The comparatively lower scores on evaluative items should not be interpreted as evidence of inherent cognitive limitations. Rather, they reflect curriculum orientation, limited exposure to academic registers, and community-based oral learning practices in Papua. This finding underscores the importance of designing assessments that gradually scaffold evaluative literacy tasks while remaining sensitive to local contexts.

These results also signal that national literacy assessments must avoid over-reliance on decontextualized, urban-centric stimuli. A more equitable approach would integrate culturally familiar content while simultaneously providing pathways for students in remote areas to engage with higher-order literacy skills.

4. CONCLUSION

The psychometric evaluation of the contextual literacy instrument, which integrates images and texts for primary school pupils in Papua, demonstrates that the tool is of high quality and suitable for application in 3T (Frontier, Remote, and Disadvantaged) areas. Of the 20 items, 18 satisfied the minimum validity requirement ($r \geq 0.30$), confirming their effectiveness in measuring key reading skills. The KR-20 reliability coefficient of 0.79 indicates substantial internal consistency. Item difficulty was well distributed, with most items categorized as moderate, and the discrimination indices showed that the majority of items could effectively differentiate student performance. These findings highlight that a regionally contextualized and visually assisted assessment approach enhances readability, engagement, and fairness for Papuan students. The tool also successfully identified specific literacy gaps, particularly in writing skills and evaluative reasoning.

This study, however, is limited by its small pilot nature with only 32 students, which restricts the generalizability of its findings. According to psychometric standards, further testing with larger and more diverse samples is required to confirm the robustness of the instrument and to ensure its applicability across different cultural and linguistic environments.

Despite these limitations, the study offers three key practical implications. First, there is a strong need for the development of contextualized literacy instruments on a larger scale so they can complement and strengthen national assessments. Second, teacher training and professional development are essential to equip educators with the skills to apply, interpret, and maximize the benefits of such tools in classroom practice. Third, the findings suggest that integration of culturally sensitive literacy instruments into national education policy is crucial to build an assessment system that is not only psychometrically rigorous but also inclusive and equitable.

In conclusion, while this pilot project represents an initial step, it demonstrates the promise of culturally grounded and multimodal literacy instruments in advancing fairer and more effective literacy assessment in Indonesia's most diverse and under-resourced regions. Expanding the scale of development, enhancing teacher capacity, and embedding such tools into national policy frameworks can ensure sustainable improvements in both the quality of assessment and the educational outcomes of children in Papua and other 3T regions.

Acknowledgments: The author would like to express sincere gratitude to the Faculty of Teacher Training and Education (FKIP), Sanata Dharma University, for the academic support provided during the research. No external funding or material support was received for this study.

REFERENCES

Alam, K., Suryatna, U., & Kusumadinata, A. A. (2020). Improving student literacy skills through visual communication media. *Indonesian Journal of Social Research (IJSR)*, *2*(2), 126–133.

Arikunto, S. (n.d.). *Prosedur Penelitian: Suatu Pendekatann Praktik. Jakarta: Rineka Cipta. 2013.*

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

Au, K. H., & Raphael, T. E. (2000). Equity and literacy in the next millennium. *Reading Research Quarterly*, *35*(1), 170–188.

Bansilal, S., & Debba, R. (2012). Exploring the role of contextual attributes in a Mathematical Literacy assessment task. *African Journal of Research in Mathematics, Science and Technology Education*, *16*(3), 302–316.

Basaraba, D., Yovanoff, P., Alonzo, J., & Tindal, G. (2013). Examining the structure of reading comprehension: Do literal, inferential, and evaluative comprehension truly exist? *Reading and Writing*, *26*(3), 349–379.

Brown, R. (1980). *Contributions of the National Assessment to Understanding the Problems of Literacy and Equity.*

Cormack, P., & Comber, B. (2013). High-stakes literacy tests and local effects in a rural school. *The Australian Journal of Language and Literacy*, *36*(2), 78–89.

Fikri, A. H., & Lestari, L. P. (2024). The The Influence Of Visual Literacy In Improving Reading Comprehension Of Grade 8 Students At SMPN 01 Buay Madang: The Influence Of Visual Literacy In Improving Reading Comprehension Of Grade 8 Students At SMPN 01 Buay Madang. *ENJEL: English Journal of Education and Literature*, *3*(2), 204–208.

Foster, R. (2025). *Culturally Informed Literacy Curriculum for Underserved Urban Elementary Students*.

Hajaroh, M., Purwastuti, L. A., & Nurhayati, R. (2021). Development of the Evaluation Instrument of the Child-Friendly School Policy in Elementary School. *International Journal of Instruction*, *14*(3), 327–340.

Hanifah, I., Hamidah, I., Suntini, S., Trang, T. T. T., & Zulfitriyani, Z. (2025). Develovment of literacy teaching materials in elementary schools based on local wisdom in Kuningan. *BAHASTRA*, *45*(2), 190–205.

Harsono, I., Sutanto, H., Purwadinata, S., Astuti, E., Wafik, A. Z., & Ekonomi, F. (2024). The Effect of Economic Factors, Health Conditions, and Access to Education on Social Inequality and Poverty in West Papua. *West Science Social and Humanities Studies*, *2*(01), 174–181.

Keane, K. (2025). Text-based discussion interactions and engagement with three different text types. *The Journal of Educational Research*, *118*(2), 185–199.

Kurniawan, I. (2023). *EYD Ejaan yang Disempurnakan*. Nuansa Cendekia.

Kurniawati, K., Wahyuriningsih, I., & Stotlikova, R. (2025). Development of Visual Literacy Assessment Instrument for Elementary School Students. *DIDAKTIKA: Jurnal Pendidikan Sekolah Dasar*, *8*(1), 65–76.

Mery, Y., Newby, J., & Peng, K. (2011). Assessing the reliability and validity of locally developed information literacy test items. *Reference Services Review*, *39*(1), 98–122.

Miller, E., Webster, V., Knight, J., & Comino, E. (2014). The use of a standardized language assessment tool to measure the language development of urban Aboriginal preschoolers. *International Journal of Speech-Language Pathology*, *16*(2), 109–120.

Mufidah, I., & Harsiati, T. (2023). Progress in International Reading Literacy Study-Based Reading Competency Assessment Instrument with Culturally Educative Texts for Primary School Students. *Jurnal Penelitian Pendidikan IPA*, *9*(SpecialIssue), 677–685.

Mulyasa, H. E. (2023). *Implementasi kurikulum merdeka*. Bumi Aksara.

Nasional, D. P. (2013). Kurikulum 2013. *Direktorat Jendral Manajemen Pendidkan Dasar Dan Menengah Pembinaan TK SD*.

Nunnally, J. C. (1975). Psychometric theory—25 years ago and now. *Educational Researcher*, *4*(10), 7–21.

Nurussaniah, N., Setyosari, P., Kuswandi, D., & Ulfa, S. (2025). Psychometric Validation of an

Analytical Skills Test in Physics Using the Rasch Model. *Journal of Baltic Science Education*, *24*(3), 522–537.

Oktavia, P., Sa'odah, S., & Ginanjar, R. R. (2023). Analysis of Misuse of Eyd in Elementary School Students of SD Negeri Sangiang Jaya. *Enrichment: Journal of Multidisciplinary Research and Development*, *1*(8), 458–462.

Othman, O. (2025). *Revealing Students' and Teachers' Perceptions of Challenges and Strategies in Arabic Reading Comprehension*.

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, *41*(3), 93–97.

Pearson, A. W., & Lumpkin, G. T. (2011). Measurement in family business research: How do we measure up? *Family Business Review*, Vol. 24, pp. 287–291. SAGE Publications Sage CA: Los Angeles, CA.

Preston, J. P., & Claypool, T. R. (2021). Analyzing assessment practices for Indigenous students. *Frontiers in Education*, *6*, 679972. Frontiers Media SA.

Prof.Dr.Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*.

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 177–192.

Robertson, A., Curtis, P. M., & Dann, C. (2018). Outcomes of a Collaborative Contextualised Learning Approach to Teacher Professional Development in Papua, Indonesia. *International Education Journal: Comparative Perspectives*, *17*(2), 88–101.

Shorrocks-Taylor, D., & Hargreaves, M. (1999). Making it clear: A review of language issues in testing with special reference to the National Curriculum Mathematics Tests at Key Stage 2. *Educational Research*, *41*(2), 123–136.

Sidauruk, J. M., Susilowati, M., & Akbar, K. K. (2025). Indonesia's Struggle with Education Inequality: Is Reform the Answer? *Indonesia Discourse*, *2*(1).

Sugiyono, S. (2018). Metode Penelitian Pendidikan Pendekatan Kualitatif, Kuantitatif dan R & D. *Alfabeta, Bandung*, *4*.

Swargiary, K. (n.d.). *Literacy and Development: A Global Perspective*.

Syafruddin, S., Wahyuni, S., Ananda, R., & Rachmaningsih, D. M. (2025). Students' Perceptions of The Indonesian Language Course and Their Correlation with The Need for Culturally Responsive Digital Learning Materials. *SAWERIGADING*, *31*(1).

Tam, T., & Jiang, J. (2015). Divergent urban-rural trends in college attendance: State policy bias and structural exclusion in China. *Sociology of Education*, *88*(2), 160–180.

Teeluckdharry, N. B., Teeroovengadum, V., & Seebaluck, A. (2021). Scale development in marketing research. *Handbook of Research Methods for Marketing Management*, 17–50.

Widiyanti, F., & Susilayati, M. (2023). VALIDITY AND RELIABILITY OF ISLAMIC AND INDONESIAN-BASED CONTEXTUAL SCIENCE LITERACY ASSESSMENT INSTRUMENTS. *INSECTA: Integrative Science Education and Teaching Activity Journal*, *4*(1), 19–28.

Wijaya, A. W. A., Siantoro, A., Triwibowo, H., Asih, R. A., Sambo, M., & Layuk, M. (2025). Literacy improvement for remote primary school students in Papua, Indonesia: The'Wahana Literasi'program. *Issues in Educational Research*, *35*(1), 395–421.