

Conceptual framework for the Development Model of Numeracy Assessment of Local Cultural Contexts Based on CAT

Sutamrin¹, Abdul Rahman², Alimuddin³

^{1,2,3} Universitas Negeri Makassar, Indonesia

Corresponding Author: rahmanmallala@gmail.com

Abstract

This study aims to develop a conceptual framework for a numeracy assessment model grounded in local cultural contexts and supported by Computer Assisted Test (CAT). The proposed model addresses the persistently low numeracy proficiency among Indonesian students, as evidenced by national and international assessments such as PISA and the National Assessment, as well as the scarcity of culturally relevant and contextually meaningful assessment instruments. By integrating constructivist learning theory, ethnomathematics, authentic assessment principles, and item response theory (IRT), this framework is designed to render numeracy assessments more contextual, adaptive, and psychometrically reliable. The research yielded a comprehensive development model comprising the following stages: (1) planning through systematic literature review and field observation; (2) design of numeracy competency grids and culturally contextualized test items rooted in Bugis-Makassar traditions; (3) expert validation to ensure content and cultural validity; (4) digitization of instruments into a CAT platform; and (5) pilot testing to collect empirical data on item characteristics, reliability, and validity. This study is expected to make a significant contribution to enhancing students' numeracy literacy by promoting authentic, culturally embedded assessments. Furthermore, it supports the broader adoption of digital assessment technologies in educational settings. By engaging students with assessment experiences anchored in their local cultural heritage, the model aims to foster numeracy competencies that are not only academically robust but also critically reflective, contextually grounded, and applicable to real-world challenges.

Keyword : Computer Assisted Test (CAT), Numeracy assessment, Local cultural, context Ethnomathematics, Item Response Theory (IRT)



INTRODUCTION

Numeracy is among the foundational competencies essential for thriving in the 21st century, as it underpins the ability to process, interpret, and apply quantitative information across diverse real-world contexts. However, Indonesian students continue to demonstrate persistently low levels of numeracy proficiency. According to the Programme for International Student Assessment (PISA) 2018, Indonesia's mean mathematics score was 379, which is significantly below the OECD average of 489 (OECD, 2019; Schleicher, 2019). This result indicates that the majority of students have not attained Level 2 proficiency, the minimum threshold required to comprehend and solve basic mathematical problems encountered in daily life (World Bank, 2020). These findings are consistent with the research of Alimuddin et al. (2024) in Makassar, which revealed that junior high school students' mathematical literacy remains at a basic level and requires targeted intervention. In contrast, reading literacy scores in the same study were relatively higher. This persistent gap underscores the urgent need for innovative, contextually grounded, and meaningful assessment approaches to foster numeracy development.

The shift in Indonesia's education policy from the National Examination (UN) to the National Assessment (AN) further highlights the strategic prioritization of literacy and numeracy as core pillars of foundational learning. The Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) explicitly positions literacy and numeracy as key indicators of meaningful learning outcomes and benchmarks for national education quality improvement (Kemendikbud, 2021). Nevertheless, data from the 2022 National Assessment indicate that over 50% of students nationwide—including those in South Sulawesi—remain below minimum competency thresholds in both domains (Pusmendik, 2022). This discrepancy between policy aspirations and actual student performance signals a critical need for assessment instruments that are not only psychometrically sound but also culturally relevant and locally contextualized. As noted by Trisnawati et al. (2022) and Siregar et al. (2024), embedding local cultural experiences into numeracy assessments enables students to meaningfully interpret mathematical concepts through familiar and authentic contexts.

Twenty-first century competencies require individuals to think critically, creatively, collaboratively, and communicatively, all of which are underpinned by robust literacy and numeracy skills. Morocco (as cited in Abidin, 2018) contends that literacy—including numeracy literacy—constitutes the bedrock for achieving 21st-century readiness. This view aligns with Trilling and Fadel's (2009) assertion that mastery of foundational literacies (reading, numeracy, and technological) is indispensable for navigating global challenges. Unfortunately, Indonesia's performance in these domains remains suboptimal. PISA 2018 reported reading and science





scores of 371 and 396, respectively, both of which fall below the OECD average of 487 (OECD, 2019). Under such conditions, Indonesian students face significant barriers to developing higher-order thinking skills (HOTS), which are crucial for academic and professional success in an increasingly complex global landscape. Thus, enhancing numeracy is not merely an academic imperative but also a strategic investment in cultivating a generation equipped to thrive in the 21st century.

Nationally, the 2022 National Assessment reaffirms that student numeracy achievement remains below minimum competency standards at both national and provincial levels, including in South Sulawesi (Pusmendik, 2022). More than half of students struggle to apply mathematical concepts across varied contexts or engage in reasoning to solve non-routine problems. This gap reflects a systemic misalignment between curricular expectations and classroom realities. In contrast, a growing body of research demonstrates that integrating local cultures into mathematics instruction enhances student motivation, conceptual understanding, and engagement (Ayuningtyas & Setiana, 2019; Jabali et al., 2020; Rizal et al., 2021). Ethnomathematics, as pioneered by D'Ambrosio (2001), offers a pedagogical approach that bridges abstract mathematical constructs with students' lived experiences through culturally resonant contexts. Consequently, developing culturally grounded numeracy assessments emerges as a vital strategy—not only to improve learning outcomes but also to reinforce students' cultural identity and appreciation of local wisdom.

Although numerous studies have successfully developed ethnomathematics-based teaching materials, modules, and learning media that enhance student engagement, research on culturally contextualized assessment instruments remains notably scarce (Ayuningtyas & Setiana, 2019; Jabali et al., 2020; Rizal et al., 2021). Assessment serves a dual function: it not only evaluates learning outcomes but also shapes and drives the learning process itself (Black & Wiliam, 2018). In this regard, Rahman et al. (2018) emphasize the transformative potential of digital assessment technologies, such as web- and Android-based platforms, to enhance the efficiency, accuracy, and reliability of instruments, thereby meeting the demands of modern, adaptive, and flexible evaluation systems. In South Sulawesi, the rich cultural heritage of the Bugis–Makassar people—characterized by symbolic artifacts, traditional practices, and everyday activities—offers fertile ground for developing authentic, contextually embedded numeracy assessments. Integrating such cultural elements not only enhances content relevance and student engagement but also fosters cultural preservation and identity (Siregar et al., 2024).

This study, therefore, seeks to address this critical gap by designing and implementing a numeracy assessment model rooted in Bugis-Makassar cultural contexts and delivered via





Computer Assisted Test (CAT). The resulting instrument is intended to be valid, reliable, adaptive, and deeply contextual, thereby advancing both educational measurement and culturally responsive pedagogy in Indonesia..

METHOD

The purpose of this study is to analyze and construct a conceptual framework for developing a numeracy assessment model grounded in local cultural contexts and delivered via Computer Assisted Test (CAT). This study serves as a preliminary phase in the broader development of a culturally contextualized, CAT-based numeracy assessment instrument. The conceptual framework is developed through a systematic, theory-driven, and rational approach (Mirriahi et al., 2015), integrating comprehensive literature reviews, critical analysis of prior research findings, and preliminary empirical investigations. The proposed development model encompasses two interrelated components: (1) the design and validation of a culturally contextualized numeracy assessment instrument, and (2) the development of a digital platform based on CAT principles to administer and adapt the assessment dynamically according to student responses.

Data were collected through direct classroom observations and semi-structured, in-depth interviews with key stakeholders, including teachers, students, and school-level policymakers. These activities aimed to identify existing gaps in numeracy assessment practices, explore contextual needs, and determine the knowledge and skill competencies required to support effective, culturally responsive learning. This study adopts the Plomp development model (Plomp & Nieveen, 2013), which comprises three iterative phases: (1) preliminary research, (2) prototyping, and (3) field testing. The preliminary research phase focused on problem identification, needs analysis, and goal formulation. Data were gathered through literature reviews, field observations, and qualitative instruments, including questionnaires and interviews. The collected data were analyzed to inform the design of the assessment model and to evaluate its potential effectiveness in addressing the identified gaps. The outcomes of this phase provide the theoretical and empirical foundation for the subsequent prototyping and validation stages, ensuring that the final CAT-based numeracy assessment instrument is not only psychometrically sound but also culturally relevant, pedagogically appropriate, and technically feasible for implementation in Indonesian educational settings.

RESULTS AND DISCUSSION

The results of observations, literature studies and interviews conducted with mathematics



teachers revealed that there is still a lack of numeracy skills that are still far from the ideal conditions expected in mathematics learning. These findings highlight the various challenges and barriers faced by teachers in their efforts to implement quality and meaningful learning.

Table 1. Gap of Ideal Conditions with Facts of Numeracy Ability

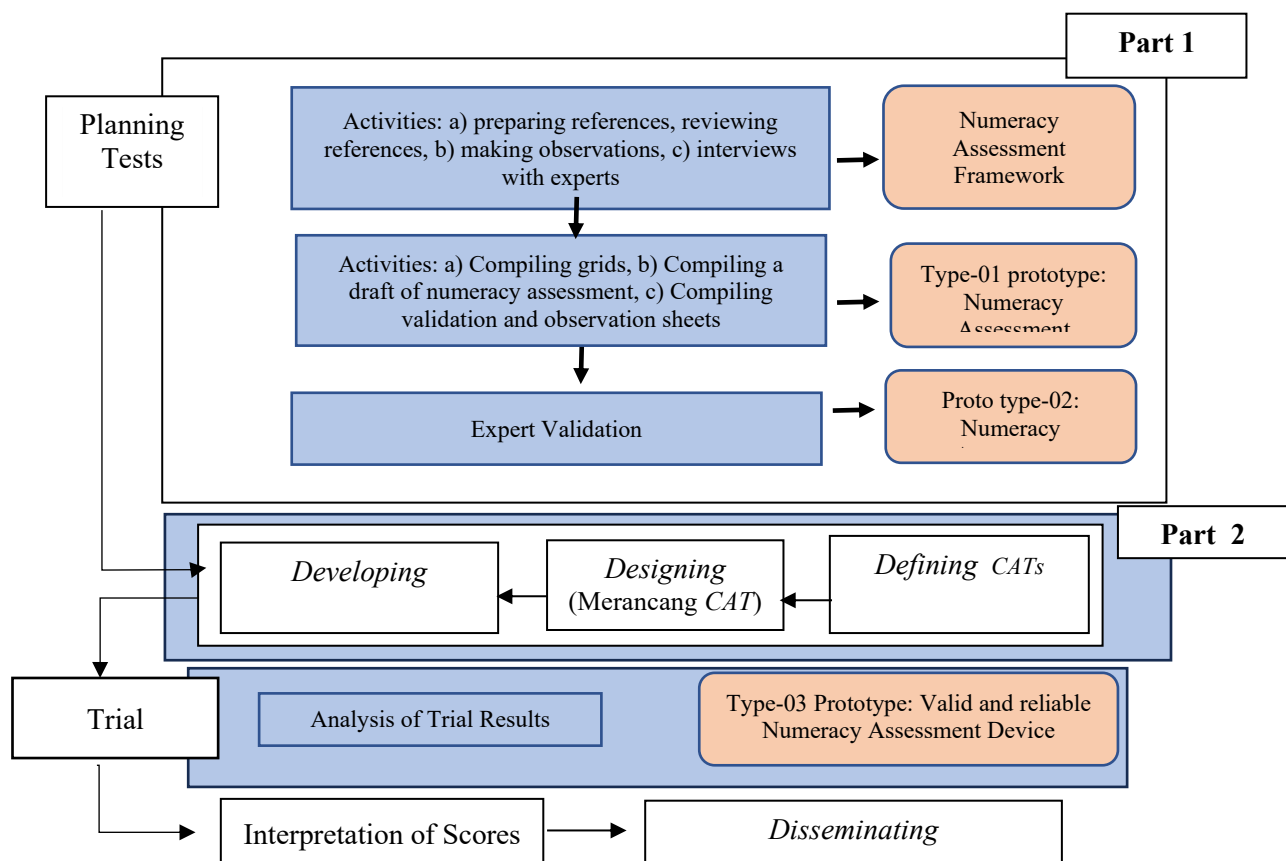
Ideal Conditions	Fact	Necessity
Students are able to apply numeracy in a real and relevant context.	Indonesia's PISA score is still below the global average (379 vs 489), indicating the low ability of numeracy applications in daily life [(OECD, 2023); (Cipta et al., 2023)].	It is necessary to develop a numeracy assessment based on the local culture of South Sulawesi so that students can understand and apply mathematics in a context that is familiar to their lives.
Students are able to solve non-routine numeracy problems that require high reasoning.	Many students have difficulty in HOTS-based numeracy problems because learning still focuses on memorization and basic operations [(Wahyuni, 2023); (Cipta et al., 2023)].	This dissertation encourages the development of numeracy instruments that train reasoning skills through contextual problems that require critical thinking.
Numeracy learning takes advantage of the local cultural context.	Ethnomathematical approaches and the use of local culture in numeracy problems have been shown to increase student engagement and understanding [(Trisnawati et al., 2022); (Siregar et al., 2024)].	The instrument in the dissertation uses the Bugis-Makassar culture as the main context to strengthen the relationship between mathematics and the daily lives of students in South Sulawesi.
Teachers are able to develop technology-based numeracy assessments (CAT: Computer Assisted Test).	The use of CAT is still limited due to a lack of training and infrastructure despite being recognized as effective for assessment efficiency	This dissertation presents a prototype of CAT-based digital numeracy assessment that is ready to be used and tested in schools in South

	[(Kurniadi et al., 2023); (Susanto et al., 2022)].	Sulawesi.
Schools use digital numeracy assessments for individualized evaluation and learning improvement.	The Class AKM has been prepared to provide individual feedback, but the implementation has not been evenly distributed [(Amalia, 2023)].	The assessment products developed in the dissertation can help educational units map students' numeracy skills digitally and personally.

Analysis of Table 1 indicates an urgent need to develop a numeracy assessment model that integrates local cultural contexts and is supported by Computer Assisted Test (CAT) technology. This need arises from a persistent gap between the expected level of students' numeracy proficiency and the actual performance observed in educational practice. While students are expected to apply mathematical concepts to real-world situations, results from both international (e.g., PISA) and national assessments (e.g., National Assessment) reveal that the majority continue to struggle with non-routine problems requiring higher-order reasoning skills. This challenge is further compounded by the limited integration of local cultural contexts into mathematics instruction and assessment. Consequently, developing culturally grounded numeracy assessments represents a critical intervention to bridge this gap. By embedding elements of Bugis-Makassar culture — which are deeply rooted in students' daily experiences — assessments can be rendered more meaningful, engaging, and identity-affirming, thereby fostering both academic competence and cultural pride.

The assessment design developed in this study provides a comprehensive, step-by-step framework for educators and schools to construct instruments that are not only psychometrically valid and reliable but also culturally responsive. The framework comprises a series of interconnected stages: (1) planning based on curriculum and cultural analysis, (2) development of culturally contextualized assessment items, (3) expert validation for content and cultural appropriateness, (4) digitization of items into a CAT-compatible platform, and (5) pilot testing to generate empirical evidence on item validity, reliability, and psychometric characteristics. Following item calibration, student performance is interpreted through a four-tiered competency scale, offering a nuanced and comprehensive profile of numeracy proficiency. In the final implementation phase, the assessment model is disseminated to partner schools to support the institutionalization of regular, contextually relevant numeracy evaluations. The ultimate output of this study is a fully operationalizable numeracy assessment framework grounded in the cultural

context of South Sulawesi and delivered via a CAT platform. This framework contributes to enhancing the quality, relevance, and adaptability of numeracy assessment in Indonesia. The structure of the development model is visually summarized in Figure 1 below.



Planning Development

The development process begins with a comprehensive planning phase, which serves as the foundational stage for all subsequent activities. During this phase, a systematic literature review was conducted to examine theoretical frameworks and evidence-based practices in numeracy assessment and the integration of local cultural contexts into mathematics education. Complementing this, field observations and semi-structured interviews were carried out with teachers, students, and educational policymakers to identify authentic needs and contextual challenges within school settings.

The synthesized findings from the literature review and empirical observations informed the development of an assessment grid that specifies three core dimensions: (1) numeracy domains (e.g., numbers, algebra, geometry, data and uncertainty), (2) cognitive demand levels



(e.g., recall, application, reasoning), and (3) culturally embedded contexts drawn from Bugis–Makassar traditions. This grid functions as the primary blueprint for item development, ensuring that the resulting assessment instrument comprehensively measures numeracy competencies while remaining contextually relevant to students’ lived experiences.

Item Development and Instrument Validation

Following the planning phase, assessment items were drafted in alignment with the established grid. The instrument integrates four key numeracy domains: numbers, algebra, data and uncertainty, and geometry and measurement. Each domain is enriched with culturally authentic contexts from South Sulawesi, including:

- Bugis weaving motifs to explore patterns, symmetry, and geometric transformations;
- Calculations involving materials for songkok recca (traditional headwear) to contextualize numerical operations and proportional reasoning;
- Scale modeling of phinisi ships to frame problems in geometry and measurement;
- Data analysis of rice harvests during the Mappadendang tradition to situate statistical reasoning and uncertainty within a meaningful cultural practice.

To capture the breadth of numeracy skills, the instrument employs a variety of item formats, including multiple-choice, complex multiple-choice, matching, true-false, and short-answer questions. This diversity enables the assessment not only of procedural fluency but also of higher-order competencies such as critical reasoning, conceptual connection-making, and interpretation of culturally embedded quantitative information. Upon completion of the initial item draft, a dual-track validation process was implemented to ensure content validity, cultural authenticity, and psychometric appropriateness. The instrument was reviewed by panels of experts in mathematics education and ethnomathematics, who evaluated:

- Alignment with numeracy assessment objectives;
- Clarity and precision of language;
- Cognitive demand and appropriateness of difficulty level;
- Depth of cultural integration — ensuring that cultural elements are not merely decorative but functionally embedded within the mathematical task.

For instance, a numeracy item contextualized within Bugis weaving must require students to perform calculations or engage in reasoning that is intrinsically tied to the structure or pattern of the weave. Similarly, an algebraic problem framed within the songkok recca trade must necessitate the application of algebraic thinking to solve a realistic economic scenario rooted in that cultural practice. Feedback from validators was systematically analyzed and used to





iteratively revise the instrument. This cyclical process continued until a refined prototype was achieved — one that demonstrated strong content validity, cultural authenticity, and readiness for empirical field testing..

Development of the Computer Assisted Test (CAT) Application

Following the finalization of the assessment instrument, the subsequent phase involves the development of a Computer Assisted Test (CAT) platform to facilitate its digital implementation. This development process adheres to the 4D instructional design model (Define, Design, Develop, Disseminate; Thiagarajan, Semmel, & Semmel, 1974), ensuring a systematic and iterative approach.

- Define Stage: A needs analysis was conducted to identify functional and technical requirements for a digital assessment system capable of delivering culturally contextualized numeracy items. Key considerations included adaptability, cultural representation fidelity, and compatibility with school infrastructure.
- Design Stage: The user interface was designed to be intuitive and accessible for both students and educators. Concurrently, adaptive algorithms were planned to dynamically assign item weights based on difficulty levels, ensuring that each student receives questions tailored to their proficiency level.
- Develop Stage: A web-based application was constructed to support multiple item formats (e.g., multiple-choice, matching, short-answer) and equipped with security features such as locked-down browser functionality to minimize academic dishonesty. The platform was optimized for efficiency, adaptability, and scalability, aligning with the demands of modern, technology-enhanced assessment in educational settings.

This CAT application not only enables efficient and individualized assessment delivery but also ensures that cultural contexts remain meaningfully embedded within the digital experience..

Digitization of Assessment Items

Validated assessment items were subsequently digitized for integration into the CAT platform. This process involved converting textual content, visual elements (e.g., diagrams, cultural artifacts), and instructional prompts into digital formats compatible with the system's architecture. Rigorous verification procedures were implemented to ensure:

- Readability and clarity of item stems and response options;





- Visual fidelity and resolution of culturally embedded images (e.g., phinisi ships, weaving patterns);
- Consistency between digital presentation and the original validated paper-based drafts.

Digitization serves as the critical bridge between traditional assessment formats and computer-based delivery. A well-executed digitization process ensures that students encounter no technical barriers during assessment, while educators retain full control over item management through a centralized digital item bank.

Instrument Trials

After digitization, the instrument underwent pilot testing to generate empirical data on its psychometric properties. The primary objectives of this phase were to evaluate:

- Construct validity: The extent to which the instrument measures the intended numeracy constructs;
- Reliability: Internal consistency and stability of scores across items and administrations;
- Item characteristics: Including difficulty index, discrimination power, and distractor effectiveness.

Data from the pilot were analyzed using classical test theory (CTT) and, where applicable, item response theory (IRT) to identify poorly performing items. These items were subsequently revised or eliminated based on empirical evidence. The iterative nature of this process ensured that only valid, reliable, and representative items were retained for the final instrument. This trial phase is essential not only for refining the instrument but also for providing empirical evidence of its effectiveness in authentic educational settings. The outcomes serve as the foundation for broader implementation and contribute to the validation of culturally grounded assessment models within digital adaptive frameworks.

Score Interpretation and Diagnostic Utility

Once the instrument has been validated and confirmed to meet psychometric standards of reliability and validity, the subsequent phase involves the interpretation of student assessment scores. Scores are classified into four competency levels:

- Requires Special Intervention — indicating significant gaps in foundational numeracy skills;
- Basic — demonstrating minimal proficiency, with limited ability to apply concepts in context;
- Capable — showing consistent ability to solve routine problems and interpret basic quantitative information;



- Proficient — reflecting advanced competence, including the ability to reason critically and solve non-routine, culturally embedded problems.

This tiered classification provides educators with a clear, actionable profile of each student's numeracy proficiency. Teachers can utilize these diagnostic insights to tailor instructional strategies— implementing targeted remediation for students at lower levels while offering enrichment and extension activities for those demonstrating advanced competence. Thus, the assessment transcends its traditional summative function and serves as a formative, diagnostic tool that actively supports personalized learning and pedagogical improvement.

Dissemination and Implementation in Educational Settings

The final phase of the development model is the dissemination of the validated CAT-based assessment instrument and its supporting digital platform to partner schools. Instruments that have successfully undergone all prior stages — planning, expert validation, digitization, and empirical pilot testing — are deployed for regular use in monitoring and evaluating students' numeracy development. This implementation is designed to support schools in conducting assessments that are not only technologically efficient and adaptive but also deeply contextualized within local cultural frameworks. Furthermore, the digital platform enables longitudinal tracking of student progress, allowing schools to generate ongoing competency maps and identify cohort- or individual-level trends over time.

By institutionalizing this culturally responsive, technology-enhanced assessment model, the study contributes meaningfully to the enhancement of numeracy education quality in Indonesia. It also reinforces the integration of local cultural identity into formal evaluation systems — transforming assessment from a mere measurement exercise into a vehicle for cultural affirmation and pedagogical innovation.

DISCUSSION

This study presents a comprehensive conceptual framework for developing numeracy assessments grounded in local cultural contexts and supported by Computer Assisted Test (CAT). The framework was developed through a systematic, multi-stage process encompassing test planning, grid design, expert validation, prototype refinement, and digital implementation, as illustrated in Figure 1. Its development follows a theory-based, logical progression that integrates four foundational perspectives: ethnomathematics (D'Ambrosio, 2001), authentic assessment (Wiggins, 1993), constructivism (Piaget, 1972; Vygotsky, 1978), and Item Response Theory (IRT/Rasch; Embretson & Reise, 2000). Thus, the study's outcomes extend beyond merely



documenting the persistent gap in Indonesian students' numeracy proficiency — as evidenced by low scores in international assessments such as PISA and national evaluations (OECD, 2019; Pusmendik, 2022) — to offering a concrete, actionable solution: a framework for developing contextual, adaptive, valid, and reliable assessment instruments.

The proposed framework comprises two interrelated components. The first focuses on the development of a culturally contextualized assessment instrument, beginning with theoretical and empirical groundwork — including literature reviews and field observations — leading to the construction of an assessment grid that aligns four core numeracy domains (numbers, algebra, geometry & measurement, data & uncertainty) with culturally authentic contexts drawn from Bugis–Makassar traditions. The output is a validated prototype instrument, wherein mathematical tasks are meaningfully embedded within cultural practices — such as Bugis weaving motifs to explore symmetry and geometric patterns, calculations of materials for songkok recca to contextualize numerical operations, scale modeling of phinisi ships to frame geometry and measurement problems, and analysis of rice harvest data in the Mappadendang tradition to situate statistical reasoning. Crucially, cultural elements are not decorative illustrations but functionally integrated into the cognitive demands of each item — consistent with ethnomathematical principles and authentic assessment. The second component centers on the development and implementation of a CAT platform, designed following the 4D model (Define, Design, Develop, Disseminate; Thiagarajan et al., 1974), which supports adaptive item delivery, multiple question formats (multiple-choice, complex multiple-choice, matching, short-answer), and secure administration through features such as locked-down browsers. The platform enables IRT-based psychometric analysis of item characteristics — including difficulty, discrimination, and distractor effectiveness — and generates diagnostic student profiles mapped onto four competency levels: Requires Special Intervention, Basic, Capable, and Proficient. Together, these components form a continuous, iterative development cycle — from conceptualization to scalable implementation — designed not only to address assessment gaps but also to provide schools with a practical, technology-enhanced tool for culturally responsive numeracy evaluation.

The theoretical foundation of this framework lies in the synergistic integration of constructivism, ethnomathematics, authentic assessment, and IRT. Constructivism informs the design of assessments that enable students to build mathematical understanding through culturally familiar experiences, making numeracy tasks more meaningful and accessible. Ethnomathematics legitimizes the use of local artifacts, symbols, and traditions as authentic sources of mathematical knowledge, ensuring that cultural elements are not superficial additions but integral to the learning process. Authentic assessment principles ensure that tasks reflect real-





life applications, thereby enhancing relevance and student engagement. Meanwhile, IRT provides the psychometric backbone for adaptive, reliable, and valid measurement through the CAT platform. This theoretical coherence ensures that every stage of the framework — from planning and validation to digitization and deployment — is both conceptually robust and empirically grounded.

This research makes significant contributions to both academic discourse and educational practice in Indonesia. In terms of content innovation, it offers the first comprehensive model for developing numeracy assessments rooted in Bugis–Makassar cultural practices, fostering deeper conceptual understanding while strengthening students’ cultural identity. Technologically, the integration of CAT enables adaptive, efficient, and scalable delivery of culturally contextualized assessments, while also functioning as a diagnostic tool that provides teachers with actionable data to inform instruction and intervention. Furthermore, the framework can serve as a reference for schools and policymakers seeking to strengthen the implementation of the National Assessment — particularly in the numeracy domain — by incorporating local cultural dimensions, thereby bridging the gap between national standards and local realities. As such, this study not only addresses an academic gap in culturally grounded numeracy assessment but also delivers a scalable, evidence-based solution with tangible implications for improving educational evaluation in Indonesia.

Despite its strengths, the framework has several limitations that warrant consideration. First, its current cultural scope is limited to the Bugis–Makassar context; generalization to other regions requires careful adaptation and validation to ensure cultural authenticity and pedagogical appropriateness across Indonesia’s diverse ethnomathematical landscape. Second, pilot testing was conducted in a limited number of partner schools; although initial psychometric results are promising, large-scale trials involving broader and more diverse student populations are needed to establish external validity and reliability. Third, practical challenges remain, including limited access to digital devices, unstable internet connectivity, and varying levels of teacher readiness to operate CAT platforms (Kurniadi et al., 2023; Susanto et al., 2022). Successful nationwide implementation will require coordinated investments in infrastructure, professional development, and policy alignment. Moreover, while the framework is complete in design, its full adoption in educational systems necessitates integration with national assessment structures, curriculum standards, and teacher training programs — areas that demand further research and stakeholder collaboration.

These limitations, however, open valuable avenues for future research. Subsequent studies should prioritize adapting and validating the framework for other Indonesian cultural





contexts — such as Javanese, Balinese, Minangkabau, or Papuan traditions — to develop a nationally representative, culturally pluralistic assessment model. Large-scale validation across multiple provinces is also essential to strengthen the empirical basis of the instrument and evaluate its effectiveness in diverse educational settings. Additionally, future research could explore mechanisms to align the framework with national assessment systems, such as the Asesmen Nasional, ensuring that locally contextualized assessments contribute meaningfully to education policy and monitoring. Finally, investigating models for teacher professional development — particularly in designing, implementing, and interpreting culturally grounded CAT-based assessments — will be critical to ensuring the sustainability and fidelity of implementation..

CONCLUSION

This study has developed a comprehensive conceptual framework for culturally grounded numeracy assessment supported by Computerized Adaptive Testing, constructed through a logical, rational, and theory-driven approach. The model emerges from a critical analysis of the persistent gap between the expected level of students' numeracy proficiency and the reality of low achievement, as evidenced by national and international assessment results. To address this gap, the framework integrates the four core domains of numeracy — numbers, algebra, geometry and measurement, and data and uncertainty — with authentic cultural contexts drawn from Bugis–Makassar traditions, thereby enhancing the relevance, authenticity, and meaningfulness of assessment items. Furthermore, the incorporation of Computerized Adaptive Testing significantly enhances the validity, reliability, and efficiency of assessment delivery through the application of Item Response Theory, enabling adaptive item presentation tailored to individual student ability. Digitization facilitates not only dynamic administration but also the generation of diagnostic profiles across four competency levels, providing educators with actionable insights to inform targeted instructional interventions. This research also affirms that embedding local culture in assessment instruments not only strengthens students' conceptual engagement with mathematics but also fosters cultural preservation and identity. Nevertheless, the study is not without limitations: its cultural scope is currently confined to the Bugis–Makassar context; empirical trials involved a limited sample size; and implementation is constrained by uneven digital infrastructure and varying levels of teacher readiness. Future research should therefore expand the model's applicability to other cultural contexts across the country, conduct large-scale validation with diverse student populations, and explore pathways for integrating such assessments into national education evaluation systems. Ultimately, this framework is positioned





as a scalable, contextually responsive strategy to sustainably enhance students' numeracy literacy in alignment with both global standards and local cultural values.

AUTHOR'S CONTRIBUTION

All authors declare that there is no conflict of interest in this article. Author 1 conducted field research and compiled data analysis. Author 2 is responsible for the overall writing of the article and the preparation of the relevant literature review. Author 3 contributes to the writing and provides a critical review of the article.

THANKS

This research did not receive financial support in the form of research grants. The author would like to express his deep gratitude to the academics who have made valuable contributions to this research process.

REFERENCES

- Abidin, Y. (2018). *Learning Design in the Context of the 2013 Curriculum*. Bandung: Refika Aditama.
- Alimuddin, A., Mulbar, U., & Nursalam, N. (2024). Asymmetric patterns correlation between mathematical literacy and reading literacy. *Al-Khawarizmi: Journal of Mathematics and Natural Sciences Education*, 12(1), 126–139.
<https://doi.org/10.24256/jpmipa.v12i1.3627>
- Anwar, Z., Hamid, A., Mahmuri, H., Liani, A. M., & Nurmala, R. (2023, December). Analysis of Mathematical Critical Thinking Skills Through a Semiotic Approach. In *5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023)* (pp. 60-69). Atlantis Press.
- Ayuningtyas, D., & Setiana, L. (2019). Development of ethnomathematics-based mathematics teaching materials. *Journal of Mathematics Education*, 13(2), 101–112.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575.
<https://doi.org/10.1080/0969594X.2018.1441807>
- D'Ambrosio, U. (2001). *Ethnomathematics: link between traditions and modernity*. Belo Horizonte: Autentica.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Jabali, M., Rahayu, S., & Lestari, N. (2020). Mathematics learning media based on local culture to



- improve student learning outcomes. *Journal of Mathematics Education Research*, 4(1), 45–57.
- Ministry of Education and Culture. (2017). *National Literacy Strategy: National Literacy Movement*. Jakarta: Ministry of Education and Culture.
- Ministry of Education and Culture. (2021). *National Assessment as an Evaluation of Education Quality*. Jakarta: Ministry of Education, Culture, Research, and Technology.
- Kurniadi, A., Rahman, F., & Susilawati, L. (2023). Implementation of CAT-based digital assessments in secondary schools. *Journal of Educational Technology*, 25(1), 33–45.
- Morocco, N. in Abidin, Y. (2018). *Learning Design in the Context of the 2013 Curriculum*. Bandung: Refika Aditama.
- Nurmala, R., Mulbar, U., & Darwis, M. (2025). *Panduan desain pembelajaran in action pada pembelajaran matematika di sekolah menengah pertama*. CV. Ruang Tentor.
- OECD. (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris: OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- Piaget, J. (1972). *The Psychology of the Child*. New York: Basic Books.
- Center for Educational Assessment (Pusmendik). (2022). *Report on the Results of the 2022 National Assessment*. Jakarta: IAARD and Books, Ministry of Education and Culture.
- Rahman, A., Yusuf, I., Manda, D., & Barlian, E. (2018). The development of Android and web-based logical thinking measurement tools. *Journal of Physics: Conference Series*, 1028(1), 012168. IOP Publishing. <https://doi.org/10.1088/1742-6596/1028/1/012168>
- Rizal, M., Andriani, D., & Fauzi, A. (2021). Numeracy module based on local wisdom: Effectiveness and practicality. *Journal of Mathematics Education*, 15(1), 77–89.
- Schleicher, A. (2019). *PISA 2018: Insights and Interpretations*. Paris: OECD Publishing.
- Siregar, A., Hidayat, R., & Lestari, N. (2024). Development of numeracy questions based on cultural context. *Journal of Mathematics Education Research*, 8(1), 55–70.
- Susanto, H., Widodo, A., & Nurhaliza, R. (2022). The effectiveness of Computer Assisted Test in improving assessment efficiency. *Journal of Educational Evaluation*, 10(2), 99–110.
- Syahid, N. F., & Mattoliang, L. A. (2024, February). The Effect of Problem-Based Learning Model with Realistic Mathematics Education Approach on Mathematical Communication Skills. In *Proceeding of International Conference on Multidisciplinary Research* (Vol. 6, No. 1, pp. 57-64).
- Trilling, B., & Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times*. San Francisco: Jossey-Bass.
- Trisnawati, I., Hidayat, R., & Lestari, N. (2022). Integration of ethnomathematics in numeracy



- learning. *Journal of Mathematics Education*, 14(2), 121–134.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Wiggins, G. (1993). *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco: Jossey-Bass.
- World Bank. (2020). *Indonesia's PISA 2018 Results: Brief*. Washington, DC: World Bank Group.

