JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

The Test of Minimum Numeracy Competency for Level 2 Using an Online Quizizz Platform: Instrument Development

Fahrul Islam[1], Syahrul[2], Patahuddin[3]

[1, 2, 3]Department of Educational Research and Evaluation, Universitas Negeri Makassar, Indonesia
*Email: Fahrulislam932@gmail.com

ARTICLE INFO

Keywords:
Numeracy
Assessment
Online Quizizz
Instrument development
Elementary students

ABSTRACT

Purpose – This study aims to develop and test the reliability of the Level 2 Numeracy AKM instrument presented through an online quiz platform. The development of this instrument aims to address the need for a practical, easily accessible digital assessment tool that can enhance student engagement, particularly in efforts to strengthen the numeracy skills of elementary school students in Indonesia.

Methodology - This study employs a Research and Development (R&D) approach, adapting the ADDIE model and Retnawati's instrument development framework (analysis, design, development, user trials, and evaluation). The instrument, comprising 20 items, was designed to cover key numeracy indicators. Data were collected from 38 fourth-grade students across four elementary schools and two expert validators. The analysis included content validity assessment using Gregory's model, media feasibility analysis, reliability testing with McDonald's Omega, and a descriptive analysis of student performance and feedback.

Findings – The research confirmed the high quality of the developed instrument. The developed instrument demonstrates very high content validity, as indicated by a Cronbach's alpha coefficient of 1.00 and an average validator rating of 3.45 on a scale of 4. Its reliability level is also strong, as evidenced by a McDonald's Omega value of 0.81, reflecting good internal consistency. Student responses were also very positive. The practicality score indicating that the instrument is considered engaging, easy to use, and convenient to access. In the field trial, students were able to solve fundamental problems. However, most still had difficulties when dealing with more complex material, especially fraction comparisons and algebraic concepts.

Contribution –Overall, this study makes an important contribution in the form of a practical and validated model for developing digital-based numeracy assessment instruments that align with national AKM standards and international evaluation frameworks, such as PISA.

Received 29 July 2025; Received in revised form 12 August 2025; Accepted 15 December 2025
Jurnal Eduscience (JES) Volume 12 No. 6 (2025)
Available online 30 December 2025

INTRODUCTION

A fundamental shift toward competency-based learning and digital integration marks the global landscape of 21st-century education. This instrument provides teachers with an effective means of identifying gaps in students' numeracy skills, allowing learning to be focused and tailored to the principles of age-appropriate learning development. This research also offers an instrument development model that can be applied more broadly to produce modern assessment tools that not only meet psychometric standards but also provide an enjoyable and motivating digital learning experience. Overall, this study makes an important contribution in the form of a practical and validated model for developing digital-based numeracy assessment instruments that align with national AKM standards and international evaluation frameworks, such as PISA.

In this context, Indonesia continues to face various challenges in mathematics education. The results of PISA (OECD, 2023) and TIMSS (Hadi & Novaliyosi, 2019) indicate that many students have not reached an adequate level of numeracy proficiency. In fact, 82% of students in PISA 2022 did not achieve Level 2 in mathematics, which indicates difficulties in applying basic concepts to context-based problems. This situation reflects long-standing pedagogical issues in classrooms, where teaching has focused more on procedural skills rather than conceptual understanding or problem-solving abilities (Perdana & Suswandari, 2021; Maulidia et al., 2023).

Theoretically, this study is grounded in Self-Determination Theory (SDT) (Deci & Ryan, 2000) and the Technology Acceptance Model (TAM) (Davis, 1989), which together explain how student motivation and acceptance are formed in the context of digital assessments incorporating gamification elements. SDT emphasizes that learning engagement increases when students feel autonomy, can demonstrate competence, and feel connected to their learning environment—all aspects that can be supported through gamification features such as instant feedback, progress tracking, and reward systems. Meanwhile, TAM states that students' perceptions of the usefulness and ease of use of a platform greatly influence their willingness to utilize it in learning. By integrating these two approaches, the Quizizz-based formative assessment in this study was designed not only to measure numeracy competence but also to increase student motivation and engagement through an interactive and enjoyable digital experience. In response, the Ministry of Education, Culture, Research, and Technology of Indonesia introduced the National Assessment (*Asesmen Nasional/AN*) framework, which includes the Minimum Competency Assessment *(Asesmen Kompetensi Minimum/AKM)* as a diagnostic tool for evaluating literacy and numeracy. Class-level AKM, as an adaptation at the educational unit level, was developed to help teachers map student learning gaps and implement differentiated learning according to the teaching at the Right Level (TaRL) approach (Asrijanty, 2021; Pusmenjar, 2020). However, despite this reform being progressive, various challenges in its implementation are still encountered. Many schools still rely heavily on rote learning practices reminiscent of the former National Examination (UN), and digital readiness across regions remains uneven (Amanda & Nurjannah, 2022). In Soppeng Regency, South Sulawesi, for example, initial findings indicate that students tend to answer AKM-type questions mechanically without truly understanding them. At the same time, teachers are still constrained by limited digital infrastructure.

These issues illustrate a gap between national policy aspirations and classroom realities. To date, no digital numeracy assessment model integrating gamification and formative approaches has been systematically developed for the Indonesian elementary context. However, the combination of these two elements holds potential to increase motivation, engagement, and the authenticity of assessment in capturing students' numeracy skills. Hence, the development of a technology-based assessment model suited to Indonesian primary schools is urgently needed.

To address this gap, the present study aims to design, develop, and validate a Level 2 Numeracy AKM instrument based on the *Quizizz* platform for fourth-grade students. Internationally, Quizizz has been shown to enhance student engagement, motivation, and cognitive abilities through its interactive displays and immediate feedback (Zhao, 2019; Li et al., 2023). The instrument developed in this study is expected to be a

valid, reliable, and practical tool for teachers to assess numeracy skills, close learning gaps, and encourage student engagement in mathematical problem solving. Thus, this study strengthens digital formative assessment practices in line with national AKM policies while enriching the global discourse on technology-assisted competency-based learning.

METHODOLOGY

Research Design

This study applied the ADDIE Model as the main framework, with Retnawati's (2016) procedure integrated into the development and evaluation stages to ensure psychometric accuracy. In the Analysis stage, AKM numeracy indicators and content domains were determined based on the national assessment framework. The Design phase then produced a test blueprint and item specifications. In the Development stage, draft items were compiled and validated by experts in accordance with Retnawati's guidelines to ensure content validity. Next, the Implementation stage was carried out through instrument testing using the Quizizz platform on elementary school students.

Finally, the Evaluation stage consisted of item analysis, including the difficulty index, discrimination index, and reliability, following Retnawati's criteria. The integration of the ADDIE model and Retnawati's framework ensured that the instrument development process was both pedagogically systematic and psychometrically robust.

Participant

The content validation in this study involved two expert validators, namely a mathematics education lecturer and an experienced elementary school teacher. Although the number of validators was limited, it was considered sufficient because both experts had strong academic qualifications and professional experience directly relevant to AKM numeracy constructs and technology-based assessment development. The validation process is conducted systematically using structured assessment sheets with measurable indicators to minimize subjectivity. Bias is minimized through discussion and revision of items based on the agreement of the validators. Retnawati (2016) explains that at least two competent experts can carry out content validation, and Sugiyono (2019) emphasizes that the quality and relevance of the validators' expertise are more important than their number. Therefore, the involvement of two validators is considered sufficient to ensure the suitability of the instrument content. The second group consisted of the subjects in the field trial: 38 fourth-grade students from four public elementary schools located in the Citta District, Soppeng Regency.

Table 1. Distribution of the Schools

No	School Name	Number of Students (n)
1	SD Negeri 259 Tanacellae	8
2	SD Negeri 93 Tanete	12
3	SD Negeri 99 Labae	12
4	SD Negeri 269 Allangkiangeng	6
	Total	38

The data collection process was executed in two distinct phases to gather both theoretical and empirical evidence. The first phase involved expert validation. The complete instrument package, including the test blueprint, all 20 items, and scoring rubrics, was submitted to the two expert validators. They conducted a thorough review and provided both quantitative ratings and qualitative written feedback on the content and media implementation.

The second phase was the empirical field trial, conducted in May 2025. The 38 student participants were guided to access the AKM test on the Quizizz platform using school computers or available devices. They completed the assessment independently under standard classroom conditions. Immediately upon finishing

the test, they were asked to complete the student response questionnaire to capture their immediate perceptions and experiences with the digital assessment tool.

Instrument

The central instrument of this research was the newly developed Level 2 Numeracy AKM test, delivered through the Quizizz online platform. The test consisted of 20 items carefully crafted to measure a spectrum of competencies across the four official numeracy domains: Bilangan (Numbers), Geometri dan Pengukuran (Geometry and Measurement), Aljabar (Algebra), and Data dan Ketidakpastian (Data and Uncertainty). To mirror the complexity and format of the actual AKM, the test incorporated five different question types: standard multiple-choice (7 items), complex multiple-choice where students could select more than one correct answer (5 items), matching (3 items), short-answer (3 items), and open-ended essay/explanation (2 items). A suite of supporting instruments was utilized to collect validation and feedback data:

Content and Media Validation Sheet

This instrument is a structured assessment rubric given to expert validators. They use a 4-point scale to assess the validity of each item's content, including its suitability to the competency, the accuracy of the construct in measuring the targeted skill, and the clarity of language use. Additionally, a special section is provided for evaluating media aspects, including interface quality, interactivity level, and technical performance of the instrument when used on the Quizizz platform.

Student Response Questionnaire

This questionnaire consists of 10 items that students complete after the field trial. Using a 4-point Likert scale (1 = Strongly Disagree to 4 = Strongly Agree), this questionnaire is designed to assess the practicality of the instrument from the user's point of view, covering the clarity of the questions, the level of difficulty, the level of engagement, and their experience while using the Quizizz-based instrument.

Data Analysis

Content Validity

In addition to reporting the coefficients, this study also interprets the strength of validity and reliability. Following Guilford's (1956) guideline, commonly applied in educational research (see also Retnawati, 2016), validity coefficients ranging from 0.80–1.00 are categorized as very strong, 0.60–0.79 as strong, 0.40–0.59 as moderate, 0.20–0.39 as weak, and 0.00–0.19 as very weak. Accordingly, items with validity indices of ≥ 0.40 are considered adequate for use, whereas items below 0.20 are usually revised or discarded. A similar principle applies to reliability, where a higher Cronbach's Alpha value (closer to 1.00) indicates stronger internal consistency, and a reliability value of 0.70 or higher is generally regarded as acceptable.

In this study, the coefficients obtained were not only presented numerically but also interpreted in terms of their categories. According to Guilford (1956) and Retnawati (2016), coefficients between 0.80–1.00 are categorized as very strong, 0.60–0.79 as strong, 0.40–0.59 as moderate, 0.20–0.39 as weak, and 0.00–0.19 as very weak. Referring to these criteria, items with a validity coefficient ≥ 0.40 are considered suitable for use, while values below 0.20 are usually revised or eliminated. A similar principle applies to reliability, where the closer the number is to 1.00, the better the internal consistency. In educational research, a reliability coefficient of 0.70 or higher is typically considered to meet the minimum standard for reliability.

Media Feasibility

The scores from the media validation sheet were aggregated and converted into a percentage to provide an overall feasibility score, which was then interpreted against a standardized rubric (e.g., <53% = Not Feasible, >85% = Very Feasible).

Reliability

The internal consistency of the 20-item test was determined by calculating the McDonald's Omega (omega) coefficient. This advanced statistical measure was selected over Cronbach's Alpha because it provides a more accurate estimate of reliability for tests containing a mix of dichotomous and polytomous items. An omega value above 0.80 is indicative of high reliability.

Student Performance and Practicality

The raw scores from the AKM test were analyzed descriptively to compute the mean score for each item and for each of the four numeracy domains. This enabled the creation of a detailed profile of the student's strengths and weaknesses. Concurrently, the data from the student response questionnaire was analyzed to calculate the mean score for each statement, providing a quantitative measure of the instrument's practicality and user-friendliness.

FINDINGS

The developed Level 2 Numeracy Minimum Competency Assessment (AKM) instrument demonstrated strong psychometric and practical quality.

Validity and Reliability of the Instrument

Expert validation using the Gregory model produced a maximum coefficient (1.00), indicating complete agreement among validators regarding the relevance and clarity of all 20 items. The average assessment score of 3.45, which falls into the highly valid category, further confirms that the items developed are in line with the AKM numeracy indicators and the established learning objectives. This high validity reflects that the instrument successfully operationalizes the intended constructs—numerical reasoning, geometry, algebraic thinking, and data interpretation—without content bias.

The internal consistency analysis using McDonald's Omega yielded a coefficient of 0.81, exceeding the 0.80 standard. This result indicates strong internal consistency and reliability in measuring numeracy competencies across items with mixed formats.

Student Competency Profile and Interpretation.

The average student numeracy score was 75.39, showing adequate performance. Students demonstrated strengths in understanding whole numbers, performing basic operations, and interpreting simple data. However, consistent weaknesses were apparent in the ability to compare fractions, recognize algebraic patterns, and visualize three-dimensional shapes. Further analysis showed that these obstacles were not only related to numerical ability but also to cognitive aspects and contextual understanding. Many students still rely on memorizing procedures rather than conceptual reasoning, which limits their ability to apply thinking strategies when encountering contexts that differ from what they are accustomed to. Additionally, in some items, the complexity of the language—for example, multi-step or comparison-based instructions—is believed to have impacted students' understanding. Additionally, cultural and contextual unfamiliarity—such as tasks involving less familiar everyday situations—may have contributed to reduced accuracy in problem-solving.

These findings underscore the importance of teachers integrating explicit instruction in problem representation, language scaffolding, and contextualized learning experiences. The instrument thus serves not only as an assessment tool but also as a diagnostic guide for designing targeted interventions that strengthen students' reasoning and comprehension skills.

Practicality and Student Reception

From a pedagogical perspective, the findings of this study emphasize the importance of using explicit teaching strategies—such as comparing fractions with visual models—as well as simplifying the language in story problems and utilizing interactive media to bridge abstract reasoning to concrete reasoning. Thus, this instrument not only describes students' current abilities but also helps identify learning focuses that can be used as a reference for teachers in designing more targeted interventions.

Overall, the developed instrument proved to be valid, reliable, practical, and has substantial diagnostic value, providing teachers with an effective formative assessment tool to map learning gaps and support evidence-based teaching practices.

The results of the student questionnaire also showed positive responses, confirming the practicality and high level of engagement in using this instrument. The overall average score reached 3.54 out of 4, with the visual aspects and student enjoyment receiving the highest scores (3.8). Meanwhile, statements regarding the level of difficulty received the lowest score (3.0). Still, they fell within the "agree" category, indicating that the level of challenge was appropriate for the majority of students.

Table 2: Details of Expert Validation Scores for Each Question Item

No.	Domain	Competency	Validator 1	Validator 2	Average	Category
1	Numbers	Knowing the position of fractions on a number line	3	3	3.00	Valid
2	Numbers	Knowing the position of fractions on a number line	3	3	3.00	Valid
3	Numbers	Understanding whole numbers	4	4	4.00	Very Valid
4	Numbers	Comparing fractions and whole numbers	4	4	4.00	Very Valid
5	Numbers	Operations with whole numbers	4	4	4.00	Very Valid
6	Numbers	Understanding mixed fractions	4	4	4.00	Very Valid
7	Numbers	Understanding whole numbers	4	4	4.00	Very Valid
8	Geometry	Characteristics of 2D shapes	4	4	4.00	Very Valid
9	Geometry	Characteristics of 3D shapes	3	3	3.00	Valid
10	Geometry	Standard units	3	3	3.00	Valid
11	Geometry	Using directions/maps	3	3	3.00	Valid
12	Algebra	Solving simple equations	3	3	3.00	Valid
13	Algebra	Number patterns	3	3	3.00	Valid
14	Algebra	Number patterns	4	4	4.00	Very Valid
15	Data & Uncertainty	Determining more likely events	4	4	4.00	Very Valid
16	Data & Uncertainty	Determining more likely events	4	4	4.00	Very Valid
17	Data & Uncertainty	Simple data presentation	3	3	3.00	Valid
18	Data & Uncertainty	Simple data presentation	3	3	3.00	Valid
19	Data & Uncertainty	Simple data presentation	3	3	3.00	Valid
20	Data & Uncertainty	Simple data presentation	3	3	3.00	Valid

The students' responses to the use of the Quizizz-based numeracy assessment instrument showed a very positive reaction across almost all evaluation aspects. The overall average score ranged from 3.0 to 3.8, indicating that all indicators fell within the "agree" to "strongly agree" categories.

In general, students found the content and language of the questions easy to understand (average score 3.6), indicating that the wording and instructions in the test were suitable for the language development level of elementary school students. The readability and clarity of the question context became one of the key factors contributing to the instrument's success. Furthermore, the statement regarding the ease of understanding the wording of the questions (score 3.6) reinforced the finding that the use of communicative and straightforward language helped students avoid mistakes caused by ambiguous sentences.

In terms of visual presentation, students gave the quiz's appearance and design an excellent rating, with a score of 3.8, emphasizing that the gamification elements and attractive visuals on the Quizizz platform played a significant role in increasing learning engagement. Students also reported feeling happy and motivated while taking the quiz (score 3.8 for enjoyment and 3.6 for motivation to learn mathematics). This illustrates that integrating game-like features—such as instant scoring, leaderboards, and appealing colors—can transform students' perception of assessment from a tense activity into an enjoyable and stimulating experience.

From a technical perspective, indicators related to ease of access and use of quizzes received a score of 3.6, while application stability without technical disruptions scored 3.5. These findings suggest that the majority of students can effectively utilize the platform, although a small number may still encounter challenges, such as connection issues or device limitations.

Interestingly, the lowest score (3.0) appeared for the statement, "The questions are not too hard and not too easy, just right for me." Although still within the "agree" category, this indicates that some students perceived a level of variation in the challenge. Therefore, the instrument can be considered to have a moderate difficulty level, which is ideal for diagnostic assessments as it helps differentiate between varying levels of student ability.

Overall, the interpretation of Table 2 confirms that the Quizizz-based Numeracy Assessment instrument is efficient, engaging, and easy to use for elementary school students. The combination of communicative language, interactive visuals, and questions relevant to the learning context makes this assessment not only a tool for measuring student achievement but also a learning tool that encourages them to think critically, become more involved, and actively build understanding in the mathematics learning process.

Table 3: Details of the Students' Average Response Scores

No.	Statement	Average	Category
1	I can easily understand the questions in this quiz.	3.6	Strongly Agree
2	The words in the questions are easy to understand and not confusing.	3.6	Strongly Agree
3	The images or tables in the questions help me understand better.	3.3	Agree
4	The questions are neither too hard nor too easy. Just right for me!	3.0	Agree
5	I was motivated to try to answer all the questions in this quiz.	3.6	Strongly Agree
6	The quiz display is visually appealing and well-designed.	3.8	Strongly Agree
7	I could easily open and run this quiz.	3.6	Strongly Agree
8	The quiz ran smoothly, without frequent lags or errors.	3.5	Strongly Agree
9	I felt happy while doing this quiz.	3.8	Strongly Agree
10	I am more motivated to learn mathematics through this online quiz.	3.6	Strongly Agree

Table 3 shows that students gave very positive responses toward the Quizizz-based numeracy assessment, with average scores ranging from 3.0 to 3.8 (categories "Agree" to "Agree Strongly"). Most students found the questions straightforward and easy to understand (3.6), supported by helpful visuals (3.3). The quiz's appearance and design received the highest score (3.8), indicating strong engagement and enjoyment during the activity. The difficulty level (3.0) was considered appropriate, showing that the items

were well-balanced for diagnostic use. Technically, students found the quiz easy to access (3.6) and smooth to operate (3.5). Overall, these results confirm that the instrument is practical, user-friendly, and motivating, effectively combining assessment and learning in an engaging digital format.

Student Numeracy Profile

Field trials provide a reasonably comprehensive picture of the numeracy skills of fourth-grade students. The overall average score of 75.39 indicates adequate general competence. However, analysis based on domains and specific competencies provides a more detailed understanding of students' strengths and variations in performance.

Areas of Strength

Students appear to excel in basic and procedural skills. The best performance was observed in the Numbers domain, particularly in understanding integers (average score of 86.84) and completing multi-digit number operations (average score of 84.21). In the Geometry and Measurement domain, students also demonstrated strong abilities in identifying the characteristics of 2D shapes (average 84.21). They were also proficient in extracting information from simple infographics within the Data dan Ketidakpastian (Data) domain.

Areas for Improvement

The assessment clearly identified that students struggled with tasks requiring higher-order thinking and the application of concepts in non-routine contexts. The most significant challenges arise in the Numbers domain, particularly when students are asked to compare fractions (60.53), and in the Algebra domain when identifying visual number patterns (63.16). Other difficulties were observed in Geometry and Measurement, specifically in interpreting 3D shapes from 2D nets (65.79%) and calculating the surface area of 3D objects (68.42%). These findings suggest that, although students are pretty strong in routine calculations, they still encounter challenges in applying concepts to non-standard situations that require conceptual reasoning and a deeper understanding.

Table 4: Details of Students' AKM Performance Based on Domain and Competency

Item	Question Summary	Domain	Competency	Score	Category
1	Determining the heaviest chili paste ingredient	Numbers	Knowing the position of fractions on a number line	78.95	Good
2	Calculating the average speed of a motorcycle	Numbers	Knowing the position of fractions on a number line	76.32	Good
3	Matching numbers of equal value	Numbers	Understanding whole numbers	86.84	Very Good
4	Determining the correct statement about watermelon division	Numbers	Comparing two fractions	60.53	Fair
5	Determining the correct statement about soccer tournament results	Numbers	Operations with whole numbers	84.21	Very Good
6	Identifying the picture showing the siblings' share of chocolate	Numbers	Understanding fractions and mixed fractions	71.05	Good
7	Calculating the result of 53,323 - 687 - 8,897	Numbers	Understanding whole numbers	71.05	Good

Item	Question Summary	Domain	Competency	Score	Category
8	Determining the correct statement about a school sketch	Geometry and Measurement	Identifying characteristics of 2D shapes	84.21	Very Good
9	Identifying a 3D shape from its net	Geometry and Measurement	Identifying characteristics of 3D shapes	65.79	Fair
10	Calculating the area of aquarium parts	Geometry and Measurement	Recognizing standard units	68.42	Fair
11	Calculating distance on a map	Geometry and Measurement	Using directions/maps	81.58	Very Good
12	Calculating the number of seats in a stadium tribune	Algebra	Solving simple equations	78.95	Good
13	Determining the bracelet sequence that fits the pattern	Algebra	Recognizing simple number patterns	63.16	Fair
14	Calculating the number of laundry fragrance bottles	Algebra	Recognizing simple number patterns	81.58	Very Good
15	Determining the probability of colored squares on a game board	Data and Uncertainty	Determining more likely events	71.05	Good
16	Calculating the expected frequency of drawing a green ball	Data and Uncertainty	Determining more likely events	73.68	Good
17	Determining the correct statement from a tourism infographic	Data and Uncertainty	Understanding simple data presentation	76.32	Good
18	Calculating the number of tourists from Malaysia	Data and Uncertainty	Understanding simple data presentation	84.21	Very Good
19	Determining the correct bar chart from an infographic	Data and Uncertainty	Understanding simple data presentation	71.05	Good
20	Determining if tourist visits increased or decreased	Data and Uncertainty	Understanding simple data presentation	78.95	Good

DISCUSSION

The high validity and reliability achieved by the Quizizz-based formative assessment tool demonstrate that this instrument possesses strong psychometric reliability and theoretical suitability in relation to the broader concept of numeracy. In line with AKM and international frameworks, such as PISA (OECD, 2019), numeracy encompasses not only computational skills but also reasoning, problem-solving, and representation. The instrument developed in this study successfully captures these multidimensional aspects, making it consistent with national and global standards. However, good statistical validity and reliability do

not guarantee comprehensive coverage of the construct, as some elements, such as contextual reasoning and linguistic interpretation, may not be optimally represented yet.

The results show that although students have mastered basic procedures well, they still experience difficulties in higher-order reasoning, especially in fractions, algebraic patterns, and three-dimensional visualization. These findings align with previous studies that reveal Indonesian students tend to rely on procedural memorization rather than conceptual reasoning when faced with non-routine problems (Perdana & Suswandari, 2021; Maulidia et al., 2023), and that factors such as language complexity and unfamiliar contexts also impede understanding. Furthermore, the results of this study are consistent with international evidence on the effectiveness of gamification-based assessment, which suggests that gamification can increase student motivation and engagement, although its benefits are highly dependent on the appropriate pedagogical integration (Li et al., 2023; Díaz, 2024). Similarly, digital formative assessment research highlights the importance of real-time feedback and learning analytics in facilitating adaptive instruction (Zhang et al., 2023; Banihashem et al., 2022). Positioned within this body of literature, the present study offers novel insight by integrating formative assessment principles with gamification within the AKM numeracy framework for Indonesian elementary schools—an area still underexplored in existing research (Istiqlal et al., 2024; Safitri, 2024a, 2024b; Safitri et al., 2024, 2025; Safitri & Ansyari, 2024).

From a pedagogical perspective, the developed instrument can function as both an assessment and diagnostic tool. It helps teachers identify specific misconceptions—such as fraction comparison or recognizing algebraic patterns—and provides immediate feedback that supports differentiated instruction. Thus, the instrument bridges the gap between policy-level expectations and classroom realities, enabling the practical implementation of competency-based learning through engaging digital platforms.

CONCLUSION

This study makes an important contribution to the development of digital formative assessment by demonstrating that Quizizz-based tools can be effectively designed, developed, and validated to measure AKM numeracy competencies at the elementary school level. In addition to establishing strong psychometric characteristics, these findings confirm that gamified digital assessments are not only appealing to students but also serve as a formative strategy that aligns with national policies (AKM) and global frameworks such as PISA, which emphasize problem-solving and higher-order thinking. This integration underscores the practical relevance of gamification platforms in supporting competency-based assessment, particularly in developing countries.

However, this study has several limitations that need to be considered. The relatively small sample size and limitation to four schools in one district may limit the generalizability of the findings. Although content validation involving only two expert validators is adequate for the initial stage, it has the potential to introduce subjectivity. Additionally, the instrument only covers some of the numeracy indicators, thus not fully representing the multidimensional nature of numeracy as defined in the international framework. Recognizing these limitations helps to ensure that the interpretation of the results remains proportionate. Future research needs to expand the scope of the pilot to a larger and more diverse population, involve more expert validators, and consider applying a more adaptive item difficulty model to capture student ability variation more accurately. The next direction of development could also explore integrating this gamified assessment into national-scale assessments or linking it to international standards such as PISA, thereby increasing the resilience, scalability, and global relevance of game-based formative assessment tools to support numeracy learning.

This study contributes to the development of educational assessment by providing a valid, reliable, and user-friendly digital instrument to help teachers identify learning gaps and design learning experiences that meet students' needs. In addition to its psychometric strengths, this instrument represents a pedagogical innovation that bridges national assessment policies with classroom practices, demonstrating that gamification-based formative assessment can enhance student motivation and understanding of mathematics.

REFERENCES

Admadja, I. P., & Marpanaji, E. (2016). Pengembangan multimedia pembelajaran praktik individu instrumen pokok dasar siswa SMK di bidang keahlian Karawitan. *Jurnal Pendidikan Vokasi, 6*(2), 173. https://doi.org/10.21831/jpv.v6i2.8107

Agnesa, O. S., Hanidar, E., Afifi, N., & Rahmadana, A. (2023). Pelatihan pembelajaran literasi sains menggunakan keterampilan berpikir tingkat tinggi sebagai penyesuaian terhadap instrumen asesmen kompetensi madrasah. *I-Com: Indonesian Community Journal, 3*(1), 88–102.

Agustin, N., Annisa Ianna, Y., H., A., G., R., T., & R., R. A. (2023). *Media digital untuk pembelajaran*. Penerbit Cahya Ghani Recovery.

Ahmadi, F., & Ibda, H. (2020). *Konsep dan aplikasi literasi baru di era revolusi industri 4.0*. CV Pilar Nusantara.

Al Mawaddah, A. W., Hidayat, M. T., Amin, S. M., & Hartatik, S. (2021). Pengaruh penggunaan media pembelajaran Quizizz terhadap hasil belajar siswa pada mata pelajaran matematika melalui daring di sekolah dasar. *Journal of Basic Education, 5*(5), 3109–3116. https://doi.org/10.31004/basicedu.v5i5.1288

Amanda, N. A. J., & Nurjannah. (2022). Analisis asesmen dan intervensi pelaksanaan ANBK dalam meningkatkan mutu pendidikan internal di SD Negeri 016 Loa Kulu. *Edukasi Tematik: Jurnal Pendidikan Sekolah Dasar, 3*(1), 41–44.

Ambarwati, D., & Kurniasih, M. D. (2021). Pengaruh problem based learning berbantuan media YouTube terhadap kemampuan literasi numerasi siswa. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 5*(3), 2857–2868. https://doi.org/10.31004/cendekia.v5i3.829

Anwari, M. R., Yunus, M., Istiqamah, I., Syakir, A., & Hamidah, J. (2021). Pelatihan PKB berbasis peserta didik bagi guru Bahasa Indonesia SMA. *Jurnal Abdidas, 2*(3), 627–633. https://doi.org/10.31004/abdidas.v2i3.334

Aprinastuti, C. (2023). *Special book for media tutorial on ICT-based learning*. Stile BOOK.

Ardiansyah, M. (2022). Efektivitas penggunaan platform Quizizz dalam meningkatkan minat dan pemahaman konsep matematika. *SAP (Susunan Artikel Pendidikan), 6*(3). http://dx.doi.org/10.30998/sap.v6i3.9892

Arkadiantika, I., Ramansyah, W., Effindi, M. A., & Dellia, P. (2020). Pengembangan media pembelajaran virtual reality pada materi pengenalan termination dan splicing fiber optic. *Jurnal Dimensi Pendidikan dan Pembelajaran, 8*(1), 29. https://doi.org/10.24269/dpp.v0i0.2298

Aulia, M. P., & Mutaqin, A. (2022). Pengembangan instrumen numerasi pada konteks pertanian untuk siswa SMP. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 6*(3), 2454–2466. https://doi.org/10.31004/cendekia.v6i3.1562

Budiutomo, T. W. (2015). Peningkatan mutu pendidikan melalui penilaian proses belajar mengajar. *Academy of Education Journal, 6*(1), 52–64. https://doi.org/10.47200/aoej.v6i1.125

Bury, B. (2017). Testing Goes Mobile – Web 2.0 Formative Assessment Tools. *Conference Proceedings: ICT for Language Learning*. Pixel.

Chesa, N., & Nafi'ah, B. A. (2022). Asesmen kompetensi minimum (AKM) kelas sekolah dasar sebagai sarana evaluasi kemampuan literasi dan numerasi siswa. *Jurnal Pendidikan Dasar, 13*(2), 67–86.

Hadi, S., & Novaliyosi. (2019). *TIMSS Indonesia (Trends in International Mathematics and Science Study)*. Prosiding Seminar Nasional & Call For Papers, 562–569.

Istiqlal, M., Istiyono, E., Widihastuti, Sari, D. K., Danni, R., & Safitri, I. (2024). Construction of Mathematics Cognitive Test Instrument of Computational Thinking Model for Madrasah Aliyah Students. *Nazhruna: Jurnal Pendidikan Islam*, *7*(2), 475–492. https://doi.org/10.31538/nzh.v7i2.4425

Khomariah, K., Roesdiana, L., & Wijayanto, Z. (2022). Analisis kemampuan literasi numerasi siswa dalam menyelesaikan soal asesmen kompetensi minimum (AKM). *Jurnal Riset Pendidikan Matematika, 2*(1), 28. https://doi.org/10.29303/jrpm.v2i1.45

Krissadi, A. D. W., & Nugraha, A. (2022). *Menjadi guru penggerak merdeka belajar*. Penerbit Lakeisha.

Li, F., Zhao, Y., & Wang, C. (2023). Gamified learning and motivation: A review of digital assessment tools. *Journal of Educational Technology Research, 17*(2), 121–135.

Listrianti, F., Papatungan, M., & Qowiyah. (2023). Pengaruh media pembelajaran interaktif berbasis digital terhadap kemampuan literasi dan numerasi siswa madrasah ibtidaiyah Azzainiyah II. *Jurnal Arjuna: Publikasi Ilmu Pendidikan, Bahasa dan Matematika, 1*(5). https://doi.org/10.61132/arjuna.v1i5.273

Maulidia, F., Hobri, & Oktavianingtyas, E. (2023). Analisis kemampuan literasi numerasi siswa SMP dalam menyelesaikan soal PISA konten change and relationship. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 7*(3), 1877–1887.

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

OECD. (2023). *PISA 2022 results: Excellence and equity in education*. OECD Publishing.

Patriana, W. D., Wulandari, M. D., & Sutama. (2021). Pengelolaan pembelajaran berorientasi literasi numerasi di sekolah dasar dalam kegiatan kurikuler dan ekstrakurikuler. *Jurnal Pemikiran dan Pengembangan Sekolah Dasar, 9*(2), 116–131.

Perdana, R., & Suswandari, M. (2021). Literasi numerasi dalam pembelajaran tematik siswa kelas atas sekolah dasar. *Absis: Mathematics Education Journal, 3*(1), 9. https://doi.org/10.32585/absis.v3i1.1385

Popham, W. J. (2005). *Classroom assessment: What teachers need to know*. Pearson Education.

Prijowuntato, S. W. (2016). *Evaluasi pembelajaran*. Sanata Dharma University Press.

Pubian, Y. M., Yulianti, D., Fitirawan, H., & Nurwahidin, M. (2023). Pengembangan model blended learning berbasis Google Site untuk meningkatkan efektivitas pembelajaran siswa. *Jurnal Teknologi Pendidikan: Jurnal Penelitian dan Pengembangan Pembelajaran, 8*(2), 392–402.

Pusmenjar. (2020). *AKM dan implikasinya pada pembelajaran*. Pusat Asesmen dan Pembelajaran, Badan Penelitian dan Pengembangan dan Perbukuan, Kementerian Pendidikan dan Kebudayaan.

Pusmendik. (2021). *Tanya jawab AKM kelas*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. https://pusmendik.kemdikbud.go.id/akmkelas/tanya-jawab

Purwati, P. D., Faiz, A., Widiyatmoko, A., Ngabiyanto, & Maryatul, S. (2021). Asesmen kompetensi minimum (AKM) kelas jenjang sekolah dasar sarana pemacu peningkatan literasi peserta didik. *Sosio Religi: Jurnal Kajian Pendidikan Umum, 19*(1), 13–24.

Purwati, P. D., Widiyatmoko, A., Ngabiyanto, N., & Kiptiyah, S. M. (2021). Pembekalan guru SD Gugus Sindoro Blora melalui workshop asesmen nasional menghadapi AKM nasional. *Journal of Community Empowerment, 1*(1), 32–40. https://doi.org/10.15294/jce.v1i1.49671

Rafika Sari, A., Yusuf, A. Y. P., Ramdhania, K. F., Martyana, M. G., Nur'aini, I., Rahmadhani, S., Filiandini, R., & Putri, R. S. (2022). Adaptasi teknologi untuk meningkatkan penguatan kemampuan literasi dan numerasi siswa melalui aplikasi AKM-kelas berbasis desktop dan Android. *AMMA: Jurnal Pengabdian Masyarakat, 1*(10), 1283–1291.

Safitri, I. (2024a). Should Likert Data be Transformed Using Summated Rating Scale? A Confirmatory Factor Analysis Study on The Continuous Learning. *Jurnal Eduscience (JES)*, *11*(3), 790–801. https://doi.org/10.36987/jes.v11i3.6412

Safitri, I. (2024b). Transformation of Student Science Achievement Through Project-Based Learning (PjBL): A Meta Analysis Study. *Scientiae Educatia: Jurnal Pendidikan Sains*, *13*(2). https://doi.org/DOI: 10.24235/sc.educatia.v13i2.18694

Safitri, I., Andriyanti, E., Ansyari, R., Harahap, R. D., Harahap, R. H., & Istiqlal, M. (2025). Character Education in Indonesia: Do Integrated Islamic Schools Outperform Public Schools? *Journal of Education Culture and Society*, *16*(1), 251–270.

Safitri, I., & Ansyari, R. (2024). The Calibration of Science Achievement Test Based on Integrated Islamic Curriculum. *8th International Conference on Education and Multimedia Technology (ICEMT)*, 329–336. https://doi.org/10.1145/3678726.3678727

Safitri, I., Lestarani, D., Imtikhanah, R. D. N. W., Akbarini, N. R., Sari, M. W., Fitrah, M., & Hapsan, A. (2024). *Teori Pengukuran dan Evaluasi*. CV. Ruang Tentor.

Sani, R. A. (2021). *Strategi belajar mengajar*. INA-Rxiv.

Zhang, Y., Díaz, C., & Banihashem, S. (2023). Technology-enhanced formative assessment: Trends and challenges. *Journal of Educational Measurement and Evaluation, 45*(1), 67–84.

Zhao, F. (2019). Using Quizizz to integrate fun multiplayer activity in the accounting classroom. *International Journal of Higher Education, 8*(1), 37–43. https://doi.org/10.5430/ijhe.v8n1p37