Journal of General Education and Humanities
Vol. 5, No. 3, June 2026, pp. 3231 – 3247, https://doi.org/10.58421/gehu.v5i3.1114
ISSN 2963-7147

Contextual Design of Literacy and Numeracy–Based Assessments: A Systematic Literature Review

Mursyidah J. Parandrengi [1], Duano Sapta Nusantara[2], Rohati[3], Da Tien Nguyen[4], Kgaladi Maphutha[5]
[1,2,3]Universitas Jambi, Jambi, Indonesia
[4]Hanoi Metropolitan University, Hanoi, Vietnam
[5]University of Limpopo, Limpopo, South Africa

Article Info

Article history:

Received 2026-02-03
Revised 2026-03-30
Accepted 2026-05-23

Keywords:

Context
Development
Literacy
Numeracy
SLR

ABSTRACT

The increasing emphasis on literacy and numeracy requires assessment tasks that reflect real-world contexts. However, existing studies indicate that contextual assessment is often implemented at a superficial level, with limited use of second-order contextualization that supports deeper reasoning. This study aims to systematically examine how context is used in the design of literacy and numeracy-based assessments. This research employs a Systematic Literature Review (SLR) of 17 articles published between 2015 and 2025. Bibliometric analysis using VOSviewer and content analysis were conducted to examine context types, context levels (zero, first, and second order), and question design. The findings show that most studies apply context at the zero and first-order levels, while only a few reached second-order contextualization. Context use is dominated by local settings, with contexts such as tourism remaining underexplored. In addition, digital technology is mostly used in non-interactive formats. This study recommends developing PISA-based questions in tourism contexts at the second-order level, as well as integrating interactive digital technology to enhance students' abilities to formulate, employ, and interpret. This study contributes theoretically by linking context type, context level, and cognitive processes within mathematical literacy frameworks, and practically by guiding the design of more authentic assessment tasks.

Corresponding Author:

Duano Sapta Nusantara
Faculty of Teacher Training and Education, Universitas Jambi, Indonesia
Email: duanosaptanusantara@unja.ac.id

1. INTRODUCTION

In the context of rapid digital advancement, the increasing availability of data, and complex global challenges such as climate change, pandemics, and economic instability, the demands placed on individuals have shifted significantly. Modern society requires individuals who are not only knowledgeable but also capable of thinking critically, making informed decisions, and actively engaging as responsible citizens [1]. In this regard, literacy

and numeracy have become fundamental competencies, enabling individuals to interpret information, evaluate data, and respond effectively to real-world situations [2], [3]. These competencies extend beyond academic achievement, playing a crucial role in everyday life and professional contexts where data interpretation and decision-making are essential.

Although closely related, numeracy and mathematics are conceptually distinct. Mathematics tends to develop through abstraction, whereas numeracy emphasizes the application of mathematical knowledge in real-life contexts [4]. Similarly, numeracy involves the ability to formulate, use, and interpret mathematics in various contexts, supported by structured reasoning and argumentation [1], [5]. This highlights that the core of literacy and numeracy lies not only in mastering concepts but also in applying them meaningfully in contextually rich situations [6].

However, empirical evidence indicates that these competencies have not yet been optimally developed. The PISA 2022 results show that Indonesian students' performance remains relatively low, with most students still operating at basic levels of reading and reasoning [7]. This suggests that many students have not yet developed the ability to interpret, evaluate, and apply knowledge in complex contexts. One contributing factor is the limited availability of assessment tasks that reflect real-world problem-solving demands, as well as teachers' preparedness in designing such tasks [8], [9].

Prevailing assessment practices further influence this condition. In many classrooms, tasks are still dominated by routine and procedural questions that are detached from meaningful contexts [10]. As a result, students often experience difficulties when required to transfer their knowledge to unfamiliar or real-world situations [11]. Although contextual problems have been introduced in some studies, their implementation often remains superficial, with context functioning merely as a narrative backdrop rather than as a driver of reasoning and decision-making.

Ideally, literacy and numeracy-based assessments should integrate content and context in a meaningful way, allowing students to connect mathematical ideas with real-life experiences and socially relevant issues [1], [12], [13]. Within the PISA framework, this integration is reflected in the processes of formulating, employing, and interpreting mathematics in real-world situations [1]. Furthermore, the concept of context levels (zero, first, and second order) provides a useful theoretical lens to evaluate how deeply context influences students' reasoning processes. In well-designed assessments, context should not only support the presentation of problems but also guide the formulation of models, the selection of assumptions, and the interpretation of results.

Nevertheless, existing research has not yet fully addressed this expectation. Previous studies have explored various contextual approaches, ranging from local cultural settings to global issues such as pandemics and environmental challenges [14], [15], [16], [17], [18], [19]. While these studies demonstrate the potential of contextual tasks to enhance engagement and relevance, they also reveal inconsistencies in how context is conceptualized and implemented. In many cases, contextualization remains at a surface level and does not significantly affect the task's cognitive demands. Only a limited number of studies have attempted to position context as a central element that shapes problem formulation and interpretation.

This indicates a clear gap in the literature. There is still a limited systematic understanding of how context is structured, how different levels of contextualization are applied, and how these aspects relate to the design of literacy and numeracy-based assessment tasks. Without such understanding, efforts to develop authentic and higher-order assessments risk remaining fragmented and inconsistent.

Therefore, this study aims to systematically analyze how context is used in the development of literacy- and numeracy-based assessment tasks by integrating bibliometric and content analyses. Specifically, this study examines the types of context used, the levels of contextualization applied, and the characteristics of question design to identify patterns, limitations, and opportunities for improvement. The following research questions guide this study:

RQ1: What is the profile of the research literature on the use of context in developing literacy and numeracy-based questions?

RQ2: How is context used in the design and development of literacy and numeracy-based questions, and what types of contexts are applied?

RQ3: What gaps and recommendations can be identified from existing studies regarding the use of context in assessment design?

This study is expected to contribute both theoretically and practically. Theoretically, it provides a more structured understanding of contextual assessment by linking context levels with numeracy frameworks. In practice, it offers guidance to educators and test developers in designing more authentic, context-rich assessment tasks that better support students' higher-order thinking skills and align with international standards.

2. METHOD

This study used an SLR approach to identify and assess relevant research, requiring a level of rigor beyond what is typically required [20]. This approach was chosen because it provides integrated empirical evidence on how context is used in the development of literacy and numeracy-based questions. The SLR was implemented through three main steps: review planning, review implementation, and reporting the review results.

2.1. Review Planning

The planning stage begins with determining the study focus and formulating the research problem (RM1–RM3). Next, a review protocol is developed, including the search strategy, data source selection, and inclusion and exclusion criteria. To obtain comprehensive results, a combination of keywords was used: ("*literacy*" AND "*numeracy*"), ("*literacy*" AND "*test development*"), and ("*numeracy*" AND "*test development*"). The literature search was conducted using the Scopus database as the primary source due to its wide coverage of high-quality, peer-reviewed international journals. Scopus was selected to ensure the reliability and credibility of the retrieved studies.

2.2. Review Implementation

This stage included a literature search, screening, data extraction, and content analysis. The initial search yielded several articles from the Scopus database. After filtering for duplicates and applying exclusion criteria, a set of relevant articles was identified for further analysis. Table 1 presents the inclusion and exclusion criteria used in this study.

Table 1. Inclusion and Exclusion Criteria for SLR Literacy and Numeracy-Based Test Development

Criteria	Description
Title word	"Literacy"
	"Numeracy"
	("Literacy" AND "Numeracy")
Keyword	("Literacy" AND "Test Development")
	("Numeracy" AND "Test Development")
	- One of the selected keywords does not appear in the title, abstract, keywords, or full text.
	- Not written in English.
Exclusion	- Publisher unknown.
	- Outside the field of mathematics.
	- Full text not available/link inaccessible.
	- One of the selected keywords appears in the title, abstract, keywords, or full text.
	- Written in English.
Inclusion	- Published in one of the selected databases.
	- Is a journal article.
	- Published between 2015 and 2025.

The initial search in the Scopus database yielded 9,616 articles based on the defined keywords. After applying inclusion and exclusion criteria and removing irrelevant records, 125 articles were identified for further screening. Subsequently, title and abstract screening reduced the number to 39 articles. A full-text assessment was then conducted, resulting in 17 articles that met all eligibility criteria and were included in the final analysis. Initial screening was performed using keywords, titles, and relevant databases. This resulted in the identification of 9,616 articles from Scopus using the keywords in Table 1, yielding the results presented in Table 2.

Table 2. Article Search Results by Title, Keywords, and Abstract In Scopus

Title Word	Keywords	Number of Articles
	"Literacy" and "Numeracy"	3681
Literacy	"Literacy" and "Test Development"	5485
Numeracy	"Numeracy" and "Test Development"	450
	Total of all articles	9616

The selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure transparency and methodological rigor. This process involved four main stages: identification, screening, eligibility assessment, and final inclusion of studies. Each stage was conducted systematically to filter relevant articles according to predefined inclusion and exclusion criteria. The article selection process is represented in Figure 1.

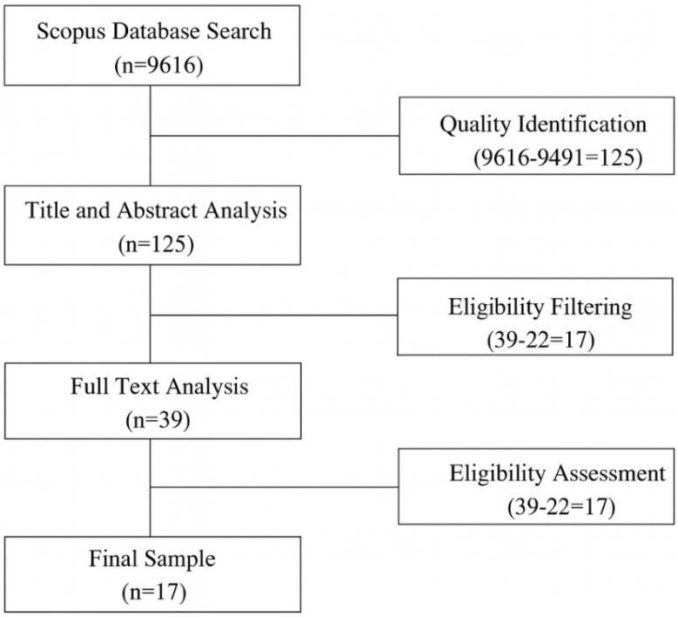

Figure 1. PRISMA-Based Selection Process of Contextual Approaches to Developing Literacy and Numeracy Test Items

After selecting 17 relevant articles for further review and analysis (see Figure 1), each article was coded along several dimensions, including title, author(s), year of publication, country, context, method, and question design. Furthermore, word clouds and VOSviewer bibliometric software were used to examine the literature and network profiles quantitatively. The data were then used to analyze the context and question design, identify gaps, and generate recommendations for further action from the literature.

2.3. Reporting Review Results

The final stage of this research includes compiling the analysis results, which are divided into three main sections. First, a literature profile, covering author trends, countries of origin, and methods used in related studies. Second, the use of context in item development, covering types such as personal, social, scientific, and cultural, along with item design and the design strategies applied. Third, research gaps and recommendations, which outline the limitations identified and opportunities for further research development. All analysis results are presented in tables, publication trend graphs, and a conceptual model that illustrates the relationship between context, item design, and the literacy and numeracy skills measured. All these steps are carried out as shown in the diagram in Figure 2 [21].

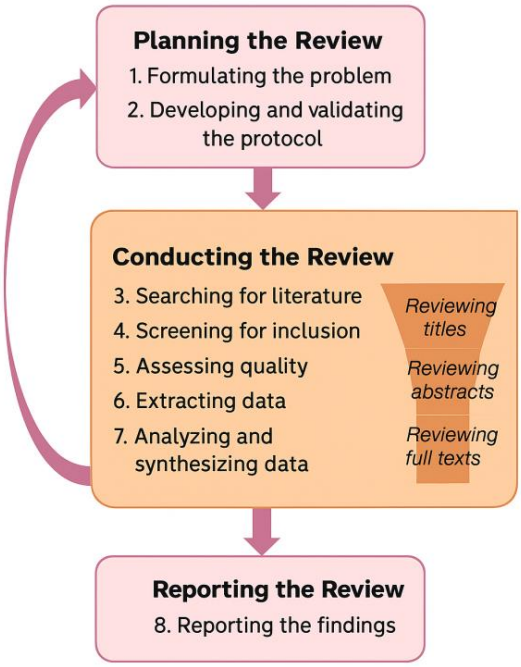

Figure 2. SLR Steps of Contextual Approaches to Developing Literacy and Numeracy Test Items

3. RESULTS AND DISCUSSION

3.1. Results

In this study, three types of analysis were conducted: research profile and network analysis, qualitative content analysis, and analysis of research gaps and recommendations. This section presents the results of the literature review on the development of literacy- and numeracy-based tests. The distribution of publications over time is shown in Figure 3. A total of 17 articles were identified, with publication trends increasing from 2020 to 2025.

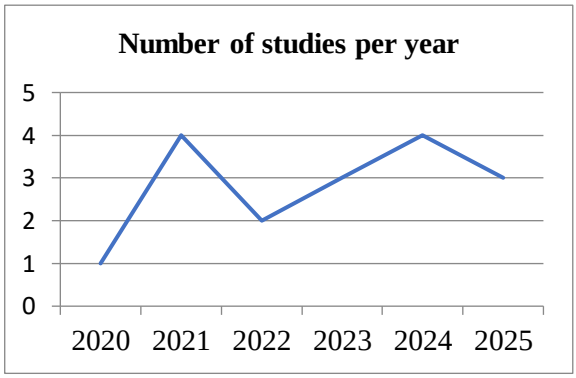

Figure 3. Number of Researches Per Year in Literacy and Numeracy-Based Test Development

A total of 53 authors contributed to the selected studies. Among them, 45 are from Indonesia, 3 from the United States, 3 from Israel, 1 from Malaysia, and 1 from Norway. International collaboration is limited, with only one collaboration identified between Indonesia and Malaysia.

Table 3. Top Five Authors of Literacy and Numeracy-Based Test Development Research

Author	Country	Number of Article	Link Strength
Zulkardi	Indonesia	6	20
Ratu Ilma Indra Putri	Indonesia	4	18
Duano Sapta Nusantara	Indonesia	2	13
Sisca Puspita Sepriliani	Indonesia	2	13
Meryansumayeka	Indonesia	1	11

Table 3 presents the top five authors based on the number of publications and link strength, as determined through VOSviewer analysis. Zulkardi contributed six articles with a link strength of 20, followed by Ratu Ilma Indra Putri with four articles and a link strength of 18. Duano Sapta Nusantara and Sisca Puspita Sepriliani each contributed two articles with a link strength of 13, while Meryansumayeka contributed one article with a link strength of 11.

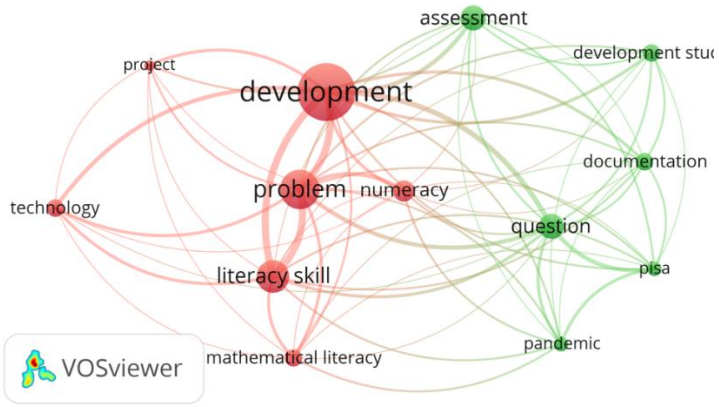

Figure 4. VOSViewer Network Visualization for Literacy and Numeracy-Based Test Development Research

The relationships between keywords were analyzed using VOSViewer. Figure 4 presents the keyword co-occurrence network, which forms two main clusters. The first cluster includes the keywords development, problem, numeracy, literacy skills, mathematical literacy, technology, and project. The second cluster includes the keywords assessment, documentation, question, development study, PISA, and pandemic. The keyword "*development*" appears as the most dominant node, followed by "*problem*", "*literacy skill*", "*numeracy*", "*assessment*", and "*question*".

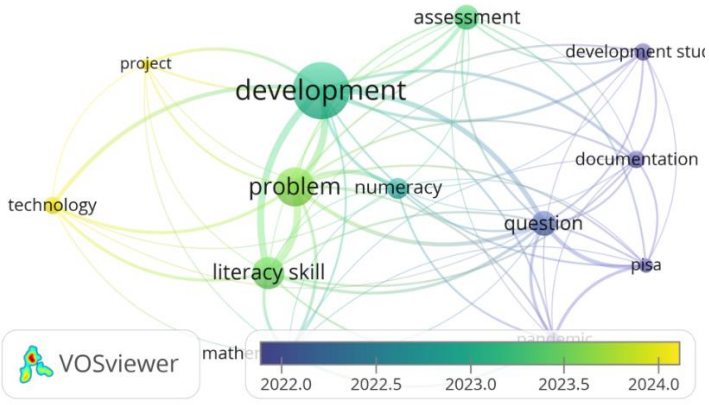

Figure 5. VOSViewer Visualization Overlay for Literacy and Numeracy-Based Test Development Research

Figure 5 shows the overlay visualization of keyword trends over time. The earlier period is characterized by keywords such as PISA, question, documentation, and pandemic. In the following period, keywords such as assessment, problem, and numeracy appear. More recent trends include keywords such as technology, project, and development.

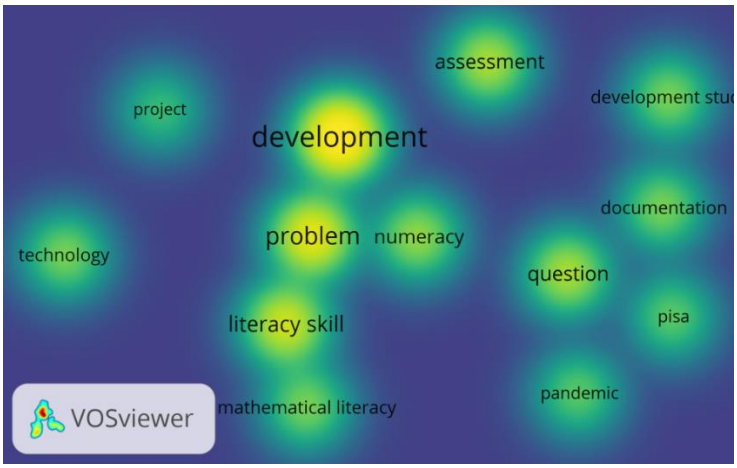

Figure 6. VOSViewer Density Overlay for Literacy and Numeracy-Based Test Development Research

Figure 6 presents the density visualization of keyword occurrences. Keywords such as *"development", "numeracy",* "*problem*", and "*literacy skill*" development, problem, literacy skills, and numeracy appear with the highest density. Keywords such as "*technology*", "*project*", "*assessment*", "*question*", and "*PISA*" appear at moderate density, while "pandemic" and "documentation" appear at lower density.

Content analysis was conducted to examine the use of context and question design in the selected studies. Table 4 presents the classification of studies based on context, context level, and question design.

The reviewed studies show variation in the types of context used, including local contexts such as cultural practices, traditional food, games, and public transportation, as well as global contexts such as COVID-19, climate change, social issues, and environmental problems. Based on context levels, the studies are classified into three categories: zero-order, first-order, and second-order context.

In zero-order contexts, the context functions only as background information without influencing the problem-solving process. In first-order contexts, the context is used to identify relevant variables and support the construction of mathematical models. In second-order contexts, the context serves as the primary basis for determining variables, assumptions, and the interpretation of results.

The distribution of studies shows that most are categorized as first-order, followed by zero-order, with only a few in the second-order category. The question designs identified include PISA-like tasks, numeracy-based tasks, and technology-supported formats such as e-modules, adaptive digital books, and project-based learning models, with technology use largely remaining non-interactive.

Table 4. Context, Context Level, and Question Design Used in the Article

No.	Title	Authors	Year	Country	Context	Context Level	Desain Soal
1.	Reading text for school literacy movement in mathematics learning	Effendi et al. (2020)	2020	Indonesia	Futsal	-	literacy
2.	Improving PISA-like questions through trialmusi video context: lrt, damri and transmusi	Agustina et al. (2021)	2021	Indonesia	LRT, Damri, Transmusi	-	PISA
3.	Designing PISA-like mathematics task using a COVID-19 context (PISACOMAT)	Nusantara et al. (2021)	2021	Indonesia	COVID-19 (Large-Scale Social Restrictions (LSSR) And Panic Buying)	Second Order	PISA
4.	Lower secondary students' encounters with mathematical Literacy	Bolstad (2021)	2021	Norwegia	Social Life, Work-Life, Citizenship	-	Numeracy
5.	Mathematical media literacy in the COVID-19 pandemic And its relation to school mathematics education	Heyd-Metzuyanim et al. (2021)	2021	Israel	COVID-19	-	PIAAC
6.	Development of e-module using Bengkulu contexts to improve literacy skills of junior high school students	Susanta et al. (2021)	2022	Indonesia	Bengkulu	First Order	Numeracy
7.	The development of PISA-based numerical problem using the context of religious day during the pandemic	Sepriliani et al. (2022)	2022	Indonesia	Religious day	First Order	PISA
8.	Development of PISA-like activities using the inquiry-based learning model and the context of religious holidays during the pandemic	Sepriliani et al. (2023)	2023	Indonesia	Religious day	Zero Order	PISA
9.	Mathematics literacy task on number pattern using Bengkulu context for junior high school students	Susanta et al. (2023)	2023	Indonesia	Bengkulu	First Order	Numeracy
10.	Development of PISA-type questions and activities in a smartphone context	Mouli et al. (2023)	2023	Indonesia	Smartphone	First Order	PISA
11.	The development of high school students' statistical literacy across grade level	Kurnia et al. (2024)	2024	Indonesia	Shoes Production, Household Waste, Test Scores, 100m Race	Zero Order	TIMMS
12.	Developing steam-teaching module in supporting students' literacy ability in elementary school	Susanto et al. (2024)	2024	Indonesia	ICT	-	Numeracy

https://doi.org/10.58421/gehu.v5i3.1114

No.	Title	Authors	Year	Country	Context	Context Level	Desain Soal
13.	Examining the role of context in statistical literacy assessment	Phadke et al. (2024)	2024	United States	Climate Change, Immigration, Race-Related Issues, And The COVID-19 Pandemic.	First Order	BLIS & MBLIS
14.	Integrating traditional food and technology in statistical learning: a learning trajectory	Ramadhani et al. (2024)	2024	Indonesia	Lemang Batok Traditional Food	Second Order	Numeracy
15.	Developing adaptive digital book (ADB) in enhancing students' numeracy literacy ability	Yerizon et al. (2025)	2025	Indonesia	Seesaw Game	Zero Order	Literacy
16.	Steam approach in project-based learning to develop mathematical literacy and students' character	Supianti et al. (2025)	2025	Indonesia	Temperature, Wind-Driven Toy Cars	Second Order	Numeracy
17.	Numeracy skill development in prospective mathematics teachers: challenges and opportunities in real-world contexts	Winarni et al. (2025)	2025	Indonesia	Save Our Water	First Order	Numeracy

3.2. Discussion

The results of the analysis of 17 articles show that research on the development of literacy and numeracy-based questions has a consistent pattern in two main dimensions, namely context and question design.

Although previous studies have emphasized the importance of contextual assessment, the existing frameworks show several limitations. First, many approaches do not clearly distinguish between different levels of contextualization, resulting in superficial integration of context in assessment tasks. Second, current frameworks often focus on task development without systematically linking context to cognitive processes such as reasoning, modeling, and interpretation. Third, the alignment between contextual design and numeracy frameworks, particularly the PISA processes of formulating, employing, and interpreting, remains inconsistent across studies.

These limitations suggest that context is frequently treated as an additional feature rather than as a core component that shapes problem-solving processes. As a result, many assessment tasks fail to support higher-order thinking skills and authentic mathematical reasoning fully.

Context

In terms of context, research shows wide variation between local and global contexts. Local contexts used include futsal [22], Bengkulu [23] and [24], lemang batok [25], holidays [18] and [26], games [27], seesaw games [28], and public transportation [29], which are used to connect mathematics to students' culture and environment. Students who learn through culturally contextualized instruction demonstrate stronger conceptual understanding and enhanced mathematical problem-solving abilities [30]. The use of global contexts such as COVID-19 LSSR and panic buying [17], COVID-19 [31], climate change [32], social issues such as social life, work-life, and citizenship [33], household waste [34], save our water [35], smartphones [36], and ICT [37] are used to foster critical awareness and social literacy in students. Climate change itself is a real-world problem we are currently experiencing [38]. This context emphasizes that developing literacy and numeracy-based questions serves not only as a measure of mathematical ability but also as a means of building social relevance and deeper meaning in learning, thereby fostering critical thinking skills. Critical thinking skills are a crucial component for students in mathematics learning [39]. Moreover, integrating local wisdom as a meaningful contextual foundation further deepens students' understanding of mathematical concepts, as problems are embedded in familiar cultural and everyday experiences, thereby enabling learners to interpret, model, and solve mathematical problems more effectively.

Context can be categorized into three levels of use: zero-order, first-order, and second-order [40]. It is categorized as zero order because the context is only used as a background story that allows students to take direct action according to instructions, and it is not used to interpret results or support mathematical reasoning. It is categorized as first order because the context helps students find or select information, variables, or relationships that are important for building mathematical models, and it is also used to assess whether the mathematical results obtained are appropriate. Meanwhile, it is categorized as second

order because the context is the main source for determining the variables, assumptions, or relationships needed in formulate problems mathematically, and it is used to evaluate whether the results or mathematical arguments are relevant and adequate when linked back to the initial situation. Most developed tasks fall into the first-order use of context, meaning that although the context is explicitly stated, it remains directly relevant to solving the task [41].

Several articles mention the context but do not include the developed questions. Of the articles that do include questions, only three can be categorized as second order, six as first order, and three as zero order. The questions are categorized as second-order because the context is used as the primary source for determining variables and assumptions in solving them [17], [25], [27]. The questions are categorized as first-order because the context is used only to determine the variables for building mathematical models [18], [23], [24], [32], [35], [36]. Meanwhile, the questions are categorized as zero-order because the context is used only as a story background, allowing students to work on the questions without having to pay attention to it [26], [28], [34].

Question Design

From a question-design perspective, research shows a strong trend toward developing PISA-like tasks that emphasize contextual problem-solving and the use of real-world data. Furthermore, numeracy- and literacy-based tasks are emerging that integrate reading, interpreting information, and quantitative reasoning. Recent trends indicate innovations in technology-based question design, such as e-modules, adaptive digital books, and project-based STEAM learning models. These innovations indicate a shift in research direction toward developing questions that are more interactive, adaptive, and relevant to 21st-century skills. To create a good education, educators must incorporate the 4C skills (creative thinking, critical thinking, problem solving, communication, and collaboration) into every subject [42]. There is a correlation between the 4C skills, numeracy literacy, and students' mathematics learning outcomes [42].

Students lacked experience in tackling PISA-style contextual problems, and teachers generally offered routine, lower-level exercises [43]. To strengthen their international performance, Indonesian students need greater exposure to mathematics tasks modeled after PISA [2]. The success of Indonesian students in solving PISA problems is influenced by the evaluation system and by teachers' ability to develop students' numeracy [44]. Thus, it is necessary to develop PISA-type mathematics questions to improve students' literacy and numeracy skills.

The use of technology in task design can also be implemented through effective, purposeful integration rather than simple adoption, as this approach has been identified as a primary driver of increased motivation, engagement, and academic achievement [45]. One platform that may be utilized is MathCityMap (MCM). The MCM Project is a mathematics trail initiative conducted in various urban locations and supported through a mobile application or a printed guide [46]. Modelling type mathematics problems may be developed from real situations in the local environment. Teachers can design site-based tasks associated with particular places, and these tasks can be interconnected to form routes linking multiple

task locations [47]. The integration of MCM into task design enables the development of contextual, real-world mathematical problems that align with the principles of PISA assessments. Within the setting of a mathematics trail supported by the MathCityMap application, the study examines the impact of positive and negative feedback on students' processes of verification and elaboration [48]. Through its location-based activities and authentic scenarios, MCM provides a rich environment for generating tasks that mirror the real-life relevance emphasized in PISA-type items.

Based on these findings, this study proposes a conceptual model for contextual assessment design that integrates three key components: context type, context level, and cognitive process. In this model, context type (local or global) provides the situational relevance, while context level (zero, first, second order) determines the depth of contextualization. These elements are then linked to cognitive processes, particularly the PISA framework of formulating, employing, and interpreting, which represent different stages of mathematical reasoning.

In this framework, higher-order contextualization occurs when context operates at the second-order level, directly influencing the formulation of variables, assumptions, and interpretations. This interaction enables assessment tasks to move beyond procedural problem-solving toward authentic mathematical modeling. Conversely, lower levels of contextualization limit the role of context and reduce the task's cognitive demands.

This conceptual model contributes to the existing literature by providing a more structured understanding of how context can be systematically integrated into assessment design. It bridges the gap between contextual task development and theoretical frameworks of numeracy, offering a clearer guideline for designing assessment tasks that promote higher-order thinking skills.

Research Gaps and Recommendations for the Future

A review of 17 articles shows that the development of literacy and numeracy-based questions is still dominated by the use of context at the zero and first order levels, where context serves only as a backdrop or provides basic information for building mathematical models. Only a small portion of research truly applies context at the second-order level, where context becomes the primary source for determining variables, assumptions, reasoning, and evaluating the relevance of solutions. Furthermore, although climate change and tourism are rich, real-world contexts close to students' lives, their use in previous research has tended to be superficial and has not been considered as contexts that require in-depth mathematical modeling. On the other hand, the use of digital technology in presenting questions, such as interactive platforms, data simulations, or adaptive media, is still limited to simple e-module or digital book formats, without exploiting the full potential of technology to present real-time data or complex contextual dynamics. This condition indicates that PISA-like question design, which actually requires the process of formulating, employing, and interpreting, has not been fully implemented authentically in the tourism context. Thus, there is a significant research gap in producing digital questions in the tourism context that are second-order, authentic, and able to stimulate students' higher-order reasoning.

Based on these gaps, further research needs to guide the development of questions in the context of climate change and tourism, processed at the second-order level, so that the context is not merely a way to beautify the questions, but becomes the primary source in determining variables, assumptions, and solution interpretations. Furthermore, the use of digital technology needs to be expanded through interactive platforms that leverage data visualization, environmental change simulations, or automated evaluation systems, so that questions can be adaptive, contextual, and better support students' critical thinking skills. Further research is also recommended to develop a standardized PISA-like question design framework specifically for the context of tourism, which includes steps for question construction, context validation, formulation of cognitive demands, and development of reasoning-based rubrics. The integration of numeracy literacy and 4C skills through digital media is also important for further research, for example, by developing data-driven collaborative activities or exploring authentic situations that encourage students to formulate and evaluate their models.

Experimental or quasi-experimental research can also be conducted to test the effectiveness of second-order digital problems in improving numeracy literacy, problem-solving, and student engagement. This way, future development of mathematics problems can be more meaningful, contextual, and relevant to the challenges of the 21st century.

4. CONCLUSION

This study provides a systematic examination of how context is used in the design of literacy and numeracy-based assessments. The findings reveal that, while contextual approaches are widely adopted, the depth of contextualization remains limited, with most studies positioning context at surface levels rather than as a core element that shapes reasoning processes. This indicates that the potential of context to support meaningful mathematical thinking has not yet been fully realized.

This study contributes by proposing a structured perspective on contextual assessment that integrates context type, context level, and cognitive processes within the numeracy framework. By explicitly linking these elements, this study offers a clearer conceptual foundation for understanding how context can function beyond narrative support and instead become a driver of higher-order reasoning in assessment design.

The implications of this study are relevant for both practice and research. For educators and assessment developers, there is a need to design tasks that move toward deeper contextualization and integrate real-world complexity with appropriate cognitive demands. At the same time, the findings highlight the importance of aligning assessment design with established frameworks such as PISA to ensure that tasks support the processes of formulating, employing, and interpreting. However, this study is limited by the scope of the selected articles and the use of a single primary database, which may not capture all relevant research in this field.

Future research is recommended to expand the scope of analysis by incorporating multiple databases and exploring empirical validation of context-based assessment models. Further studies should also investigate the effectiveness of second-order contextualization and the integration of interactive digital environments in enhancing students’ literacy,

numeracy, and higher-order thinking skills. In addition, future research is encouraged to explore underutilized contexts, such as tourism, which offer rich and authentic real-world situations but have received limited attention in existing studies. Such contexts can support more meaningful, interdisciplinary learning experiences when integrated into assessment design. Ultimately, this study underscores the importance of developing more authentic, theoretically grounded assessment practices that better align with the demands of 21st-century learning and real-world problem-solving.

ACKNOWLEDGEMENTS

The author expresses sincere appreciation to all researchers whose work has been cited in this systematic literature review. Their scholarly contributions have provided valuable insights and a strong foundation for the analysis and synthesis presented in this study.

REFERENCES

[1] OECD, *PISA 2022 Assessment and Analytical Framework*. 2023. [Online]. Available: https://www.oecd.org/en/publications/pisa-2022-assessment-and-analytical-framework_dfe0bf9c-en.html

[2] D. S. Nusantara, Zulkardi, and R. I. I. Putri, "Designing PISA-like mathematics problem in COVID-19 pandemic (PISAComat)," *J. Phys. Conf. Ser.*, vol. 1657, no. 1, 2020, doi: 10.1088/1742-6596/1657/1/012057.

[3] A. Fauzan *et al.*, "Realistic mathematics education (RME) to improve literacy and numeracy skills of elementary school students based on teachers' experience," *Infin. J.*, vol. 13, no. 2, pp. 301–316, 2024, doi: 10.22460/infinity.v13i2.p301-316.

[4] M. Goos and K. O'Sullivan, "Numeracy across the curriculum," *Oxford Res. Encycl. Educ.*, no. November, 2022, doi: 10.1093/acrefore/9780190264093.013.1530.

[5] D. S. Nusantara, Z. Zulkardi, and R. I. I. Putri, "Investigating students' mathematical argumentation ability in solving PISA-like mathematics tasks on change and relationships," *Math. Educ. J.*, vol. 19, no. 4, pp. 807–826, 2025, doi: 10.22342/mej.v19i4.pp807-826.

[6] U. Lusinda, F. T. Pasaribu, T. Gustiningsi, and D. S. Nusantara, "Pengembangan soal matematika tipe PISA konten uncertainty and data menggunakan konteks museum siginjai Jambi," *Numeracy*, vol. 12, no. 1, pp. 99–117, 2025, doi: 10.46244/numeracy.v12i1.3128.

[7] OECD, *PISA 2022 Result (Volume 1): The State of Learning and Equity in Education*, vol. I. 2023. [Online]. Available: https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en.html

[8] A. Fauzan, Y. Harisman, G. Boriboon, H. Syaputra, and Hafizatunnisa, "Enhancing mathematical literacy: A comparative study of mathematics learning curricula in Thailand and Indonesia," *Infin. J.*, vol. 14, no. 3, pp. 753–780, 2025, doi: 10.22460/infinity.v14i3.p753-780.

[9] R. Kiraam, D. S. Nusantara, and F. T. Pasaribu, "Students' errors in solving PISA questions on space and shape content," *Numeracy*, vol. 12, no. 2, pp. 149–162, 2025, doi: 10.46244/numeracy.v12i2.3436.

[10] D. A. Lapele, "Identifikasi kendala calon guru dalam menerapkan pembelajaran matematika realistik," *J. Cendekia J. Pendidik. Mat.*, vol. 6, no. 3, pp. 3126–3135, 2022, doi: 10.31004/cendekia.v6i3.1055.

[11] I. Juniarmi, Rohati, and D. S. Nusantara, "Students ' analogical reasoning in solving number pattern problems," vol. 4, no. 4, pp. 2227–2241, 2025, doi: 10.58421/gehu.v4i4.735.

[12] Zulkardi *et al.*, "How students work with pisa-like mathematical tasks using COVID-19 context," *J. Math. Educ.*, vol. 11, no. 3, pp. 405–416, 2020, doi: 10.22342/jme.11.3.12915.405-416.

[13] B. W. Pangestika, N. Diah, S. Lestari, A. Cahya, D. Kurniati, and A. Suwito, "Culturally Responsive Teaching : Improving Middle School Numeracy in Linear Functions," vol. 5, no. 1, pp. 851–861, 2026.

[14] T. Gustiningsi, R. I. I. Putri, Zulkardi, and Hapizah, "Developing a PISA-like mathematical problem : Using traditional food context," *Int. J. Educ. Curric. Appl.*, vol. 6, no. 3, pp. 324–337, 2023.

[15] D. S. Nusantara, Zulkardi, and R. I. I. Putri, "Designing Pisa-like mathematics problem using a COVID-19 transmission map context," *AIP Conf. Proc.*, vol. 2438, no. 1, p. 20005, Oct. 2021, doi: 10.1063/5.0071596.

[16] D. S. Nusantara, F. P. Ibarra, F. T. Pasaribu, and U. Lusinda, "Designing JUMPISA : A PISA-based jumping task to support students ' statistical literacy using Jambi context," *J. Instr. Math.*, vol. 6, no.

2, pp. 96–116, 2025, doi: 10.37640/jim.v6i2.2518.
[17] D. S. Nusantara, Zulkardi, and R. I. I. Putri, "Designing pisa-like mathematics task using a COVID-19 context (PISAComat)," *J. Math. Educ.*, vol. 12, no. 2, pp. 349–364, 2021, doi: 10.22342/JME.12.2.13181.349-364.
[18] S. P. Sepriliani *et al.*, "The development of PISA-based numerical problem using the context of religious day during the pandemic," *Math. Educ. J.*, vol. 16, no. 2, pp. 157–170, 2022, doi: 10.22342/jpm.16.2.16010.157-170.
[19] R. I. Sistyawati *et al.*, "Development of PISA types of questions and activities content shape and space context pandemic period," *Infin. J.*, vol. 12, no. 1, pp. 1–12, 2023, doi: 10.22460/infinity.v12i1.p1-12.
[20] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," *SSRN Electron. J.*, vol. 10, no. 2010, 2012, doi: 10.2139/ssrn.1954824.
[21] I. Risdiyanti, Z. Zulkardi, R. I. I. Putri, R. C. I. Prahmana, and D. S. Nusantara, "Ratio and proportion through realistic mathematics education and pendidikan matematika realistik Indonesia approach: A systematic literature review," *J. Elem.*, vol. 10, no. 1, pp. 158–180, 2024, doi: 10.29408/jel.v10i1.24445.
[22] K. Nia, S. Effendi, R. I. I. Putri, and P. Yaniawati, "Reading tect book for school literacy movement in mathematics learning," *J. Pendidik. Mat.*, vol. 14, no. 2, pp. 145–154, 2020, doi: 10.22342/jpm.14.2.6731.145-154.
[23] A. Susanta, H. Sumardi, and Zulkardi, "Development of e-module using Bengkulu contexts to improve literacy skills of junior high school students," *Math. Educ. J.*, vol. 16, no. 2, pp. 171–186, 2022, doi: 10.22342/jpm.16.2.17698.171-186.
[24] A. Susanta, H. Sumardi, E. Susanto, and H. Retnawati, "Mathematics literacy task on number pattern using Bengkulu context for junior high school students," *J. Math. Educ.*, vol. 14, no. 1, pp. 85–102, 2023, doi: 10.22342/JME.V14I1.PP85-102.
[25] R. Ramadhani, R. C. I. Prahmana, Soeharto, and A. Saleh, "Integrating traditional food and technology in statistical learning: A learning trajectory," *J. Math. Educ.*, vol. 15, no. 4, pp. 1277–1310, 2024, doi: 10.22342/jme.v15i4.pp1277-1310.
[26] S. P. Sepriliani, Zulkardi, and Somakim, "Development of PISA-like activities using the Inquiry-based learning model and the context of religious holidays during the pandemic," *Math. Educ. J.*, vol. 17, no. 1, pp. 37–54, 2023, doi: 10.22342/jpm.17.1.17765.37-54.
[27] I. I. Supianti, P. Yaniawati, E. Bonyah, A. W. Hasbiah, and N. Rozalini, "STEAM approach in project-based learning to develop mathematical literacy and students' character," *J. Mat. Educ.*, vol. 14, no. 2, pp. 283–302, 2025, doi: 10.22460/infinity.v14i2.p283-302.
[28] Yerizon, Arnellis, Suherman, I. M. Arnawa, C. Sa'dijah, and L. Anwar, "Developing adaptive digital book (ADB) in enhancing students' numeracy literacy ability," *Infin. J.*, vol. 14, no. 3, pp. 571–586, 2025, doi: 10.22460/infinity.v14i3.p571-586.
[29] A. Agustina and Zulkardi, "Improving PISA-like questions through trialmusi video context: LRT, damri, and transmusi," *Math. Educ. J.*, vol. 15, no. 1, pp. 91–102, 2021, doi: 10.22342/jpm.15.1.11989.91-102.
[30] Khairullah, R. C. I. Prahmana, and N. R. N. Peni, "Wura Bongi Monca dance as a promising context for fostering problem-solving skills in learning set theory," *J. Elem.*, vol. 11, no. 4, pp. 828–844, 2025, doi: 10.29408/jel.v11i4.30579.
[31] E. Heyd-Metzuyanim, A. J. Sharon, and A. Baram-Tsabari, "Mathematical media literacy in the COVID-19 pandemic and its relation to school mathematics education," *Educ. Stud. Math.*, vol. 108, no. 1–2, pp. 201–225, 2021, doi: 10.1007/s10649-021-10075-8.
[32] S. Phadke, M. Beckman, and K. L. Morgan, "Examining the role of context in statistical literacy assessment," *Stat. Educ. Res. J.*, vol. 23, no. 1, pp. 1–17, 2024, doi: 10.52041/SERJ.V23I1.529.
[33] O. H. Bolstad, "Lower secondary students' encounters with mathematical literacy," *Math. Educ. Res. J.*, vol. 35, no. 1, pp. 237–253, 2023, doi: 10.1007/s13394-021-00386-7.
[34] A. B. Kurnia, T. Lowrie, and S. M. Patahuddin, "The development of high school students ' statistical literacy across grade level," *Math. Educ. Res. J.*, vol. 36, no. 1, pp. 1–29, 2024, doi: 10.1007/s13394-023-00449-x.
[35] S. Winarni, Kamid, J. Marzal, and Asrial, "Numeracy skill development in prospective mathematics teachers: Challenges and opportunities in real-world contexts," *J. Math. Educ.*, vol. 16, no. 1, pp. 91–110, 2025, doi: 10.22342/jme.v16i1.pp91-110.
[36] M. F. Mouli, Zulkardi, and R. I. I. Putri, "Development of PISA-type questions and activities in a smartphone context," *Math. Educ. J.*, vol. 17, no. 1, pp. 1–20, 2023, doi: 10.22342/jpm.17.1.19386.1-20.
[37] E. Susanto, A. Susanta, Rusnilawati, and S. R. B. Ali, "Developing STEAM-teaching module in

supporting students' literacy ability in elementary school," *Math. Educ. J.*, vol. 18, no. 3, pp. 349–366, 2024, doi: 10.22342/jpm.v18i3.pp349-366.

[38] M. J. Parandrengi and C. Hiltrimartin, "Students' errors in solving climate change context mathematical modeling problems," *J. Elem.*, vol. 9, no. 2, pp. 388–402, 2023, doi: 10.29408/jel.v9i2.11422.

[39] A. Z. N. Fajrina, M. Zaidan, and F. Mulyatna, "Analisis kemampuan berpikir kritis matematis siswa kelas VII dalam menyelesaikan soal perbandingan," *J. Pendidik. Guru Mat.*, vol. 3, no. 1, pp. 277–292, 2024, doi: 10.33387/jpgm.v3i1.5740.

[40] F. A. Salgado, "Developing a theoretical framework for classifying levels of context use for mathematical problems," *Proc. 39th Annu. Conf. Math. Educ. Res. Gr. Australas.*, no. 1, pp. 110–117, 2016, [Online]. Available: https://files.eric.ed.gov/fulltext/ED572410.pdf

[41] A. W. Kohar, E. B. Rahaju, and A. Rohim, "Prospective teachers' design of numeracy tasks using a physical distancing context," *J. Math. Educ.*, vol. 13, no. 2, pp. 191–210, 2022, doi: 10.22342/jme.v13i2.pp191-210.

[42] A. M. B. Harsono, R. C. Murti, and R. D. Cahya, "Hubungan keterampilan 4C dan kemampuan literasi numerasi peserta didik hasil belajar matematika," *Aksioma*, vol. 12, no. 3, pp. 3299–3308, 2023, doi: 10.24127/ajpm.v12i3.7162.

[43] D. S. Nusantara, Zulkardi, and R. I. I. Putri, "Designing PISA-like mathematics problem relating change and relationship using physical distancing context," *J. Phys. Conf. Ser.*, vol. 1663, no. 1, 2020, doi: 10.1088/1742-6596/1663/1/012004.

[44] D. Yansen, R. I. I. Putri, Zulkardi, and S. Fatimah, "Developing PISA-like mathematics problems on uncertainty and data using asian games football context," *J. Math. Educ.*, vol. 10, no. 1, pp. 37–46, 2019, doi: 10.22342/jme.10.1.5249.37-46.

[45] A. Aspandi and M. A. Muttaqin, "Transforming Teacher Roles in Indonesia ' s Digital Era : Enhancing Learning Effectiveness and Student Engagement," vol. 4, no. 4, pp. 1495–1509, 2025.

[46] A. N. Cahyono, Masrukan, Mulyono, M. Ludwig, S. Jablonski, and D. X. K. Oehler, "Indonesia-Germany MathCityMap training: Shifting mobile math trails teacher training to a hybrid environment," *J. Math. Educ.*, vol. 14, no. 1, pp. 55–68, 2023, doi: 10.22342/JME.V14I1.PP55-68.

[47] A. N. Cahyono, Y. L. Sukestiyarno, M. Asikin, Miftahudin, M. G. K. Ahsan, and M. Ludwig, "Learning mathematical modelling with augmented reality mobile math trails program: How can it work?," *J. Math. Educ.*, vol. 11, no. 2, pp. 181–192, 2020, doi: 10.22342/jme.11.2.10729.181-192.

[48] S. Jablonski, "Students' verification and elaboration in outdoor mathematics: The role of digital feedback in MathCityMap," *Digit. Exp. Math. Educ.*, vol. 10, no. 1, pp. 132–157, 2024, doi: 10.1007/s40751-024-00137-w.