International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 3, June 2026, pp. 2608~2617
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i3.39078

Illuminative evaluation of mathematics curriculum implementation in improving students' numeracy achievement

Jose Bonatua Hasibuan[1], Deni Darmawan[1], Suhendra[2], Deni Kurniawan[1]
[1]Faculty of Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[2]Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Article Info

Article history:

Received Feb 11, 2026
Revised May 5, 2026
Accepted May 15, 2026

Keywords:

Curriculum implementation
Illuminative evaluation
Mathematics curriculum
Mixed methods
Numeracy achievement

ABSTRACT

Persistent evidence from national and international assessments indicates that students' numeracy achievement remains low, suggesting a gap between intended curriculum goals and classroom implementation. This study conducts an illuminative evaluation of secondary mathematics curriculum implementation as an instructional system, examining how curriculum enactment relates to students' numeracy achievement in the Indonesian secondary context. Employing a sequential explanatory mixed-methods design, the quantitative phase assessed the numeracy performance of 288 secondary students across content domains, cognitive levels, and item formats, while the qualitative phase investigated instructional practices, assessment culture, and the school learning milieu to explain the observed achievement patterns. The findings indicate uneven numeracy achievement, with performance concentrated at procedural levels and declining from knowing to applying and reasoning. Students perform relatively better on objective formats but demonstrate a limited ability to justify solutions in open-ended tasks. Qualitative evidence further indicates a misalignment between reasoning-oriented curriculum intentions and efficiency-driven classroom practices that emphasize procedural accuracy. These findings provide evidence-based insights into aligning curriculum design, classroom instruction, and assessment practices to strengthen reasoning-oriented numeracy learning.

Corresponding Author:

Jose Bonatua Hasibuan
Faculty of Education, Universitas Pendidikan Indonesia
Dr. Setiabudhi Street, Bandung, Indonesia
Email: josehasibuan@upi.edu

1. INTRODUCTION

Strengthening numeracy has become a central agenda in global education, particularly within the sustainable development goals framework, where quality education is viewed as a driver of social and economic development [1], [2]. Numeracy is no longer understood merely as an arithmetic skill but as the capacity to interpret data, reason quantitatively, model real-world situations, and make evidence-based decisions in complex contexts [3], [4]. In contemporary mathematics education, numeracy is therefore conceptualized as involving modeling, interpreting multiple representations, and engaging in reflective reasoning rather than routine procedural execution.

Despite this global emphasis on numeracy, international and national assessments continue to reveal challenges in Indonesian students' numeracy achievement. Results from Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA) consistently indicate performance below the international average, particularly on tasks requiring reasoning, interpretation, and application rather than routine procedures [5]–[7]. National data from the national assessment and the education report card (*rapor pendidikan*) similarly show that although minimum competency levels have gradually improved, a large proportion of students remain below expected numeracy standards, with substantial regional disparities. Empirical studies further report that students perform relatively better on routine tasks but experience difficulty with contextual problems, data interpretation, and higher-order reasoning demands [8]–[10].

These patterns suggest that numeracy achievement cannot be explained solely by individual cognitive factors but is closely related to instructional and assessment conditions in classroom contexts [11]. In many secondary mathematics classrooms, learning remains dominated by procedural exercises, textbook-oriented instruction, and efficiency-driven assessment practices [12], particularly objective and multiple-choice formats of assessment that prioritize correct answers and time efficiency. This orientation creates a structural misalignment between the intended curriculum, which emphasizes reasoning and contextual numeracy, and the enacted curriculum experienced by students in daily classroom practice [11], [12]. Consequently, students have limited opportunities for modeling, justification, and written reasoning—key components of higher-order numeracy [13]. Existing numeracy studies in Indonesia often examine PISA-type tasks, specific cognitive domains, or isolated instructional interventions. However, relatively little research investigates how curriculum implementation, instructional practices, and assessment culture interact as a systemic process shaping numeracy achievement in authentic classroom contexts.

To address this gap, this study conducts an illuminative evaluation of secondary mathematics curriculum implementation and examines how curriculum enactment relates to students' numeracy achievement in classroom contexts. Specifically, the study: i) analyzes students' numeracy performance across content domains, cognitive levels, and item formats and ii) links these patterns with qualitative evidence on instructional planning, teaching practices, and assessment culture within the school learning milieu [14]. A sequential explanatory mixed-methods design integrates quantitative achievement structures with qualitative investigation to interpret numeracy outcomes as products of instructional processes rather than solely indicators of individual ability [15]–[18].

In the Indonesian context, these issues relate closely to national education policy. Law No. 20/2003 positions education as a means to develop intellectual capacity, character, and life skills, while curriculum plays a central role in achieving these goals [19]. The implementation of Kurikulum Merdeka further emphasizes contextual learning, interdisciplinary approaches, and the strengthening of literacy and numeracy competencies. Within this reform, mathematics learning is framed not only as content mastery but also as reasoning, representation, and problem solving, reflecting perspectives that emphasize alignment among curriculum, pedagogy, and assessment to support higher-order competencies in authentic classroom contexts [11]. This study adopts the illuminative evaluation framework proposed by Parlett and Hamilton [15], which emphasizes interactions among the instructional system, the learning milieu, and students' learning experiences in real classroom contexts. Within this perspective, curriculum implementation is understood as a lived instructional ecology rather than merely a formal policy document, and numeracy achievement is interpreted as an outcome shaped by enacted curriculum, assessment culture, and classroom practice.

Building on this perspective, the study contributes to numeracy and curriculum research in three ways. First, it provides an illuminative evaluation linking curriculum enactment, instructional practices, and assessment culture with the structures of students' numeracy achievement in Indonesian secondary education. Second, it integrates quantitative achievement analysis with qualitative evidence on instructional practice and the learning milieu, offering a systemic interpretation of numeracy outcomes beyond individual cognitive performance. Third, the findings provide policy-relevant insights for strengthening alignment between curriculum intentions, classroom instruction, and assessment systems within reforms such as Kurikulum Merdeka and national assessment policies. Based on this framework, the study addresses the following research questions:

- How is students' numeracy achievement structured across content domains, cognitive levels, and item formats in secondary schools?
- How is the mathematics curriculum implemented as an instructional system within the school learning milieu?
- How does curriculum implementation explain the observed patterns of students' numeracy achievement from an illuminative evaluation perspective?

2. METHOD

2.1. Research design and procedures

This study employed a sequential explanatory mixed-methods design, integrating quantitative and qualitative phases within an illuminative evaluation perspective. The research was conducted during the second semester of the 2024–2025 academic year in secondary schools in Riau Province, Indonesia, which represents diverse contexts in terms of school type (general and vocational), geographic distribution, and variation in educational resources. Consequently, the findings are interpreted within this provincial context and may not be generalized to other regions.

The quantitative phase mapped students' numeracy competence and identified achievement patterns requiring deeper explanation. Based on these results, two schools with the highest average numeracy scores were purposively selected for the qualitative phase using criteria of numeracy performance, representation of school types, and institutional willingness to participate in classroom observations and interviews. Although this approach enabled an in-depth illuminative investigation of functioning instructional systems and curriculum enactment, the purposive selection may limit representativeness and introduce contextual bias.

Participants in the qualitative phase included mathematics teachers, students who had participated in the numeracy test, and vice principals responsible for curriculum coordination. Student participants were enrolled in selected intact classes, present during the test administration, and completed the numeracy test under standardized conditions. Teacher respondents were certified mathematics teachers actively teaching at the secondary level in the sampled districts. Schools involved in the qualitative phase agreed to participate in classroom observations, interviews, and document analysis. Data collection emphasized natural classroom contexts and focused on instructional planning, teaching practices, assessment implementation, and students' learning experiences within the school learning milieu.

The study proceeded in four stages: i) quantitative measurement of students' numeracy achievement and teacher perceptions; ii) a bridging phase using open-ended teacher responses to identify instructional practices and triangulate survey findings; iii) qualitative data collection through interviews, classroom observations, and document analysis; and iv) integrative interpretation linking quantitative achievement patterns with qualitative explanations to understand curriculum implementation as an instructional system within the learning milieu. Because the design was cross-sectional, the findings represent conditions at a single time point rather than longitudinal development. Ethical procedures included informed consent, voluntary participation, confidentiality assurance, and anonymized data coding for research purposes.

2.2. Data acquisition and analysis

The quantitative sample consisted of 288 secondary students from senior high schools and vocational schools selected through cluster sampling. Schools were grouped by type and geographic location, clusters were proportionally selected to represent both categories, and intact classes served as sampling units. In addition, 152 mathematics teachers from twelve districts and cities completed a questionnaire addressing curriculum development and numeracy-oriented instructional practices.

The numeracy test consisted of thirty items across three content domains—algebra, geometry and measurement, and data and uncertainty—and three cognitive levels: knowing, applying, and reasoning, reflecting contemporary perspectives that distinguish procedural knowledge from higher-order reasoning and mathematical justification [20], [21]. Six item formats were included: multiple choice, multiple-complex, matching, short answer, true–false, and open-ended items. Open-ended tasks were included to elicit students' written reasoning, which is central to reflective numeracy competence [8], [22]. Objective items were scored dichotomously, while open-ended responses were assessed using an analytic rubric evaluating completeness, accuracy, and coherence of reasoning. Scores were converted to a 0–100 scale, and the test was administered in classrooms for ninety minutes under standardized supervision.

Content validity was established through expert judgment by three mathematics education specialists. Empirical validity was examined through pilot testing using item–total correlations, discrimination indices, and distractor effectiveness analysis. Internal consistency reliability was assessed using Cronbach's alpha, indicating acceptable reliability (α=0.82). The teacher questionnaire consisted of Likert-scale and open-ended items describing instructional practices and perceived best practices in numeracy-oriented teaching, with open-ended responses functioning as bridging data linking quantitative and qualitative phases.

Quantitative data were analyzed using descriptive and inferential statistics. Normality testing with the Shapiro–Wilk test indicated non-normal distributions across several domains and cognitive levels; therefore, non-parametric tests were primarily employed. Measures of central tendency and dispersion were calculated, followed by Friedman tests for comparisons across domains and cognitive levels and Wilcoxon signed-rank tests with Bonferroni correction for post-hoc analysis. Two-way analysis of variance (ANOVA) was additionally applied for selected comparisons by school type and geographic location due to its robustness to moderate normality violations in large samples. All analyses were conducted at a significance level of 0.05 using IBM SPSS statistics version 26. Effect sizes complemented statistical significance, with Kendall's W reported for Friedman tests, effect sizes ($r=Z/\sqrt{N}$) for Wilcoxon tests, and partial eta squared (η^2p) for ANOVA.

The qualitative phase was conducted in the two purposively selected schools. Participants included mathematics teachers, students who had participated in the numeracy test, and vice principals responsible for curriculum coordination. Classroom observations were conducted in repeated cycles; in each school, one grade X and one grade XI class were observed across three instructional meetings. Observations were carried out by the principal researcher and a trained research assistant using a structured protocol addressing instructional strategies, teacher–student interaction, task characteristics, and assessment practices. Field notes were independently documented and compared after each session, and interview recordings were transcribed verbatim.

Qualitative data were analyzed using thematic analysis combining inductive coding with a theoretically informed framework derived from illuminative evaluation [15], following established thematic analysis procedures [23]. The analysis involved familiarization with transcripts and documents, open coding, thematic categorization, and interpretive synthesis guided by the constructs of instructional system, learning milieu, and students' learning experiences. Trustworthiness was ensured through source triangulation (teachers, students, and documents), technique triangulation (interviews, observations, and document analysis), peer debriefing, audit trail maintenance, and member checking with selected participants.

3. RESULTS AND DISCUSSION

The findings of this study are presented through an integrated interpretation of quantitative and qualitative evidence within a sequential explanatory mixed-methods design. Quantitative results provide the structural pattern of students' numeracy achievement, while qualitative findings explain how instructional practices, assessment culture, and the school learning milieu shape those outcomes. The analysis is organized around five major dimensions: overall achievement profile, numeracy domains, cognitive levels, item formats, and curriculum implementation as an instructional system.

3.1. Profile of students' numeracy achievement

The numeracy test administered to 288 secondary students in Riau Province reveals substantial variation in competence, with overall performance concentrated at the lower end of the achievement distribution. Scores range from 0.00 to 81.82, and both the mean (29.18) and median (27.27) fall well below the midpoint of the scale. The interquartile spread further indicates considerable disparities in student performance, suggesting that numeracy attainment remains uneven across learners. This bottom-heavy distribution indicates that mastery-level numeracy competence remains limited, with performance concentrated at foundational rather than advanced levels. Patterns of wide dispersion combined with low central tendency have similarly been reported in studies of mathematical literacy and PISA-type performance, where a small group of higher achievers coexists with a large proportion of students operating at basic procedural levels [24], [25].

Grouping the scores into proficiency categories reveals the pattern more clearly. More than three-quarters of students (77.43%) fall into the very low and low proficiency bands, whereas fewer than 9% reach the high or very high categories. This bottom-heavy distribution indicates that numeracy difficulties are systemic rather than limited to a small subgroup. The overall pattern reflects a structure of competence where procedural familiarity may exist for some learners, but higher-level application and reasoning skills remain underdeveloped. Such a distribution mirrors broader assessment trends indicating that students tend to perform better on routine procedures than on tasks involving modeling, data interpretation, or complex reasoning [1], [3], [26].

The study examined the normality of the overall numeracy score distribution using the Shapiro–Wilk test. The results indicated a significant deviation from normality ($p<0.001$), thereby justifying the use of non-parametric inferential analyses in subsequent sections. Substantively, the non-normal distribution reflects a clustering of performance at the lower end with only a small proportion of higher achievers, reinforcing the conclusion that numeracy proficiency remains limited and unevenly distributed. Similar skewed achievement patterns have been observed in large-scale assessments, where performance gaps between basic and advanced proficiency levels reflect structural differences in exposure to reasoning-oriented tasks [24], [27]. These baseline findings establish the structural profile of students' numeracy competence and provide the foundation for interpreting differences across domains, cognitive levels, assessment formats, and curriculum implementation in the following sections.

3.2. Numeracy achievement across content domains and cognitive levels

Analysis across content domains reveals a hierarchical pattern in students' numeracy achievement. The geometry and measurement domain shows the highest central tendency, followed by algebra, while data and uncertainty records the lowest performance. However, the relative advantage in geometry and measurement appears to reflect procedural familiarity rather than strong conceptual or representational understanding [13]. Even the comparatively stronger domain remains within a generally low proficiency range when interpreted against contemporary numeracy expectations emphasizing reasoning and contextual interpretation rather than routine computation [3], [7]. Similar domain-level patterns have been observed in studies of mathematical literacy, where geometry tasks often support formula-based procedures while data-related tasks require higher levels of interpretation and contextual reasoning [1], [24].

Qualitative evidence helps explain these domain differences. Classroom observations and student interviews indicate that instruction in geometry and measurement frequently emphasizes formula application through repetitive practice. Students report confidence when solving routine tasks but difficulty when problems require visual interpretation, contextual modeling, or non-standard reasoning. Thus, the quantitative advantage in this domain reflects procedural consolidation reinforced by classroom practice and assessment formats prioritizing formula-based problem solving. Research on geometry learning similarly suggests that procedural fluency in formula use does not necessarily translate into spatial reasoning or representational competence [28].

Algebra demonstrates moderate but unstable achievement. Document analysis and teacher interviews suggest that algebra instruction often alternates between conceptual explanation and procedural symbol manipulation. Students exposed to varied representations tend to perceive algebra as manageable, whereas others describe it as cognitively demanding due to symbolic density. This variability explains why aggregate statistics mask differences in students' learning experiences. Previous studies similarly indicate that success in algebra depends strongly on representational flexibility and conceptual connections rather than symbol manipulation alone [9], [29].

The data and uncertainty domain show the lowest achievement among all domains. Quantitative results show both the lowest central tendency and the greatest dispersion. Qualitative evidence indicates that students' main difficulties lie not in statistical computation but in the interpretation of textual stimuli, tables, and graphs. Classroom instruction rarely emphasizes data interpretation, inferential reasoning, or decision making under uncertainty [14], even though these competencies are central to contemporary numeracy frameworks in international assessments [1], [30], [31]. Statistical literacy research similarly shows that students often master procedures but have difficulty interpreting variability, uncertainty, and contextual meaning in data [32], [33].

A similar hierarchical structure also appears across cognitive levels. Performance declines consistently from knowing to applying and further to reasoning, indicating that students' achievement decreases as cognitive demand increases. Wilcoxon signed-rank tests confirm that differences among cognitive levels are statistically significant. The gap between knowing and applying is significant ($Z=10.507$, $p<0.001$, $r=0.619$), while the difference between knowing and reasoning is even larger ($Z=12.792$, $p<0.001$, $r=0.754$). The comparison between applying and reasoning also remains significant ($Z=6.927$, $p<0.001$, $r=0.408$). These effect sizes indicate that the decline in performance across cognitive levels is both statistically and educationally meaningful.

Qualitative evidence suggests that this hierarchy is closely related to classroom instructional practices. At the knowing level, students demonstrate relatively stronger performance because instruction frequently emphasizes recall, formula application, and routine procedures. Observations and interviews show that learning activities often focus on memorizing definitions, algorithmic steps, and standard solution patterns. Although this approach supports procedural fluency, it does not always promote conceptual explanation or flexible problem solving [10], [20].

At the applying level, students must transfer knowledge to contextual situations and select appropriate strategies. However, qualitative evidence indicates that contextual modeling activities and real-world problem solving are relatively limited in classroom practice. As a result, students often experience uncertainty when confronted with non-routine tasks involving verbal stimuli, contextual scenarios, or multiple representations [24], [30].

The lowest performance appears at the reasoning level. Students rarely encounter tasks requiring justification, argumentation, or evaluation of alternative solutions during regular instruction. Classroom discussions often conclude once the correct answer is obtained, with little emphasis on explaining reasoning processes or constructing mathematical arguments. Teachers also report time constraints and assessment practicality as factors limiting the use of open-ended reasoning tasks. Consequently, students have limited opportunities to develop reflective numeracy competence and mathematical argumentation skills [21], [22]. From an illuminative evaluation perspective, this hierarchy reflects the interaction between the instructional system and the learning milieu, where procedural knowledge is reinforced more strongly than higher-order reasoning [15].

3.3. Achievement by item format and assessment culture

Analysis by item format provides further insight into the relationship between assessment practices and numeracy performance. Multiple-choice items produce the highest and most stable scores. Students perceive this format as familiar and manageable, reflecting its dominance in routine classroom assessments. Teacher interviews also indicate that the frequent use of multiple-choice questions is largely motivated by administrative efficiency. Similar patterns have been reported in assessment research, where objective formats tend to favor recognition and procedural recall rather than deeper reasoning or explanation [2], [34]–[36].

Short-answer and matching items yield moderate performance. These formats typically assess direct conceptual recall or straightforward computation and rarely require extended explanation or interpretation. Consequently, they capture partial conceptual knowledge but provide limited opportunities for students to demonstrate reasoning processes or argumentation [20].

Performance declines in more complex formats such as multiple-complex and true–false items that require evaluating several statements simultaneously. Students report cognitive overload when completing these tasks, and teachers acknowledge that such formats are used less frequently in classroom assessments. Lower scores therefore reflect both higher cognitive demand and limited exposure. Research on assessment design similarly indicates that these formats require stronger metacognitive and analytical reasoning skills that procedural-oriented classrooms seldom cultivate [37].

Open-ended items record the lowest achievement levels. Students often struggle to articulate step-by-step solutions and written mathematical reasoning, while teachers report that such tasks are used infrequently due to the time required for scoring and evaluation. The limited use of open-ended tasks restricts opportunities for students to engage in mathematical communication and justification, which are essential components of reflective numeracy competence and higher-order reasoning [8], [21], [35], [36]. International assessment frameworks likewise emphasize written explanation and justification as indicators of advanced mathematical proficiency [2]. Overall, these patterns indicate that assessment culture plays a central role in shaping numeracy performance. When classroom assessments emphasize efficiency and correct answers rather than explanation and reasoning, students receive fewer opportunities to develop higher-order numeracy skills, a pattern widely discussed in research on formative assessment in mathematics education [38].

3.4. Curriculum implementation and integrated interpretation

Teacher questionnaires indicate that, at a declarative level, teachers perceive numeracy-oriented instruction and assessment positively. They report aligning objectives with numeracy competencies and recognizing the importance of contextual learning. However, qualitative data reveal a systematic gap between these perceptions and classroom enactment [12]. Previous curriculum implementation research widely documents such discrepancies between stated beliefs and enacted practice [34].

Open-ended responses often describe practices using general terms without detailed pedagogical explanation. Observations show that instruction frequently returns to expository methods and procedural exercises under time pressure. This reflects a situation in which curricular innovation is normatively accepted but only partially realized in practice, consistent with illuminative evaluation's emphasis on the gap between formal intentions and lived classroom reality [15].

The learning milieu mediates this process. Even in schools with supportive academic cultures, reasoning-focused pedagogy is not consistently institutionalized. Assessment culture remains oriented toward efficiency and correctness, reinforcing students' perception that mathematics learning primarily involves obtaining correct answers rather than constructing arguments. Research on classroom discourse similarly shows that when correctness is prioritized over explanation, opportunities for developing mathematical reasoning and communication are significantly reduced [8], [21].

Students' learning experiences provide the most direct evidence of implementation quality. Students express comfort with routine exercises but uncertainty in contextual or reasoning tasks. They seldom encounter structured opportunities to justify solutions or interpret complex representations. These experiences align closely with the quantitative hierarchy of cognitive performance and reflect patterns observed in studies where limited exposure to reasoning-oriented tasks constrains students' ability to engage in higher-order numeracy practices [1], [25], [30].

Taken together, these findings indicate that students' numeracy achievement is shaped by the interaction among the instructional system, assessment practices, and the school learning milieu [14]. From an illuminative evaluation perspective, numeracy outcomes should therefore be interpreted not solely as indicators of individual cognitive ability but as manifestations of how curriculum policies are enacted in everyday classroom practice [39]. Although curriculum policy emphasizes contextual problem solving and mathematical reasoning, classroom instruction and assessment practices frequently continue to prioritize procedural proficiency, answer-oriented performance, and rapid task completion. This misalignment helps explain the persistent decline in performance from knowing to applying and especially to reasoning, a pattern widely documented in studies of curriculum implementation and mathematics education reform in contexts where assessment systems remain predominantly procedural [27], [34].

Comparable patterns have been reported across diverse educational systems, indicating a broader systemic challenge rather than a context-specific anomaly. Research on large-scale assessment and classroom instruction consistently shows that when teaching and evaluation systems emphasize procedural accuracy over conceptual explanation and modeling, students’ higher-order reasoning competencies remain underdeveloped [24], [25]. In the present study, the dominance of objective assessment formats and routine procedural tasks reinforces a classroom culture that values correctness and efficiency over justification and mathematical communication, thereby limiting opportunities for reasoning-oriented learning.

This illuminative synthesis demonstrates that numeracy outcomes are embedded in instructional culture and shaped by interactions among curriculum structures, instructional practices, and students’ classroom experiences [15]. Qualitative evidence further indicates that teachers’ declarative alignment with numeracy-oriented curriculum goals does not consistently translate into enacted pedagogical practices. Instead, curriculum implementation is mediated by contextual factors such as assessment traditions, time constraints, and institutional expectations for instructional efficiency. As a result, the enacted curriculum tends to support procedural mastery more strongly than reflective reasoning, which helps explain the systematic decline in performance across cognitive levels.

Strengthening numeracy therefore requires systemic alignment among instructional design, classroom discourse, and assessment practices toward reasoning, modeling, and mathematical communication rather than isolated instructional adjustments. This finding reinforces contemporary perspectives that position explanation, justification, and contextual interpretation as core components of mathematical proficiency and reflective numeracy competence [21], [30]. These results also provide policy-relevant implications for the implementation of Kurikulum Merdeka and ongoing national assessment reforms, particularly the need to align classroom assessment practices with reasoning-oriented numeracy competencies.

Finally, this study contributes to educational evaluation and numeracy research in three ways. First, it applies an illuminative evaluation perspective to conceptualize numeracy achievement as an outcome of an instructional system rather than solely an individual cognitive attribute. Second, it provides integrated mixed-method evidence linking achievement structures with assessment culture, curriculum enactment, and the learning milieu in authentic secondary school contexts. Third, it demonstrates that persistent reasoning gaps are systematically shaped by procedural-oriented instructional practices and assessment traditions, offering empirically grounded insights for curriculum implementation, teacher professional development, and assessment reform.

4. CONCLUSION

This study applies an illuminative evaluation perspective to examine how the implementation of the secondary mathematics curriculum relates to students’ numeracy achievement. The integration of quantitative achievement patterns with qualitative evidence on instructional practices, assessment culture, and the school learning milieu indicates a consistent decline in performance from procedural knowledge to higher-order reasoning. Students demonstrate relatively stronger performance on routine and formula-based tasks but encounter difficulties in contextual interpretation, modeling, data reasoning, and written justification. These patterns highlight a systemic misalignment between curriculum intentions that emphasize reasoning-oriented numeracy and classroom practices that continue to prioritize procedural fluency and answer-oriented assessment. By linking curriculum enactment with observed achievement structures, this study contributes a systemic perspective to numeracy research, showing that numeracy outcomes are shaped by the interaction among instructional systems, assessment practices, and students’ classroom learning experiences.

The findings provide policy-relevant implications for the implementation of Kurikulum Merdeka and ongoing national assessment reforms. Strengthening reasoning-oriented tasks, formative assessment, and classroom discourse emphasizing explanation and mathematical communication is essential to align curriculum intentions with classroom practices and national numeracy goals. Future research should extend illuminative evaluation to broader regional contexts and examine longitudinal changes in instructional practices and numeracy development. Investigating diverse school settings and integrating classroom discourse analysis, task design studies, and large-scale assessment data with qualitative inquiry would further strengthen systemic evaluation of mathematics curriculum implementation.

FUNDING INFORMATION

The authors state that no funding was involved in this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Jose Bonatua Hasibuan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Deni Darmawan	✓	✓								✓				
Suhendra	✓	✓								✓				
Deni Kurniawan	✓	✓								✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [JBH], upon reasonable request. The data are not publicly available due to restrictions related to research ethics and participant privacy.

REFERENCES

[1] The Organisation for Economic Co-operation and Development (OECD), *PISA 2018 results (volume I): what students know and can do*. Paris: OECD Publishing, 2019, doi: 10.1787/5f07c754-en.

[2] The Organisation for Economic Co-operation and Development (OECD), *PISA 2022 results (volume I): learning in the digital world*. Paris: OECD Publishing, 2023, doi: 10.1787/7f6d1f92-en.

[3] A. Hidayatullah and C. Csíkos, "Mathematics related belief system and word problem-solving in the Indonesian context," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 18, no. 4, p. em2094, Mar. 2022, doi: 10.29333/ejmste/11902.

[4] B. Trilling and C. Fadel, *21st century skills: learning for life in our times*. San Francisco, CA: Jossey-Bass, 2009.

[5] L. Linda and I. Asyura, "Students' mathematical reasoning ability in solving post-COVID-19 PISA model math problems," *Jurnal Riset Pendidikan Matematika*, vol. 8, no. 2, pp. 140–152, Nov. 2021, doi: 10.21831/jrpm.v8i2.44739.

[6] I. Nurazizah and Z. Zulkardi, "Students' mathematical reasoning ability in solving PISA-like mathematics problem COVID-19 context," *Jurnal Elemen*, vol. 8, no. 1, pp. 250–262, Jan. 2022, doi: 10.29408/jel.v8i1.4599.

[7] A. Rivai, A. Lestari, N. P. Munir, and A. Anas, "Students' mathematical literacy in solving PISA problems observed by learning styles," *Jurnal Pendidikan Matematika*, vol. 17, no. 1, pp. 121–134, Dec. 2022, doi: 10.22342/jpm.17.1.19905.121-134.

[8] M. S. Mahmud and N. F. M. Drus, "The use of oral questioning to improve students' reasoning skills in primary school mathematics learning," *Frontiers in Education*, vol. 8, p. 1126816, May 2023, doi: 10.3389/feduc.2023.1126816.

[9] F. A. Ningsih, R. Lefrida, I. N. Murdiana, and B. Mallo, "Analysis of students' errors in solving problems involving algebraic expressions," *JMPM: Jurnal Matematika dan Pendidikan Matematika*, vol. 9, no. 2, pp. 235–247, Feb. 2025, doi: 10.26594/jmpm.v9i2.3509.

[10] A. H. Dewantara, F. A. Setiawati, and S. Saraswati, "Towards numeracy literacy development: a single-case study on the use of the living book homeschooling model," *Infinity Journal*, vol. 12, no. 2, pp. 225–242, 2023, doi: 10.22460/infinity.v12i2.p225-242.

[11] F. M. Zana, C. Sa'dijah, Susiswo, L. Anwar, and H. Zulnaidi, "Curriculum and teacher assessment practices in mathematics learning: alignment with higher order thinking skills in Indonesian secondary schools," *Journal on Mathematics Education*, vol. 15, no. 4, pp. 1311–1334, 2024, doi: 10.22342/jme.v15i4.pp1311-1334.

[12] T. N. H. Yunianta, D. Suryadi, D. Dasari, and T. Herman, "Textbook praxeological-didactical analysis: lessons learned from the Indonesian mathematics textbook," *Journal on Mathematics Education*, vol. 14, no. 3, pp. 503–524, 2023, doi: 10.22342/jme.v14i3.pp503-524.

[13] F. Nursyahidah, Wardono, S. Mariani, and K. Wijayanti, "Integrating technology, Javanese ethnomathematics, and realistic mathematics education in supporting prospective mathematics teachers' numeracy skills: a learning trajectory," *Journal on Mathematics Education*, vol. 16, no. 2, pp. 671–688, 2025, doi: 10.22342/jme.v16i2.pp671-688.

[14] V. Adelia, R. I. I. Putri, Zulkardi, and M. Doorman, "The development of web-based learning environment to enhance students' numeracy and reasoning," *Journal on Mathematics Education*, vol. 16, no. 2, pp. 651–670, 2025, doi: 10.22342/jme.v16i2.pp651-670.

[15] M. Parlett and D. Hamilton, "Evaluation as illumination: a new approach to the study of innovatory programs," Centre for Research in the Educational Sciences, University of Edinburgh, Edinburgh, U.K., 1972.

[16] R. Sims, H. Banat, P. Tran, P. Panahi, and B. Dilger, "Illuminative evaluation of an intercultural-competence-focused first-year writing curriculum," *Writing & Pedagogy*, vol. 14, no. 1, pp. 103–127, Sep. 2022, doi: 10.1558/wap.21124.

[17] C. Fourie, I. L. Potgieter, and M. Coetzee, "A qualitative illuminative evaluation of emotional intelligence training as transformative learning," *SA Journal of Industrial Psychology*, vol. 50, no. 1, pp. 1–13, Dec. 2024, doi: 10.4102/sajip.v50i0.2243.

[18] M. J. Gunio, "Determining the influences of a hidden curriculum on students' character development using the illuminative evaluation model," *Journal of Curriculum Studies Research*, vol. 3, no. 2, pp. 194–206, Nov. 2021, doi: 10.46303/jcsr.2021.11.

[19] F. Rahmawati and H. Retnawati, "An analysis of students' difficulties in solving PISA-like mathematical problems," *Journal of Physics: Conference Series*, vol. 1200, no. 1, p. 012015, Mar. 2019, doi: 10.1088/1742-6596/1200/1/012015.

[20] J. Kilpatrick, J. Swafford, and B. Findell, *Adding it up: helping children learn mathematics*. Washington, D.C.: The National Academies Press, 2001, doi: 10.17226/9822.

[21] A. Valenta, K. Rø, and S. I. Klock, "A framework for reasoning in school mathematics: analyzing the development of mathematical claims," *Educational Studies in Mathematics*, vol. 116, no. 1, pp. 91–111, May 2024, doi: 10.1007/s10649-024-10309-5.

[22] D. Kollosche, "Styles of reasoning for mathematics education," *Educational Studies in Mathematics*, vol. 107, no. 3, pp. 471–486, Jul. 2021, doi: 10.1007/s10649-021-10046-z.

[23] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.

[24] X. S. Wang, L. B. Perry, A. Malpique, and T. Ide, "Factors predicting mathematics achievement in PISA: a systematic review," *Large-Scale Assessments in Education*, vol. 11, no. 1, p. 24, Jun. 2023, doi: 10.1186/s40536-023-00174-8.

[25] L. H. Muhaimin, R. A. Sholikhakh, S. Yulianti, Ardani, A. Hendriyanto, and S. Sahara, "Unlocking the secrets of students' mathematical literacy to solve mathematical problems: a systematic literature review," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 20, no. 4, p. em2428, Apr. 2024, doi: 10.29333/ejmste/14404.

[26] M. Jannah, A. Pesik, S. Kumesan, E. Ermita, and N. Rahim, "Analysis of student errors in solving equations and inequalities problems in algebra and trigonometry based on Newman procedure," *Journal of Education and Teaching Learning (JETL)*, vol. 6, no. 1, pp. 25–40, Jan. 2024, doi: 10.51178/jetl.v6i1.1616.

[27] The Organisation for Economic Co-operation and Development (OECD), *PISA 2022 results (volume I): the state of learning and equity in education*. Paris: OECD Publishing, 2023, doi: 10.1787/53f23881-en.

[28] A. Andira, M. Darwis M, and H. Syam, "Exploring of students' ability to solve geometry problems based on Van Hiele's level of thinking," *ETDC: Indonesian Journal of Research and Educational Review*, vol. 1, no. 2, pp. 192–200, Mar. 2022, doi: 10.51574/ijrer.v1i2.181.

[29] C. Dorner, C. Ableitinger, and G. Krammer, "Revealing the nature of mathematical procedural knowledge by analysing students' deficiencies and errors," *International Journal of Mathematical Education in Science and Technology*, vol. 57, no. 3, pp. 413–434, Mar. 2025, doi: 10.1080/0020739X.2024.2445666.

[30] The Organisation for Economic Co-operation and Development (OECD), *PISA 2022 assessment and analytical framework*. Paris: OECD Publishing, 2023, doi: 10.1787/dfe0bf9c-en.

[31] J. Engel, "Statistical literacy for active citizenship: a call for data science education," *Statistics Education Research Journal*, vol. 16, no. 1, pp. 44–49, May 2017, doi: 10.52041/serj.v16i1.213.

[32] A. B. Kurnia, T. Lowrie, and S. M. Patahuddin, "The development of high school students' statistical literacy across grade level," *Mathematics Education Research Journal*, vol. 36, no. Suppl 1, pp. 7–35, Mar. 2024, doi: 10.1007/s13394-023-00449-x.

[33] C. Batanero and R. Álvarez-Arroyo, "Teaching and learning of probability," *ZDM - Mathematics Education*, vol. 56, no. 1, pp. 5–17, Feb. 2024, doi: 10.1007/s11858-023-01511-5.

[34] F. M. Zana, C. Sa'dijah, and Susiswo, "LOTS to HOTS: how do mathematics teachers improve students' higher-order thinking skills in the class?" *International Journal of Trends in Mathematics Education Research*, vol. 5, no. 3, pp. 251–260, Sep. 2022, doi: 10.33122/ijtmer.v5i3.143.

[35] S. Breuer, T. Scherndl, and T. M. Ortner, "Effects of response format on achievement and aptitude assessment results: multi-level random effects meta-analyses," *Royal Society Open Science*, vol. 10, no. 5, p. 220456, May 2023, doi: 10.1098/rsos.220456.

[36] K. Marcq, E. J. C. Donayre, and J. Braeken, "The role of item format in the PISA 2018 mathematics literacy assessment: a cross-country study," *Studies in Educational Evaluation*, vol. 83, p. 101401, Dec. 2024, doi: 10.1016/j.stueduc.2024.101401.

[37] M. V. J. Veenman, B. H. A. M. van Hout-Wolters, and P. Afflerbach, "Metacognition and learning: conceptual and methodological considerations," *Metacognition and Learning*, vol. 1, no. 1, pp. 3–14, Apr. 2006, doi: 10.1007/s11409-006-6893-0.

[38] K. Maskos, A. Schulz, S. S. Oeksuez, and K. Rakoczy, "Formative assessment in mathematics education: a systematic review," *ZDM – Mathematics Education*, vol. 57, no. 4, pp. 679–693, Aug. 2025, doi: 10.1007/s11858-025-01696-x.

[39] V. Geiger and M. Schmid, "A critical turn in numeracy education and practice," *Frontiers in Education*, vol. 9, p. 1363566, Apr. 2024, doi: 10.3389/feduc.2024.1363566.

BIOGRAPHIES OF AUTHORS

Jose Bonatua Hasibuan received his Doctor of Education degree from the Department of Curriculum Development, Faculty of Education, Universitas Pendidikan Indonesia, Indonesia. He is a civil servant serving as a mathematics teacher under the Riau Provincial Education Office, Indonesia. He earned his bachelor's degree in mathematics education and his master's degree in applied mathematics. His research focuses on mathematics learning, literacy and numeracy, as well as curriculum development and evaluation. He can be contacted at email: josehasibuan@upi.edu.

Deni Darmawan is a professor of Educational Technology and Communication Science at Universitas Pendidikan Indonesia and a lecturer at the Department of Curriculum Development, Faculty of Education, Universitas Pendidikan Indonesia. He has authored numerous books, including Instructional Technology, Quantitative Research Methods, and Information and Communication Technology Education. His research interests include bio-communication science, educational technology, information systems, and information communication. He can be contacted at email: deni_darmawan@upi.edu.

Suhendra is a professor in Productive Mathematics Learning and a lecturer at the Mathematics Education Study Program, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia. He earned his master's degree in Mathematics Education from the University of Houston, Texas, USA, and his doctoral degree in Mathematics Education from Curtin University, Perth, Australia. He can be contacted at email: suhendra@upi.edu.

Deni Kurniawan is a lecturer in Curriculum Development at the Faculty of Education, Universitas Pendidikan Indonesia. He completed both his master's and doctoral degrees in Curriculum Development. One of the courses he teaches is Secondary Education Curriculum. He can be contacted at email: deniku@upi.edu.