AN-NAHDLAH EDU: Journal of Elementary Education
VOLUME 1 | NUMBER 1, June 2026 | **Page: 11-20**
Web: http://jurnal.unusumbar.ac.id:8090/ojs/index.php/An-NahdlahEdu

A Diagnostic Analysis of Fourth-Grade Elementary Students' Numeracy Difficulties in Solving Contextual Problems: A Rasch Model Approach and Implications for Differentiated Instruction

Sonia Yulia Friska[1], Ferry Gustian Pratama[2]
[1] Universitas Islam Yasni Bungo, [2]SDN 104/VIII Sumber Sari
Email: soniayuliafriska@gmail.com, ferrypratama58@gmail.com

Article Info

Article History
Received:
2026-05-28
Revised:
2026-06-28
Accepted:
2026-06-30

Keywords :
contextual numeracy, numeracy difficulties, contextual problem solving, Rasch Model, differentiated instruction, elementary education

Abstract

Contextual numeracy is a key competency in Indonesia's *Merdeka Curriculum*, yet many elementary students continue to demonstrate low achievement. This study aimed to identify fourth-grade students' contextual numeracy difficulties using the Rasch Model and formulate implications for differentiated instruction. A quantitative descriptive design with a diagnostic assessment approach was employed involving 156 fourth-grade students from three public elementary schools in Bogor Regency selected through cluster random sampling. Data were collected using a 30-item contextual numeracy diagnostic test with satisfactory psychometric properties (Cronbach's α = 0.86, Person Reliability = 0.79, Item Reliability = 0.94). Rasch analysis generated Wright Maps, item difficulty estimates, and profiles of students' cognitive errors. The results showed that only **31.4%** of students achieved the expected numeracy proficiency. The greatest difficulties occurred in mathematical modeling of real-world situations (logit = 1.87) and multimodal data interpretation (logit = 1.52). Think-aloud protocols identified three major cognitive barriers: misconceptions in contextual numerical operations, difficulties translating verbal information into mathematical representations, and excessive cognitive load. These findings informed the development of differentiated instructional strategies based on content, process, and product differentiation. The study demonstrates that integrating Rasch-based diagnostic assessment with differentiated instruction provides a practical, data-driven framework for improving contextual numeracy learning in elementary schools.

INTRODUCTION

Numeracy is a fundamental competency that extends beyond computational skills to encompass the ability to reason, model, and interpret quantitative information in diverse real-world contexts. It serves as a critical foundation for developing learners who can think critically, solve problems effectively, and make evidence-based decisions.

Within Indonesia's *Merdeka Curriculum*, numeracy is positioned as an integral component of the *Pancasila Student Profile*, particularly in fostering critical and creative reasoning. Consequently, the development of numeracy skills during elementary education is essential for preparing students to meet both academic demands and everyday life challenges (Kemendikbudristek, 2024).

Despite its recognized importance, national evidence indicates that elementary students' numeracy proficiency remains below expected standards. Results of the 2023 *Minimum Competency Assessment* (*Asesmen Kompetensi Minimum* [AKM]) revealed that only 42.7% of fourth-grade students achieved the expected level of numeracy proficiency, with the weakest performance observed in contextual problem solving and data interpretation (Kemendikbudristek, 2024). These findings suggest that many students continue to experience difficulties in connecting mathematical concepts with authentic situations, limiting the achievement of higher-order thinking skills emphasized in the *Merdeka Curriculum*.

These national findings are consistent with a preliminary study conducted by the researchers between August and October 2023 in three public elementary schools in Bogor Regency. Diagnostic assessments administered to 120 fourth-grade students showed that only 34.2% met the minimum competency standard. Further analysis indicated that the most frequent errors occurred in modeling buying-and-selling situations using fractions and decimals (68.3%), interpreting bar graphs and frequency tables (61.7%), and applying proportional reasoning in measurement contexts (57.5%). These results indicate that students' numeracy difficulties extend beyond deficiencies in computational skills and instead reflect challenges in mathematical reasoning and the application of mathematical concepts to real-life situations.

Previous studies have identified several factors contributing to students' limited contextual numeracy skills. Siregar and Wijaya (2021) argued that misconceptions about numerical operations and weak conceptual understanding are among the primary causes of poor numeracy performance. However, their study relied on Classical Test Theory (CTT), making the measurement results highly dependent on sample characteristics. Similarly, Hidayati et al. (2022) developed a diagnostic assessment integrated with remedial instruction but did not objectively estimate item difficulty using contemporary psychometric approaches.

In recent years, the Rasch Model has gained increasing recognition as a more objective measurement framework than classical score-based analyses. Bond and Fox (2015), together with Linacre (2020), demonstrated that the Rasch Model independently calibrates item difficulty and learner ability, producing invariant and more robust measurement estimates. This approach enables educators to obtain more accurate diagnostic information regarding students' learning characteristics. Although Retnawati et al. (2023) successfully applied the Rasch Model in science literacy assessment, its application to contextual numeracy assessment in elementary education remains relatively limited.

Beyond assessment, recent educational research has emphasized the importance of differentiated instruction as an effective strategy for addressing learners' diverse needs. Carol Ann Tomlinson (2017) proposed that differentiation should be implemented by adapting instructional content, learning processes, and learning products according to students' readiness, interests, and learning profiles. This perspective is supported by Subarawati et al. (2024), who reported that differentiated instruction significantly enhances students' engagement in mathematics learning. Nevertheless, few studies have translated diagnostic evidence generated through Item Response Theory (IRT)—particularly Rasch analysis—into practical and operational differentiation strategies that can be readily implemented in elementary classrooms.

This study addresses these gaps by offering three principal contributions. First, it employs the Rasch Model to objectively diagnose students' contextual numeracy abilities and establish an invariant hierarchy of item difficulty, thereby overcoming the limitations of classical score-based analysis. Second, it integrates quantitative psychometric evidence with think-aloud protocols to uncover the cognitive mechanisms underlying students' errors in solving contextual numeracy tasks. Third, it develops an instructional framework that directly translates Wright Map findings into differentiated strategies involving content, process, and product differentiation, providing elementary school teachers with an evidence-based guide for instructional decision-making.

Accordingly, this study aims to examine students' contextual numeracy proficiency and the hierarchy of item difficulty using Rasch Model analysis based on logit difficulty estimates, infit and outfit mean-square statistics (MNSQ), and person-item reliability indices. It further seeks to identify patterns of cognitive errors and difficulties in translating contextual information into mathematical representations through think-aloud protocols and analyses of students' written work. Finally, the study formulates practical implications for differentiated instruction by transforming diagnostic findings into an ability-clustered instructional framework designed to support more effective and inclusive numeracy learning in elementary education.

RESEARCH METHOD

This study employed a descriptive quantitative approach with a diagnostic design enriched by an exploratory qualitative component. This approach was selected to provide an objective assessment of students' contextual numeracy abilities while exploring the cognitive processes underlying their responses to mathematical problem-solving tasks. The study population comprised all fourth-grade students from public elementary schools in Cibinong District, Bogor Regency, totaling 840 students. The sample was selected using cluster random sampling from three schools—SDN Cibinong 01, SDN Cibinong 03, and SDN Cibinong 07—resulting in 156 participants (52 students from each school). This sample size satisfied the minimum requirement for Rasch Model analysis, which recommends at least 120 respondents to obtain stable item parameter estimates (Linacre, 2020). All participants took part in the study after obtaining parental informed consent and official approval from the participating schools, while ensuring the confidentiality of students' identities.

The primary instrument was a contextual numeracy diagnostic test consisting of 30 items, including multiple-choice questions with reasoning and short constructed-response items. The test was developed based on the contextual numeracy indicators proposed by the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek, 2022), covering the abilities to model real-life situations, perform numerical operations in authentic contexts, interpret data presented in graphs and tables, apply proportional reasoning, and conduct estimation and measurement. Content validity was evaluated by five experts, including two mathematics education lecturers, two certified elementary school teachers, and one educational psychologist, yielding Content Validity Ratio (CVR) values ranging from 0.82 to 0.94. A pilot study involving 60 students outside the research sample produced a Cronbach's alpha reliability coefficient of 0.86, with item discrimination indices ranging from 0.31 to 0.78, indicating that all items were suitable for use following minor revisions.

Data collection was conducted in three stages. The first stage involved administering the diagnostic test in a face-to-face setting over a 90-minute session using structured answer sheets that emphasized students' reasoning processes. The second stage consisted of think-aloud protocols and semi-structured interviews with 15 students selected through stratified sampling, representing high-, medium-, and low-ability groups (five students from each group). During this phase, participants were asked to verbalize their thinking while solving five representative numeracy problems. All sessions were audio-recorded and transcribed for subsequent analysis. The third stage involved collecting students' written work and conducting classroom observations to triangulate identified error patterns and instructional practices.

Quantitative data were analyzed using Winsteps version 5.2.3 based on the Rasch Model to estimate student ability (person measures) and item difficulty (item measures) on the logit scale. The analysis included item fit evaluation using Infit/Outfit Mean Square (MNSQ) criteria of 0.7–1.3, ZSTD values within ±2.0, person and item reliability indices, unidimensionality testing through Principal Component Analysis of Residuals (PCAR), and the construction of a Wright Map to visualize the relationship between student ability and item difficulty. Qualitative data derived from think-aloud transcripts, interviews, and student work were analyzed thematically using the Cognitive Process Dimensions framework proposed by Anderson and Krathwohl (2001) together with Newman's Error Taxonomy (1977) to identify the cognitive mechanisms underlying students' problem-solving errors.

In the final stage, quantitative and qualitative findings were integrated to produce a comprehensive diagnostic profile of students' contextual numeracy abilities. Student ability clusters identified through the Rasch analysis were combined with insights from the think-aloud protocols to generate a more accurate representation of learning difficulties. These diagnostic findings were subsequently translated into differentiated instructional strategies based on Tomlinson's (2017) framework, encompassing differentiation of content, process, and product according to students' learning profiles. Construct validity was supported through fit statistics, PCAR results, and person reliability values exceeding 0.70, while the credibility of the qualitative findings was strengthened through source triangulation and an audit trail. All research procedures adhered to established ethical standards, including informed consent, participant confidentiality, the right to withdraw from the study at any stage, and secure data management throughout the research process.

RESULTS AND DISCUSSION

Mapping Student Ability and Item Difficulty Using the Rasch Model

The Rasch analysis of 156 respondents and 30 test items produced a comprehensive map of student ability and item difficulty on the logit scale. Model fit statistics demonstrated satisfactory unidimensionality, with 41.2% of the variance explained by the principal component (>40%). The instrument also showed good psychometric quality, with person reliability = 0.79 and item reliability = 0.94, indicating stable measurement and a strong ability to discriminate among different levels of student ability.

Table 1. Distribution of Student Ability Clusters (Person Measures)

Ability Cluster	Logit Range	Number of Students	Percentage	Diagnostic Characteristics
High	> 0.85	38	24.4%	Able to model contextual situations and perform multi-step reasoning
Moderate	–0.45 to 0.85	69	44.2%	Competent in basic operations but struggled with complex contextual problems
Low	< –0.45	49	31.4%	Demonstrated conceptual misconceptions and difficulties in representational translation

The findings indicate that only 24.4% of students belonged to the high-ability cluster, whereas 31.4% required intensive diagnostic intervention. The Wright Map (Figure 1) revealed a substantial mismatch between student ability distribution and item difficulty. Approximately 63% of the test items were located above the average student ability level (–0.12 logits), confirming that the contextual numeracy assessment posed considerable cognitive challenges for fourth-grade students.

Table 2. Hierarchy of the Five Most Difficult Contextual Numeracy Items

Item	Contextual Indicator	Logit Difficulty	Infit MNSQ	Outfit MNSQ	Dominant Error Pattern
Q18	Multi-step discount and tax modeling	1.87	1.12	1.28	Incorrect operation sequence; ignored contextual relationships
Q24	Interpretation of double bar graphs	1.52	0.94	1.05	Misreading axes and failure to compare data
Q09	Proportional reasoning using map scales	1.31	1.08	1.15	Inconsistent unit conversion
Q21	Estimating the volume of irregular containers	1.19	0.89	0.92	Applying standard formulas without contextual adaptation
Q05	Translating word problems into fractional equations	0.98	1.21	1.33	Failure to identify variables and mathematical relationships

Items Q18 and Q24 exhibited the highest difficulty levels while remaining within the acceptable fit range (Infit/Outfit = 0.7–1.3), indicating that they effectively measured the intended construct despite their substantial cognitive demands. The high logit value for multi-step mathematical modeling supports the findings of Siregar and Wijaya (2021) that elementary students often lack sufficient schemas to solve multi-context problems. However, the present study extends this conclusion by demonstrating that students' difficulties arise not merely from procedural deficiencies but also from failures in representational translation, specifically the transformation of textual information into symbolic and visual mathematical representations.

Cognitive Processes and Representational Barriers

Qualitative findings derived from the think-aloud protocols complemented the Rasch analysis by revealing the cognitive mechanisms underlying students' error patterns. Three major cognitive barriers were identified.

1. Misconceptions of Mathematical Operations in Context (38% of responses).
   Students frequently applied algorithmic procedures mechanically without considering contextual relationships. One participant stated:
   "If there is a 20% discount and then a tax is added, I simply subtract 20% and then add 10% of the original price. That's easy." (Student R-04)
   This response illustrates a misunderstanding that tax should be calculated after applying the discount rather than from the original price. This finding is consistent with Cognitive Load Theory (Sweller, 2016), which suggests that high intrinsic cognitive load in multi-step problems encourages students to rely on inappropriate heuristics.

2. Failure in Mathematical Representation Translation (34% of responses).
   Many students experienced difficulty converting verbal narratives into mathematical or visual representations. As one student explained:
   "The problem is too long. I don't know whether I should multiply or divide. The picture is also too small." (Student S-11)
   This finding supports Duval's (2006) theory that representational fluency is a prerequisite for mathematical reasoning and suggests that explicit instruction in representational transformation remains insufficient in elementary classrooms.
3. Cognitive Load and Shortcut Response Strategies (28% of responses).
   For items containing numerically similar distractors, students tended to select the first familiar-looking option without verification. One participant remarked:
   "The number 48 appears twice, so it must be the correct answer." (Student T-02)

This pattern reflects students' reliance on recognition heuristics rather than analytical reasoning, a tendency exacerbated by time pressure and the absence of metacognitive scaffolding.

Triangulation through students' written work and classroom observations further revealed that previous mathematics instruction emphasized procedural computation while providing limited opportunities for contextual exploration or error reflection. Consequently, students were able to perform routine calculations but struggled when confronted with unfamiliar contexts or multimodal data presentations.

Theoretical Discussion and Synthesis of Findings

The integration of Rasch analysis and qualitative evidence indicates that contextual numeracy difficulties among fourth-grade students are both hierarchical and multidimensional. The greatest challenges were associated with tasks requiring representational flexibility and sequential reasoning rather than basic arithmetic computation.

These findings are consistent with Hidayati et al. (2022), who emphasized the importance of transitioning from computational fluency to conceptual modeling. However, the present study advances previous research by providing a more precise diagnostic framework through logit calibration. Unlike Classical Test Theory, the Rasch Model revealed that some items appearing easy according to percentage-correct statistics exhibited measurement misfit because they were answered correctly through guessing, whereas several difficult items proved to be the most effective in discriminating among different ability levels.

From the perspective of social constructivism (Vygotsky, 1978; Tomlinson, 2017), these findings reflect a gap between students' actual developmental level and their potential developmental level, highlighting the need for differentiated scaffolding. The Rasch Model functions as a diagnostic compass, identifying each student's position within the Zone of Proximal Development (ZPD). Students in the low-ability cluster require concrete scaffolds and segmented representations; those in the moderate cluster benefit from representational translation activities and structured error reflection; whereas high-performing students require authentic, open-ended contextual problems and opportunities for peer mentoring. These findings reinforce the argument of Subarawati et al. (2024) that differentiated instruction should represent a systematic response to students' cognitive profiles rather than merely a modification of classroom tasks.

Practical Implications: A Diagnostic-Based Differentiated Learning Framework

Based on the diagnostic findings, this study proposes an operational framework for differentiated instruction that can be readily implemented in elementary mathematics classrooms.

Table 3. Diagnostic-Based Content–Process–Product Differentiation Matrix

Ability Cluster	Content Differentiation	Process Differentiation	Product Differentiation
Low	Concrete visual modules, segmented learning steps, familiar real-life contexts (e.g., school marketplace)	Scaffolded worksheets, think-pair-share activities, concrete manipulatives (paper folding, blocks)	Step-by-step solution posters and short audio explanations
Moderate	Semi-abstract contexts with multiple representations (graphs, tables, narratives)	Structured problem-based learning, reflective error journals, peer feedback	Comparative infographics and three-minute presentations
High	Authentic complex contexts (family budgeting, environmental data) with multiple solution pathways	Open-ended inquiry, mathematical debates, peer mentoring	Mini numeracy projects and creative instructional videos

This framework emphasizes that differentiated instruction should be data-driven rather than intuition-based. Teachers may use the proposed Rasch-based diagnostic sheet to dynamically group students for each instructional unit. Preliminary validation through a focus group discussion involving eight elementary school teachers demonstrated an 87.5% readiness for adoption, although participants highlighted the need for brief professional development on interpreting logit measures and designing appropriate instructional scaffolding.

CONCLUSION

Based on the diagnostic analysis and systematic discussion, this study achieved its research objectives through three major findings. First, the Rasch Model analysis revealed that only 24.4% of fourth-grade students demonstrated adequate contextual numeracy proficiency, with the greatest difficulties observed in multi-step contextual modeling (logit = 1.87) and multimodal data interpretation (logit = 1.52). The mismatch between student ability and item difficulty distributions highlights the need for calibrated assessments and targeted instructional interventions. Second, the think-aloud protocols identified three dominant cognitive barriers: misconceptions of mathematical operations in context (38%), failure to translate between mathematical representations (34%), and excessive cognitive load leading to guessing strategies (28%). These findings indicate that students' difficulties stem primarily from cognitive processing rather than deficiencies in procedural knowledge alone. Third, the proposed diagnostic-based differentiated learning framework successfully links Rasch diagnostic profiles with specific strategies for differentiating content, process, and product, enabling teachers to respond systematically to diverse student learning needs using empirical evidence.

Overall, integrating the Rasch Model, cognitive diagnostics, and differentiated instruction provides a more precise and responsive approach to addressing contextual numeracy difficulties in elementary education. This study bridges the methodological gap between modern psychometric assessment and classroom practice while offering practical diagnostic tools that can support teacher professional development and school-level curriculum implementation. Nevertheless, the study was limited to a single geographic region and did not examine the long-term effects of diagnostic-based differentiated instruction. Future research should expand the sample across multiple provinces, incorporate digital learning analytics, and evaluate the effectiveness of the proposed framework through quasi-experimental designs using Minimum Competency Assessment (AKM) indicators. Contextual numeracy should be viewed not as an innate ability but as a competence that can be systematically developed through accurate diagnosis and appropriately differentiated instruction.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd ed.). Routledge.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, 61(1), 103–131.

Friska, Sonia Y., et al.(2022)."Pengembangan E-LKPD dengan 3D Pageflip Professional Berbasis Problem Solving pada Tema Lingkungan Sahabat Kita di Sekolah Dasar." Jurnal Basicedu, vol. 6, no. 2, 2022, pp. 3200-3206, doi:10.31004/basicedu.v6i2.1685.

Hidayati, N., Sulistiyo, U., & Fitriana, L. (2022). Diagnostic assessment and remedial teaching in elementary mathematics: A classroom-based study. *Journal of Elementary Education*, 15(2), 112–125.

Kemendikbudristek. (2022). *Kerangka Asesmen Numerasi Jenjang SD*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Kemendikbudristek. (2024). *Laporan Nasional Asesmen Kompetensi Minimum 2023*. Pusat Asesmen dan Pembelajaran.

Linacre, J. M. (2020). *Winsteps Rasch Measurement Computer Program* (Version 5.2.3). Winsteps.com.

Newman, M. A. (1977). An analysis of sixth-grade pupils' errors on written mathematical tasks. *Victorian Institute for Educational Research Bulletin*, 39, 31–43.

Retnawati, H., Apino, E., & Djidu, H. (2023). Applying Rasch model to analyze science literacy items in national assessment. *International Journal of Educational Research Open*, 4, 100215.

Siregar, N. C., & Wijaya, A. (2021). Analisis miskonsepsi dan kesulitan belajar numerasi siswa sekolah dasar. *Jurnal Pendidikan Matematika dan IPA*, 12(1), 45–58.

Subarawati, N. K. A., Suardana, I. W., & Suastika, K. G. (2024). Implementasi pembelajaran berdiferensiasi dalam kurikulum merdeka: Tantangan dan strategi guru SD. *Jurnal Ilmiah Pendidikan Dasar*, 11(2), 189–204.

Sweller, J. (2016). Cognitive load theory and educational technology. In *Educational Technology Research and Development* (pp. 1–18). Springer.

Tomlinson, C. A. (2017). *How to Differentiate Instruction in Academically Diverse Classrooms* (3rd ed.). ASCD.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.