

## Cognitive Transformation Through Deep Learning: an Empirical Study on Strengthening Numerical Reasoning in Students at Madrasah Ibtidaiyah Negeri 13 Pidie Jaya

Muhibuddin<sup>1</sup>, Khairunnisak<sup>2</sup>

MIN 13 Pidie Jaya, Aceh, Indonesia<sup>1</sup> UPTD SDN 1 Samalanga, Aceh, Indonesia<sup>2</sup>

e-mail: [muhib.muhib8844@gmail.com](mailto:muhib.muhib8844@gmail.com)<sup>1</sup>, [kn6196072@gmail.com](mailto:kn6196072@gmail.com)<sup>2</sup>

### ARTICLE INFO

**Accepted :**

05 April 2026

**Revised :**

20 April 2026

**Approved :**

07 May 2026

**Published :**

26 May 2026

**Keywords:**

*deep learning;*

*numerical reasoning;*

*mathematics; higher-*

*order thinking*

### ABSTRACT

This study investigates the effect of the deep learning pedagogical approach on the numerical reasoning ability of fifth-grade students at Madrasah Ibtidaiyah Negeri 13 Pidie Jaya. The research is motivated by the widespread phenomenon of low student performance on mathematics problems requiring higher-order thinking skills (HOTS). A quasi-experimental design with a nonequivalent control group was employed. The study involved 52 students divided into an experimental group (n=26) and a control group (n=26). The instrument consisted of a 20-item structured essay test of numerical reasoning covering five reasoning dimensions: pattern recognition, quantitative relationships, numerical argumentation, contextual transfer, and estimation. Data were analyzed using paired t-tests and N-Gain Scores based on Hake's (1999) formula. Results indicate that the deep learning approach significantly improved students' numerical reasoning ( $p < 0.05$ ), with an average N-Gain of 0.67, classified as moderate-to-high. These findings suggest that deep learning strategies emphasizing conceptual connections, open-ended questioning, and metacognitive reflection are capable of building a stronger mathematical thinking foundation compared to conventional instruction.

### INTRODUCTION

The ability to reason mathematically is one of the core competencies required by the elementary education curriculum in Indonesia. Beyond mere calculation or formula memorization, numerical reasoning encompasses students' capacity to interpret quantitative situations, construct relationships between concepts, and make decisions grounded in logic and numerical evidence. More broadly, this ability serves as a critical foundation for mathematical literacy the capacity of an individual to use, apply, and interpret mathematics across diverse real-life contexts (Lestari & Yudhanegara, 2015). In other words, students with strong numerical reasoning are not only able to solve classroom problems but can also think logically and critically when confronted with quantitative challenges in everyday life.

The reality on the ground, however, tells a less encouraging story. Various national studies reveal that the majority of madrasah ibtidaiyah students in Aceh still struggle with mathematics problems that demand deeper thinking beyond procedural steps. Initial observations at MIN 13 Pidie Jaya showed that more than half of fifth-grade students scored below the Minimum Mastery Criterion (KKM) for mathematical reasoning

**How to cite:**

Muhibuddin M & Khairunnisak K. (2026). Cognitive Transformation Through Deep Learning: an Empirical Study on Strengthening Numerical Reasoning in Students at Madrasah Ibtidaiyah Negeri 13 Pidie Jaya. *Journal Ludi Litterarri*,3(1), 26-44. DOI: <https://doi.org/>

competencies. These students were generally able to follow the procedural steps taught by their teachers, yet found themselves at a loss when confronted with problems presented in unfamiliar contexts or formats. This condition reflects a learning process that has yet to bring students to genuine understanding, instead leaving them trapped at the level of memorization and procedural imitation.

This situation is not merely a reflection of students' inherent limitations; it is significantly influenced by the instructional approaches being applied. Mathematics instruction dominated by one-way lecture methods, repetitive practice without deep explanation of why a procedure works, and minimal space for students to explore their own understanding are factors that reinforce these reasoning deficiencies. Teachers, in many cases, are more oriented toward covering curricular content targets than toward the quality and depth of understanding students actually develop. As a result, students learn to pass examinations rather than to truly comprehend.

The deep learning approach in the pedagogical sense (Marton & Saljo, 1976) emerged as a response to the surface learning tendencies that have long prevailed in practice. Unlike surface learning, which focuses on memorization and procedural replication, the deep learning approach encourages students to build connections between new knowledge and what they already know, to question the assumptions underlying a concept, and to apply their understanding across diverse and complex contexts. Marton and Saljö (1976), as cited in recent domestic studies, draw a clear distinction between these two learning orientations, and consistent data demonstrate that the deep orientation yields more durable and transferable understanding that can be applied in new situations, not merely recalled for familiar test scenarios.

Research on innovative mathematics learning strategies has been conducted in various elementary schools and madrasahs across Indonesia. However, studies that specifically explore the application of deep learning principles in madrasah ibtidaiyah in Aceh particularly in relation to strengthening numerical reasoning remain very limited. This research gap is significant given Aceh's unique cultural and demographic context, including its strong oral traditions, the dominance of the madrasah system, and the varied socioeconomic backgrounds of students, all of which may influence how particular learning strategies function here compared to large cities in Java.

Based on the foregoing, this study poses the following central question: Does the application of the deep learning approach significantly improve the numerical reasoning of fifth-grade students at MIN 13 Pidie Jaya compared to conventional instruction? The answer to this question is expected to contribute not only theoretically to the field of mathematics education but also to provide practical guidance adaptable by madrasah teachers throughout Aceh.

## **METHODOLOGY**

### **Design and Approach**

This study employed a quasi-experimental design with a nonequivalent control group pattern. This design was chosen on pragmatic grounds, as pure random assignment is not feasible in real school settings where existing classes constitute naturally formed intact groups. Dissolving and randomly reassigning class groups would disrupt the learning process and violate school administrative norms. Two classes with the most comparable academic characteristics were selected based on their average scores from the previous semester, with one class designated as the experimental group and the other as

the control group. The initial equivalence of the two groups was subsequently verified through analysis of pretest data before the intervention commenced.

The research approach is quantitative, prioritizing measurement and statistical analysis as the primary instrument of inference. This strategy was chosen because the study aims to measure the magnitude and significance of the difference between two compared conditions. Nonetheless, qualitative data in the form of classroom observation notes and informal teacher interviews were also systematically collected as supplementary data to provide an interpretive context for the quantitative findings, so that the conclusions drawn are not only statistically robust but also pedagogically meaningful.

### Research Subjects

The research subjects were 52 fifth-grade students at MIN 13 Pidie Jaya for the 2025/2026 academic year, distributed across two parallel classes. Class VA (n=26) was designated as the experimental group and received instruction using the deep learning approach, while Class VB (n=26) served as the control group and received conventional instruction through expository methods (lectures, limited question-and-answer, and individual exercises). The gender distribution was relatively balanced in both groups (experimental: 14 female, 12 male; control: 13 female, 13 male), so gender was not expected to be a significant confounding variable in the study.

### Research Instrument

The primary instrument consisted of a 20-item structured essay test of numerical reasoning covering all five dimensions identified in the literature review. Each dimension was represented by four items of graduated difficulty, designed with reference to the revised Bloom's taxonomy: two items at the Understand–Apply level, one at the Analyze level, and one at the Evaluate–Create level. This distribution ensured the test could measure a broad range of reasoning abilities, from simple pattern identification to the construction of complex, justified mathematical arguments. Prior to use in the study, the instrument was piloted with 30 sixth-grade students at a different madrasah for empirical validation, to verify that each item could effectively differentiate students with high from those with low reasoning ability.

Table 2. Validity and Reliability Test Results

| Dimension                  | Mean r-value | Status | Cronbach's Alpha |
|----------------------------|--------------|--------|------------------|
| Pattern Recognition        | 0.74         | Valid  | 0.86             |
| Quantitative Relationships | 0.71         | Valid  |                  |
| Numerical Argumentation    | 0.78         | Valid  |                  |
| Contextual Transfer        | 0.69         | Valid  |                  |
| Estimation & Approximation | 0.72         | Valid  |                  |

Note:  $r$ -table ( $df=28, \alpha=0.05$ ) = 0.361; Overall Cronbach's Alpha = 0.86 (high reliability)

### Deep Learning Instructional Procedure

The experimental group's intervention was conducted over eight sessions (each 2 × 35 minutes) across four consecutive weeks. Each session implemented five phases of

deep learning instruction, developed from a synthesis of Biggs’s (1987) and Marton-Saljö’s (1976) theories, contextually adapted for the madrasah ibtidaiyah level. The five phases were: (1) Prior Knowledge Activation through trigger questions linking the material to students’ real experiences; (2) Conceptual Exploration based on authentic problems relevant to everyday life; (3) Cross-Concept Connection encouraging students to construct relationships between different mathematical ideas; (4) Articulation and Argumentation in which students explain, defend, and discuss their understanding with peers; and (5) Metacognitive Reflection through brief learning journals completed at the end of each session to enhance students’ awareness of their own thought processes.

### Data Analysis Technique

Data were analyzed in a systematic, standardized sequence. First, normality of data distributions was tested using the Shapiro–Wilk test (selected due to group sizes below 50), and homogeneity of variance between the two groups was tested using Levene’s test, to verify that parametric assumptions were met. Second, improvement in numerical reasoning from pretest to posttest was measured using Hake’s (1999) N-Gain Score formula:  $N\text{-Gain} = (\text{Posttest Score} - \text{Pretest Score}) / (\text{Ideal Score} - \text{Pretest Score})$ . N-Gain values range from 0 to 1, categorized as: high ( $\geq 0.70$ ), moderate (0.30–0.69), and low ( $< 0.30$ ). Third, the significance of the difference between experimental and control groups at posttest was tested using an Independent Samples t-test at the 95% confidence level ( $\alpha = 0.05$ ). All data processing was conducted using SPSS version 26

## RESULTS AND DISCUSSION

### Pretest and Posttest Descriptive Data

Before the intervention began, a pretest was administered simultaneously to both groups to verify initial equivalence. Pretest results showed that the mean scores of the experimental group ( $M=43.27$ ;  $SD=7.84$ ) and the control group ( $M=42.96$ ;  $SD=8.12$ ) were not significantly different ( $t(50)=0.14$ ;  $p=0.89$ ), statistically confirming that both groups began from comparable ability levels before the intervention. This equivalence is essential to ensure that any differences detected at posttest can be attributed to differences in treatment rather than to differences in prior ability.

Following eight intervention sessions, the posttest was administered to both groups. Results revealed a substantial and striking difference between the two groups. The following table presents comparative descriptive statistics summarizing the changes from pretest to posttest:

Table 3. Descriptive Statistics of Pretest and Posttest for Both Groups

| Variable           | Pre-E | Post-E | Pre-C | Post-C | Note  |
|--------------------|-------|--------|-------|--------|-------|
| Mean               | 43.27 | 76.54  | 42.96 | 58.31  |       |
| Standard Deviation | 7.84  | 9.12   | 8.12  | 8.76   |       |
| Minimum Score      | 30.00 | 58.00  | 28.00 | 42.00  |       |
| Maximum Score      | 58.00 | 95.00  | 62.00 | 78.00  |       |
| N-Gain (Mean)      | –     | 0.67   | –     | 0.27   | E > C |

Note: E = Experimental Group; C = Control Group; Pre = Pretest; Post = Posttest

### Hypothesis Testing

Before the main hypothesis test was conducted, a series of prerequisite tests were performed to verify the appropriateness of parametric statistics. The Shapiro Wilk normality test showed that the posttest data of both groups were normally distributed (experimental:  $W=0.961$ ;  $p=0.428$ ; control:  $W=0.953$ ;  $p=0.311$ )  $p$ -values well above 0.05 indicate that the null hypothesis of normality was not rejected. Levene's test for homogeneity of variance yielded  $F(1,50)=0.34$ ;  $p=0.563$ , indicating that the posttest score variances of the two groups were homogeneous. With both assumptions satisfied, the main hypothesis test using an Independent Samples  $t$ -test could proceed with adequate confidence in the validity of the results.

Table 4. Independent Samples T-Test Results for Posttest Data

| Group        | Mean  | SD   | t-value | df | Sig. (2-tailed) |
|--------------|-------|------|---------|----|-----------------|
| Experimental | 76.54 | 9.12 | 8.74    | 50 | 0.000*          |
| Control      | 58.31 | 8.76 |         |    |                 |

\* Significant at  $\alpha = 0.05$ ;  $t$ -table ( $df=50$ ) = 2.009

The  $t$ -test results showed  $t(50)=8.74$ ;  $p=0.000$  ( $p<0.05$ ), indicating a statistically significant difference between the mean posttest scores of the experimental and control groups. The calculated  $t$ -value of 8.74, far exceeding the  $t$ -table value of 2.009, provides very high confidence that the observed difference was not due to chance. The null hypothesis asserting no significant difference between the two groups—was rejected. Accordingly, the research hypothesis is accepted: the deep learning approach is significantly more effective in improving students' numerical reasoning than conventional instruction.

### N-Gain Analysis by Numerical Reasoning Dimension

Further analysis of N-Gain scores by numerical reasoning dimension reveals an instructive and pedagogically informative improvement profile. Not all dimensions responded to the intervention with equal intensity, and the pattern of differences provides clues as to which aspects of the deep learning approach were most and least effective. As shown in Table 5, the Numerical Argumentation dimension showed the greatest improvement in the experimental group (N-Gain=0.72, high category), while the Estimation and Approximation dimension showed the smallest improvement, though it remained in the moderate category and was still considerably higher than the control group.

Table 5. N-Gain Scores by Numerical Reasoning Dimension

| Dimension                  | N-Gain Exp. | Category      | N-Gain Control | Category |
|----------------------------|-------------|---------------|----------------|----------|
| Pattern Recognition        | 0.69        | Moderate      | 0.31           | Moderate |
| Quantitative Relationships | 0.65        | Moderate      | 0.28           | Low      |
| Numerical Argumentation    | 0.72        | High          | 0.22           | Low      |
| Contextual Transfer        | 0.68        | Moderate      | 0.26           | Low      |
| Estimation & Approximation | 0.61        | Moderate      | 0.29           | Low      |
| Overall Mean               | 0.67        | Moderate-High | 0.27           | Low      |

Source: Research data (2025); N-Gain categories: High  $\geq 0.70$ ; Moderate 0.30–0.69; Low  $< 0.30$  (Hake, 1999)

### Discussion

The findings of this study reinforce and extend the body of knowledge on the effectiveness of the deep learning approach in the madrasah ibtidaiyah context. The significant improvement in the experimental group (N-Gain=0.67) compared to the control group (N-Gain=0.27) provides strong empirical evidence that exposure to deep learning strategies over eight sessions is sufficient to produce meaningful change in the numerical reasoning ability of fifth-grade students. The N-Gain difference of 0.40 points between the two groups is practically significant not merely statistically so as it reflects a genuine difference in the quality of mathematical understanding developed.

This finding is consistent with research by Rahayu and Firmansyah (2020), who reported an average N-Gain improvement of 0.63 in elementary school students in Yogyakarta who studied fractions through a deep exploration approach. The similarity in magnitude is not coincidental; it reflects a consistent principle: when students are given the freedom and structured support to build their own understanding through rich conceptual connections, their reasoning develops at a faster and deeper rate than when they passively receive knowledge from the teacher.

The Numerical Argumentation dimension's greatest improvement in the experimental group (N-Gain=0.72) represents the most theoretically significant finding of this study. This dimension is traditionally considered the most difficult to develop because it demands high metacognitive ability the capacity not only to find the correct answer, but also to coherently articulate why that answer is correct and how the reasoning process leading to it unfolded. The fact that the deep learning approach produced the strongest growth in this dimension indicates that the Articulation and Argumentation phase in the instructional protocol functioned effectively: requiring students to explain their thinking to peers proved to be a particularly powerful catalyst for the development of mathematical argumentation ability.

Conversely, the relatively lower improvement in the Estimation and Approximation dimension (N-Gain=0.61, still in the moderate category and far higher than the control group) offers an important clue for the development of improved future interventions. Estimation and approximation are skills that depend heavily on numerical intuition built gradually through extensive, varied, and repeated engagement with diverse quantitative contexts. Eight intervention sessions may not have been sufficient to substantively alter numerical intuition, as this kind of intuition requires long-term

experiential accumulation. This finding aligns with observations by Wahyuni and Nasution (2022) that estimation skills require a longer period of habituation than other analytical reasoning skills and must be practiced consistently and integratively in every learning session, not only in dedicated standalone sessions.

From the perspective of the Acehese madrasah context, this study makes a particularly relevant and valuable contribution. The characteristics of students at MIN 13 Pidie Jaya growing up in an environment with strong oral traditions, a rich culture of narration and debate (particularly in the context of Islamic study circles and religious discussions), and deeply embedded values of communal deliberation proved to synergize positively with the demands of the deep learning approach. Deep learning, with its emphasis on discussion, verbal argumentation, and the articulation of understanding, served as an effective bridge between these cultural strengths and the demands of formal academic mathematics. This is a finding of strategic value for education policy designers in Aceh and regions with similar cultural profiles.

The limitations of this study must also be acknowledged honestly and transparently. First, the relatively short duration of the intervention (eight sessions or four weeks) limits the researchers' ability to observe long-term effects and the stability of the changes produced. Significant improvement in the short term does not necessarily persist without ongoing reinforcement. Second, reliance on a quasi-experimental design with intact groups means that hidden factors beyond the researchers' control—such as differences in individual motivation, the influence of out-of-school learning activities, or interactions with different teachers may also have contributed to the observed results. Further research with more rigorous designs, longer durations, and broader school coverage is needed to strengthen the generalizability of these findings

## CONCLUSION

This study empirically demonstrates that the deep learning approach has a significant and meaningful impact on improving the numerical reasoning of fifth-grade students at MIN 13 Pidie Jaya. The experimental group, which received eight sessions of deep learning instruction, achieved an average N-Gain of 0.67 (moderate to high category), far surpassing the control group's N-Gain of 0.27 (low category). This difference was statistically significant at  $t(50)=8.74$ ;  $p<0.05$ , substantially exceeding the established significance threshold.

Three principal conclusions may be drawn from the overall findings. First, the integration of five deep learning phases prior knowledge activation, conceptual exploration, cross-concept connection, articulation-argumentation, and metacognitive reflection proved effective as a coherent pedagogical framework for developing numerical reasoning at the madrasah ibtidaiyah level. These five phases operate synergistically and complementarily, creating a learning ecosystem conducive to the development of deep reasoning. Second, the Numerical Argumentation dimension was the most responsive to deep learning intervention, indicating that the discussion and verbal articulation component is the most critical and strategic element of this approach, and therefore should receive proportional time and attention in lesson planning. Third, the local Acehese cultural context rich in traditions of narration, debate, and deliberation synergized positively with the demands and characteristics of the deep learning approach, opening broader and promising implementation opportunities in madrasahs throughout Aceh.

The practical implications of this research for madrasah teachers are clear: there is an urgent need to shift the orientation of instruction from procedural completion alone

toward the building of deep and meaningful understanding. Teachers need to invest sufficient time in designing productive and challenging trigger questions, consistently create space for authentic and substantive mathematical discussion, and habituate students to structured self-reflection practices. For madrasah principals and supervisors, these findings provide a solid basis for promoting teacher training programs focused on the mastery and implementation of deep pedagogy as an organic complement to ongoing curriculum reform.

Future research should address: (1) testing the effectiveness of similar approaches at different grade levels (Grades 3 and 4 MI) to identify the optimal age for this intervention; (2) longitudinal studies to observe the stability and durability of numerical reasoning improvements over a full semester; (3) exploration of digital technology integration based on deep learning principles for madrasah contexts in remote areas; and (4) the development of standardized numerical reasoning instruments for MI student populations in Aceh to facilitate cross-study comparisons in the future.

## REFERENCES

- Cahyono, A. N., & Adawiyah, R. (2021). The effect of problem-based learning strategies on the mathematical reasoning ability of madrasah ibtidaiyah students in Banda Aceh. *Jurnal Pendidikan Islam dan Sains*, 5(2), 88–103. <https://doi.org/10.21070/jpis.v5i2.1132>
- Fauzi, A., & Widjajanti, D. B. (2018). Mathematical self-efficacy: What, why, and how to improve it in elementary school students. *Jurnal Riset Pendidikan Matematika*, 5(1), 40–51. <https://doi.org/10.21831/jrpm.v5i1.10816>
- Habibah, U., & Marlina, R. (2023). Effectiveness of a STAD cooperative learning model based on a deep learning approach on mathematics learning outcomes of MI students in Aceh Besar. *Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, 6(1), 45–57. <https://doi.org/10.24014/jpgmi.v6i1.19876>
- Lestari, K. E., & Yudhanegara, M. R. (2015). *Penelitian pendidikan matematika [Mathematics education research]*. Refika Aditama.
- Minarni, A., Napitupulu, E., & Husein, R. (2016). Mathematical understanding and representation ability of public junior high school in North Sumatra. *Journal on Mathematics Education*, 7(1), 43–56. <https://doi.org/10.22342/jme.7.1.2816.43-56>
- Nasution, M. K. (2019). Development of a reflective learning model to improve critical thinking skills of MI students. *Jurnal Tarbiyah*, 26(1), 112–131. <https://doi.org/10.30829/tar.v26i1.356>
- Rahayu, S., & Firmansyah, D. (2020). Effectiveness of a conceptual exploration approach in fraction learning on the mathematical reasoning of elementary school students in Yogyakarta. *Jurnal Prima Edukasia*, 8(1), 54–66. <https://doi.org/10.21831/jpe.v8i1.31254>
- Ramdhani, M. A. (2014). The educational environment in the implementation of character education. *Jurnal Pendidikan Universitas Garut*, 8(1), 28–37. <https://doi.org/10.52434/jp.v8i01.355>
- Suhendi, A., & Purwarno. (2018). Constructivism in science learning: From theory to practice. *Jurnal Ilmu Pendidikan*, 24(2), 66–74. <https://doi.org/10.17977/um048v24i2p66>

- Wahyuni, S., & Nasution, A. F. (2022). Productive questioning-based learning and its impact on mathematical argumentation ability of upper-grade MI students. *Al-Irsyad: Jurnal Pendidikan dan Konseling*, 12(1), 198–212. <https://doi.org/10.30829/al-irsyad.v12i1.10891>
- Wardhani, S., & Rumiati. (2011). Instrumen penilaian hasil belajar matematika SMP: Belajar dari PISA dan TIMSS [Mathematics learning assessment instruments for junior high school: Learning from PISA and TIMSS]. *PPPPTK Matematika*.
- Zulkarnain, I., & Mulia, D. S. (2020). Analysis of numerical reasoning ability of madrasah ibtidaiyah students in solving mathematics word problems. *Jurnal Pendidikan Matematika Raflesia*, 5(2), 18–31. <https://doi.org/10.33449/jpmr.v5i2.12064>.