

Mathematics Achievement under the Merdeka Curriculum: A Document Analysis of Grade VII Summative Assessment Results

Hizmi Wardani¹, Sahat Saragih², Mulyono³

¹Universitas Muslim Nusantara Al-Washliyah, Medan, Indonesia

^{2,3}Universitas Negeri Medan, Medan, Indonesia

Article Info

Article history:

Received 2026-09-16

Revised 2026-10-15

Accepted 2026-10-26

Keywords:

Document analysis

Geometry

Initial mathematical ability

Junior high school

Mathematics achievement

Merdeka curriculum

ABSTRACT

Mathematics achievement under the Merdeka Curriculum remains important, particularly in understanding how students' initial abilities relate to end-of-semester performance. Existing studies have examined mathematics achievement and Initial Mathematical Ability (IMA) separately, while evidence linking IMA to end-of-semester assessment performance at the junior secondary level remains limited. This study aimed to describe students' mathematics performance based on IMA and final semester assessment results and identify differences across mathematical domains. A document analysis design was employed using assessment records from 158 Grade VII students at a public junior high school in Medan. Data comprised an IMA test as the initial assessment and a final semester summative assessment as the post-assessment. Descriptive statistics and normalized gain analysis were used. The findings indicated moderate descriptive progress, with most students performing near or above the expected proficiency level. Geometry was the weakest assessed domain. These findings provide a descriptive profile of mathematics performance but do not establish the effectiveness of the Merdeka Curriculum, particularly because the two assessments may differ in content and measurement characteristics. Teachers can use domain-level assessment results to provide targeted support, especially in geometry, while future research should employ comparable longitudinal or quasi-experimental assessments.

This is an open-access article under the [CC BY-SA](#) license.



Corresponding Author:

Hizmi Wardani

FKIP Universitas Muslim Nusantara Al-Washliyah

Email: hizmiwardani@umnaw.ac.id

1. INTRODUCTION

Mathematics plays an important role in developing students' logical reasoning, critical thinking, numeracy, and problem-solving abilities, which are essential for addressing problems in academic and real-life contexts [1], [2]. These competencies are increasingly relevant to 21st-century mathematics education, which emphasizes problem solving,

creativity, computational thinking, collaboration, and the effective use of technology [3]–[5]. The development of such competencies is also closely related to critical thinking, creativity, communication, collaboration, and numeracy skills [4], [6]. At the same time, mathematical literacy enables students to interpret quantitative information and apply mathematical knowledge to real-world situations [7]. Consequently, examining students' mathematics achievement is important not only for describing learning outcomes but also for identifying areas that require further instructional attention.

At the international level, the mathematics performance of Indonesian students remains a significant concern. In PISA 2022, Indonesia obtained a mathematics score of 366, compared with the OECD average of 472, while only 18% of Indonesian students reached at least Level 2, compared with 69% across OECD countries [7]. The PISA findings therefore provide an important national context for examining mathematics achievement at the school level. Complementing this international evidence, a study involving Indonesian mathematics teachers identified several factors perceived to contribute to Indonesia's PISA 2022 mathematics performance, including the effects of the pandemic, curriculum-related issues, individual student factors, resource constraints, parental involvement, and the school environment [8]. Analysis of PISA 2022 data has also highlighted the role of cognitive activation in encouraging students to engage in deeper thinking, conceptual reasoning, and complex problem solving [9]. These findings indicate the importance of examining actual student performance in specific school contexts rather than relying solely on international or teacher-perception data.

At the junior high school level, difficulties in mathematics learning have been documented across several content domains. In algebra, seventh-grade students have experienced difficulties in understanding concepts, principles, and verbal representations [1]. Difficulties have also been reported in mathematical problem solving, including challenges in understanding concepts and developing appropriate solution strategies in Systems of Linear Equations in One Variable [10]. Similarly, students' errors in Lines and Angles have involved conceptual, procedural, and technical aspects of problem solving [11]. Taken together, these studies suggest that mathematics achievement should be examined not only as an overall score but also according to students' initial abilities and the mathematical content domains represented in assessment tasks.

One relevant dimension of this variation is Prior Mathematical Ability (KAM), which can be used to classify students according to their initial level of mathematical competence. Previous studies have differentiated students into high, medium, and low KAM groups and identified differences in mathematical literacy and problem-solving performance across these groups [12], [13]. Students with higher initial mathematical ability tend to demonstrate more complete problem-solving processes, whereas students with medium and lower initial abilities may encounter difficulties at several stages [13]. In the present study, preliminary data from 158 seventh-grade students at SMP Negeri 13 Medan show that 26 students (16.5%) are classified as high KAM (score ≥ 39.04), 112 students (70.9%) as medium KAM, and 20 students (12.7%) as low KAM (score ≤ 20.28), with a mean KAM score of 29.66 and a standard deviation of 9.38. This distribution provides a school-level

context for examining whether mathematics achievement differs across students with different levels of initial mathematical ability.

The issue of variation in KAM is particularly relevant to the implementation of the Merdeka Curriculum, which emphasizes learning that responds to students' needs, readiness, and learning characteristics [14]. However, studies have reported that mathematics teachers continue to experience challenges in understanding and implementing differentiated learning and in using digital learning media to accommodate diverse student needs [15]. Other reported challenges include learning planning, preparation of instructional materials, selection of learning models, assessment, and adaptation to students' basic abilities [16]. These findings underscore the importance of obtaining empirical evidence about students' actual achievement so that instructional evaluation is informed by observed performance rather than by curriculum expectations alone.

The Merdeka Curriculum was introduced progressively in Indonesia beginning in 2022 and subsequently established as part of the national curriculum framework through Ministerial Regulation Number 12 of 2024. For Phase D mathematics in Grades VII–IX, the expected learning outcomes encompass mathematical concepts and skills in numbers, algebra, measurement, geometry, data analysis, and probability, including their application to mathematical problem solving [17]. The curriculum promotes flexible and contextual learning with attention to students' understanding and learning progress [14]. National Assessment analyses have reported improvements in literacy and numeracy outcomes among educational units implementing the Merdeka Curriculum, although such aggregate evidence does not directly describe achievement patterns within a particular grade, school, or mathematical content domain [18].

Existing studies concerning the Merdeka Curriculum have predominantly examined implementation challenges from teachers' perspectives [15], [16], [19], the relevance of learning models to mathematics instruction [20], and the development of teaching modules or learning tools [21]. Consequently, there remains a need for school-level empirical evidence based directly on students' assessment responses, particularly evidence that simultaneously considers students' initial mathematical ability and achievement across mathematical content domains. This study addresses that gap by analyzing end-of-semester summative mathematics assessment answer sheets from 158 seventh-grade students at SMP Negeri 13 Medan. The use of assessment answer sheets provides direct documentary evidence of students' responses to the tasks administered in their school context and allows achievement to be examined without attributing observed patterns to the implementation of the curriculum itself [22], [23].

The novelty of this study lies in the integration of three descriptive perspectives within a single school-level analysis: mathematics achievement according to KAM groups, the distribution of achievement categories, and achievement across mathematical content domains represented in the summative assessment. Rather than evaluating the Merdeka Curriculum through perceptions or implementation reports, the study uses students' assessment documents as the empirical basis for describing actual performance in Grade VII, the initial grade of Phase D. This approach provides a more focused contribution to curriculum-related research by connecting curriculum expectations with evidence from

students' completed assessment tasks while maintaining the descriptive scope of document analysis.

Accordingly, this study aims to answer three complementary research questions: (1) how do seventh-grade students' mathematics learning outcomes differ across KAM groups based on the N-Gain analysis; (2) how are students' mathematics learning outcomes distributed according to the achievement categories used in the study; and (3) which mathematical content domains demonstrate the highest and lowest achievement percentages based on the analysis of summative assessment answer sheets? The findings are expected to contribute school-level descriptive evidence for mathematics learning evaluation and provide diagnostic information that can support teachers and schools in identifying differences in achievement according to initial mathematical ability and content domain.

2. METHOD

2.1. Research Design

This study employed a quantitative descriptive design using document analysis as the primary data collection strategy. The quantitative descriptive design was selected because the study aimed to provide a systematic description of students' mathematics learning outcomes using numerical data without manipulating variables or establishing causal relationships [24]. Document analysis was employed because the summative assessment answer sheets constituted primary school-level documents containing students' responses to the administered assessment tasks. The analysis followed systematic procedures while considering the authenticity and relevance of the documents [22], [23].

2.2. Research Setting, Population, and Sample

The study was conducted at SMP Negeri 13 Medan during the 2025/2026 academic year. The population comprised all seventh-grade students who participated in the end-of-semester summative mathematics assessment within the Merdeka Curriculum context, totaling 158 students. A total sampling technique was applied because the entire accessible population was included in the analysis [25]. Thus, no individual student was excluded from the dataset, and the unit of analysis was the end-of-semester summative mathematics assessment answer sheet of each of the 158 students.

2.3. Data Sources and Assessment Materials

The main data source was the Grade VII end-of-semester summative mathematics assessment answer sheet, which was prepared by a team of mathematics teachers with reference to the Learning Outcomes for Phase D of the Merdeka Curriculum. The assessment consisted of 10 questions covering five mathematical content domains, with two questions representing each domain, as shown in Table 1.

To address the methodological requirements concerning the assessment instrument, the development of the 10-item assessment should be documented in terms of its content alignment, item format, scoring procedure, and evidence of validity and reliability. In particular, the manuscript should specify whether the items were teacher-developed, collaboratively developed by the mathematics teachers, or adapted from an existing

assessment instrument. The scoring rubric or scoring rules should also be stated, including how partial or incomplete responses were treated.

Table 1. Distribution of Question Items by Mathematical Content Domain

Mathematical Content Domain	Question Number	Number of Questions
Algebra	1 and 2	2
System of Linear Equations in One Variable	3 and 4	2
Social Arithmetic	5 and 6	2
Comparison	7 and 8	2
Lines and Angles	9 and 10	2

Because each mathematical domain was represented by only two items, the domain-level percentages are interpreted descriptively rather than as estimates of a highly precise domain-specific measurement. This limitation is considered when interpreting differences between material domains.

2.4. Initial Mathematical Ability (KAM)

Before the end-of-semester summative assessment was analyzed, students were classified according to their Initial Mathematical Ability (KAM). KAM scores were obtained from an initial mathematics test administered before the learning process. The manuscript should specify whether this KAM test was developed by the teachers/researchers or adopted from a standardized or previously validated instrument, including information about its content coverage, scoring range, and validity and reliability evidence.

KAM classification was based on the mean and standard deviation of the observed KAM scores, following a classification approach commonly used in mathematics education research [26]. The classification criteria were as follows:

High KAM: $KAM \geq (\bar{x} + SD)$

Medium KAM: $(\bar{x} - SD) < KAM < (\bar{x} + SD)$

Low KAM: $KAM \leq (\bar{x} - SD)$

where $\bar{x}$ represents the mean KAM score of all students and SD represents the standard deviation of the KAM scores. This procedure establishes relative boundaries based on the observed distribution rather than fixed absolute cut-off scores, thereby describing students' initial mathematical abilities within the context of the study [26].

2.5. Data Collection Procedure

The data collection process consisted of three stages. First, the researchers obtained the students' KAM scores from the initial mathematics test. Second, the end-of-semester summative mathematics assessment answer sheets were collected and organized according to student identification codes. Third, the assessment responses were scored and entered into a dataset for descriptive analysis according to KAM group and mathematical content domain.

Student names were replaced with identification codes during data processing to maintain confidentiality. The researchers should also state explicitly that permission to

access and analyze the students' assessment records was obtained from the relevant school authority before data analysis was conducted.

2.6. Measurement and Scoring

The final assessment contained 10 items distributed equally across five mathematical content domains. Each student's responses were scored according to the predetermined scoring rules and subsequently converted to a 0–100 scale. The manuscript should explicitly report the maximum raw score, the scoring weight assigned to each item, whether all items had equal weights, and the procedure used to convert raw scores into the 0–100 scale.

Content validity should be reported based on the actual procedure used to establish the alignment between the assessment items and the Phase D Learning Outcomes. If expert judgment, a teacher review panel, a test blueprint, or another validation procedure was used, the procedure and results should be described. Reliability evidence should likewise be reported if it was calculated. Item difficulty and discrimination indices should be reported if they were available from the assessment analysis.

2.7. Data Analysis

Data analysis consisted of four complementary procedures.

Procedure 1: Analysis of the Overall Distribution of Student Achievement

Each student's total assessment score was converted to a 0–100 scale and classified into four achievement categories used in the study, as presented in Table 2. Descriptive statistics, including the mean, median, standard deviation, minimum score, and maximum score, were calculated to describe the overall distribution of student achievement.

Table 2. Categories of Mathematics Learning Outcomes

Category	Value Range	Description of Achievements
Very Proficient	90 – 100	Exceeding minimum achievement
Proficient	75 – 89	Meet minimum targets
Quite Proficient	60 – 74	Approaching minimum achievement
Need Guidance	< 60	Not yet reached the minimum achievement

Source:[14]

Procedure 2: Analysis of Achievement Percentage by Mathematical Content Domain

The achievement percentage for each mathematical content domain was calculated from the two assessment items representing that domain. The calculation used the following formula [27]:

$$\text{Achievement Percentage} = (\text{Total Score Obtained} / \text{Maximum Possible Score}) \times 100\%$$

The resulting percentages were then interpreted using the criteria presented in Table 3.

Table 3. Interpretation Criteria for Mathematical Content Achievement

Percentage	Criteria
81% – 100%	Very high
61% – 80%	High
41% – 60%	Moderate
21% – 40%	Low
0% – 20%	Very Low

Source: [27]

Procedure 3: Analysis of Achievement by KAM Group

This procedure was used to describe the distribution of students' summative assessment achievement across the high, medium, and low KAM groups. For each group, the mean summative assessment score, the percentage of students classified as Proficient or Very Proficient, and the achievement percentage for each mathematical content domain were calculated. Differences between KAM groups were presented descriptively using comparative tables and graphs to identify patterns of achievement across levels of initial mathematical ability [26].

Procedure 4: N-Gain Analysis by KAM Group

N-Gain analysis was used to describe the normalized change between students' initial mathematics performance and their final summative assessment performance [28]. The N-Gain value was calculated for each student and then averaged within each KAM group. The interpretation criteria are presented in Table 4 [28].

However, because N-Gain is most appropriately interpreted when the pre- and post-assessments measure the same or closely comparable construct, the comparability of the KAM test and the final summative assessment must be explicitly established in this study. The manuscript should report whether the two assessments used the same score range or were transformed to a common scale, whether their maximum scores were equivalent, and whether they represented comparable mathematical content domains. If the two assessments differed substantially in construct, content coverage, or scoring scale, the N-Gain results should be interpreted as a descriptive index of score change rather than as a direct measure of learning gain.

Table 4. N-Gain Interpretation Criteria

N-Gain Value	Criteria
$N\text{-Gain} \geq 0.70$	High
$0.30 \leq N\text{-Gain} < 0.70$	Moderate
$N\text{-Gain} < 0.30$	Low

Source:[28]

2.8. Ethical Considerations

The study used students' educational assessment records; therefore, confidentiality and responsible handling of student data were maintained throughout the research process. Student names were not reported in the dataset or manuscript, and identification codes were used during data analysis. Access to the assessment documents was conducted with permission from the relevant school authority. The manuscript should report the specific

institutional or school permission procedure actually obtained before the records were analyzed.

3. RESULTS

3.1. Overview of the Data

The study analyzed 158 end-of-semester summative mathematics assessment answer sheets from seventh-grade students at SMP Negeri 13 Medan during the 2025/2026 academic year. Before the semester learning activities, all students completed an Initial Mathematical Ability (KAM) test, which was used to describe their initial mathematical ability and to classify them into three KAM groups. The end-of-semester summative assessment was subsequently used to describe students' final mathematics achievement.

Table 5. Descriptive Statistics of KAM and Final Summative Assessment

Statistics	KAM / Pre-test	Final Exam / Post-test
Number of Students (n)	158	158
Average	29.66	72.85
Median	26.00	73.00
Standard Deviation	9.38	9.41
Minimum	12	45
Maximum	60	100

Table 5 shows that the mean score increased from 29.66 on the initial KAM test to 72.85 on the final assessment, representing a descriptive difference of 43.20 points. The initial scores ranged from 12 to 60, whereas the final assessment scores ranged from 45 to 100. Because the study employed a descriptive design and did not conduct an inferential test comparing the two scores, the difference is reported as a descriptive change rather than as evidence of statistically significant improvement.

3.2. Classification of Initial Mathematical Ability

KAM classification was based on the sample mean and standard deviation [26]. With a mean of 29.66 and a standard deviation of 9.38, the upper boundary for the High KAM group was 39.04, while the lower boundary for the Low KAM group was 20.28.

Table 6. Distribution of Students by KAM Group

KAM Group	Criteria	Number of Students	Percentage
Tall	$X \geq 39.04$	26	16.5%
Midle	$20.28 < X < 39.04$	112	70.9%
Low	$X \leq 20.28$	20	12.7%

The Medium KAM group constituted the largest proportion of the sample, with 112 students (70.9%), followed by the High KAM group with 26 students (16.5%) and the Low KAM group with 20 students (12.7%). This distribution indicates substantial variation in students' initial mathematical ability and provides the basis for examining differences in subsequent descriptive achievement patterns across KAM groups.

3.3. Overall Mathematics Learning Achievement

Final assessment scores were classified into four achievement categories based on the categories used in the Merdeka Curriculum learning and assessment framework [14].

Table 7. Distribution of Learning Outcomes by Achievement Category

Category	Score Range	Description	n	%
Very Proficient	90 – 100	Exceeding minimum achievement	11	7.0%
Proficient	75 – 89	Meeting minimum targets	63	39.9%
Quite Proficient	60 – 74	Approaching minimum achievement	77	48.7%
Need Guidance	< 60	Not yet reaching the minimum	7	4.4%

Overall, 74 students (46.8%) were classified as Proficient or Very Proficient, whereas 77 students (48.7%) were classified as Quite Proficient. Only 7 students (4.4%) were in the Need Guidance category. The mean final assessment score of 72.85 therefore remained within the Quite Proficient range and was close to the Proficient threshold of 75. This school-level pattern can be considered alongside national evidence reporting changes in literacy and numeracy outcomes among educational units implementing the Merdeka Curriculum [18]. However, the national evidence is based on an aggregate population and different measurement context; therefore, it should not be interpreted as directly representing the achievement of the students in the present study.

3.4. Achievement by Mathematical Content Domain

Achievement was analyzed across five mathematical content domains, each represented by two assessment items. The results are presented in Table 8.

Table 8. Achievement by Mathematical Content Domain

Mathematical Content Domain	Question Numbers	Item Mean Scores	Achievement (%)	Criteria
Algebra	1 & 2	8.00 / 6.66	73.32%	Tall
SPLSV	3 & 4	6.52 / 7.70	71.11%	Tall
Social Arithmetic	5 & 6	7.75 / 8.01	78.80%	Tall
Comparison	7 & 8	7.61 / 6.90	72.56%	Tall
Lines and Angles	9 & 10	6.79 / 6.91	68.48%	Tall

All five content domains fell within the High achievement category. Social Arithmetic recorded the highest percentage (78.80%), whereas Lines and Angles recorded the lowest (68.48%). Algebra, Comparison, and Systems of Linear Equations in One Variable recorded 73.32%, 72.56%, and 71.11%, respectively.

The relatively higher achievement in Social Arithmetic may be related to the contextual character of the material and its connection with situations that students commonly encounter in everyday life. However, the present data do not permit the contextual nature of the material to be established as the cause of the higher achievement. In contrast, the lower percentage for Lines and Angles is consistent with previous research reporting conceptual, procedural, and technical errors among junior high school students when solving problems involving lines and angles [11].

Within the Algebra domain, the mean scores for Items 1 and 2 also differed (8.00 and 6.66, respectively), indicating that achievement may vary according to the specific demands of individual items even within the same content domain. Because each domain was represented by only two items, these percentages should be interpreted as descriptive indicators of performance on the assessed content rather than comprehensive measures of mastery of the entire domain.

3.5. N-Gain by KAM Group

N-Gain was calculated to describe the relative difference between the initial KAM scores and final assessment scores using the normalized-gain approach [28]. The results are presented in Table 9.

Table 9. N-Gain by KAM Group

KAM Group	n	Mean KAM	Mean Final Assessment	Absolute Gain	N-Gain	Category
High	26	46.62	82.12	35.50	0.6646	Moderate
Moderate	112	27.63	70.77	43.14	0.5940	Moderate
Low	20	19.00	72.50	53.50	0.6605	Moderate
Overall	158	29.66	72.85	43.20	0.6140	Moderate

All three KAM groups produced moderate N-Gain values, with values of 0.6646 for High KAM, 0.5940 for Medium KAM, and 0.6605 for Low KAM [28]. The overall N-Gain was 0.6140, which was also classified as Moderate. These values describe relative score change according to the N-Gain classification and do not constitute evidence of statistical significance or causal learning effects.

The High KAM group had the highest N-Gain (0.6646), followed closely by the Low KAM group (0.6605), while the Medium KAM group recorded the lowest value (0.5940). The relatively lower N-Gain of the Medium group should be interpreted cautiously because N-Gain is mathematically affected by both the initial score and the available room for improvement. The pattern may also reflect differences in score distribution, item characteristics, or measurement properties rather than a substantive difference in learning response. In absolute terms, the Low KAM group showed the largest score increase, whereas the High KAM group showed the smallest.

The difference between absolute gain and N-Gain is important. Absolute gain indicates the raw increase in scores, whereas normalized gain expresses the increase relative to the maximum possible improvement [28]. Therefore, the larger absolute gain observed in the Low KAM group does not necessarily imply a larger normalized learning gain.

3.6. Achievement Categories by KAM Group

Table 10. Distribution of Achievement Categories by KAM Group

KAM Group	n	Very Proficient	Proficient	Quite Proficient	Need Guidance
Tall	26	8 (30.8%)	15 (57.7%)	3 (11.5%)	0 (0.0%)
Midle	112	3 (2.7%)	40 (35.7%)	62 (55.4%)	7 (6.3%)
Low	20	0 (0.0%)	8 (40.0%)	12 (60.0%)	0 (0.0%)

Table 10 shows a clear difference in the distribution of final achievement categories across the three KAM groups. Among students with High KAM, 88.5% were classified as Proficient or Very Proficient. In the Medium KAM group, most students were in the Quite Proficient category (55.4%), while the Low KAM group contained 60.0% of students in the Quite Proficient category and 40.0% in the Proficient category.

The absence of students in the Need Guidance category within the Low KAM group should not be interpreted as evidence that low initial ability was associated with superior learning outcomes. The group was relatively small ($n = 20$), and the categorical distribution does not account for the different starting points or measurement characteristics of the two assessments. The results are therefore most appropriately interpreted as descriptive evidence of different achievement distributions across initial-ability groups.

3.7. Achievement by Mathematical Content Domain and KAM Group

Table 11. Achievement by Mathematical Content Domain and KAM Group

Mathematical Content Domain	High KAM	Criteria	Medium KAM	Criteria	Low KAM	Criteria
Algebra	79.42%	High	71.96%	High	73.00%	High
SPLSV	78.65%	High	69.38%	High	71.00%	High
Social Arithmetic	89.04%	Very High	76.47%	High	78.50%	High
Comparison	85.96%	Very High	69.73%	High	71.00%	High
Lines and Angles	77.50%	High	66.29%	High	69.00%	High

Table 11 shows that the High KAM group achieved Very High performance in Social Arithmetic (89.04%) and Comparison (85.96%), while its other three domains were classified as High. The Medium and Low KAM groups were classified as High across all five domains.

Lines and Angles produced the lowest domain-level achievement within each KAM group, although the specific percentages differed across groups. The Medium KAM group recorded the lowest percentage for this domain (66.29%), followed by the Low KAM group (69.00%) and High KAM group (77.50%). This consistency across groups suggests that Lines and Angles warrants attention as an assessment domain, although the present data cannot establish the specific instructional or cognitive reasons for its lower achievement.

4. DISCUSSION

The findings provide three main descriptive insights. First, all KAM groups showed moderate N-Gain values, although the magnitude of absolute and normalized change differed across groups. Second, nearly half of the students were classified as Proficient or Very Proficient, while a similar proportion remained in the Quite Proficient category. Third, achievement differed across mathematical content domains, with Social Arithmetic showing the highest percentage and Lines and Angles the lowest.

The N-Gain results indicate that the High and Low KAM groups had very similar normalized gains despite substantially different initial scores, whereas the Medium KAM group recorded a somewhat lower N-Gain. This pattern should not be interpreted as evidence that one KAM group benefited more from instruction. N-Gain is influenced by the starting score and the maximum attainable score, and the use of different instruments for the initial

and final measurements further limits direct interpretation of the values [28]. The lower N-Gain of the Medium group may therefore reflect a combination of initial-score distribution, available room for improvement, item characteristics, and measurement differences between the KAM test and the final assessment. Because the present study did not use equivalent pre- and post-tests or an experimental comparison group, these possible explanations cannot be empirically distinguished.

The findings also illustrate the importance of distinguishing absolute gain from normalized gain. The Low KAM group had the largest absolute increase, while the High KAM group had the highest normalized gain by a small margin. These measures answer different descriptive questions and should therefore not be treated as interchangeable [28]. Previous studies have likewise shown that students with different levels of initial mathematical ability may demonstrate different profiles of mathematical literacy and problem-solving performance [12], [13]. In the present study, KAM is therefore more appropriately viewed as a diagnostic classification for describing achievement patterns than as a determinant of students' final outcomes.

The distribution of final achievement indicates that the class was concentrated around the boundary between Quite Proficient and Proficient performance. This pattern provides a more nuanced picture than the overall mean alone because students with similar class-level averages may have substantially different achievement distributions. The national evidence concerning literacy and numeracy outcomes in educational units implementing the Merdeka Curriculum provides a broader policy context [18], but the difference in population, assessment, and measurement level means that the present school-level results cannot be directly compared with or attributed to those national trends.

The content-domain analysis provides a more specific diagnostic perspective. Social Arithmetic showed the highest achievement, whereas Lines and Angles showed the lowest achievement across the overall sample. The relatively stronger performance in Social Arithmetic may be associated with the contextual familiarity of the content, but this explanation remains speculative because the study did not collect data on students' perceptions, prior experiences, or cognitive processes. The lower performance in Lines and Angles is consistent with previous research documenting conceptual, procedural, and technical errors in this mathematical domain [11]. The consistency of the relatively lower Lines and Angles scores across KAM groups suggests that the issue should not be considered exclusively a characteristic of students with lower initial mathematical ability.

The relatively low achievement in Lines and Angles also provides a basis for considering possible instructional alternatives. GeoGebra-assisted instruction has been reported as a potential means of supporting students' conceptual understanding of geometry, while systematic evidence has documented increasing interest in technology-supported ethnomathematics in Indonesia [29], [30]. These approaches are best regarded as possible directions for future instructional development rather than as interventions demonstrated to be effective by the present study. Experimental or quasi-experimental research would be needed to determine whether such approaches improve achievement in the specific school context examined here.

The psychometric characteristics of the assessment should also be considered when interpreting the domain-level findings. Each mathematical content domain was represented by only two items, meaning that a difference in performance on one item could substantially influence the percentage assigned to that domain. The difference between the two Algebra items illustrates this issue. Accordingly, the domain percentages provide useful diagnostic information about the assessed items but should not be interpreted as comprehensive estimates of students' mastery of the full mathematical domains. Future studies should employ a larger number of items per domain and report evidence of content validity, reliability, item difficulty, and discrimination.

Another important limitation concerns the use of the KAM test as the pre-test and the final summative assessment as the post-test for N-Gain analysis. N-Gain is more readily interpretable when the pre- and post-tests measure the same or closely comparable construct, use comparable score scales, and represent similar content. In this study, the KAM test was designed to classify initial mathematical ability, whereas the final assessment measured achievement after one semester across five specified content domains. Therefore, the N-Gain values should be interpreted as descriptive normalized score changes rather than as definitive measures of learning gain. Future research should use equivalent or parallel pre- and post-tests with clearly documented construct alignment if the primary objective is to estimate learning gain.

The document-analysis design provides a direct source of evidence because students' actual assessment responses were analyzed rather than relying exclusively on perceptions or self-reports. Document analysis can provide systematic evidence from existing educational records when the documents are authentic, relevant, and analyzed using an explicit procedure [22], [23]. Nevertheless, the present study was limited to one school, one grade level, one semester, and a 10-item summative assessment. The findings therefore describe the particular cohort and assessment context rather than representing all seventh-grade students or all schools implementing the Merdeka Curriculum.

Overall, the findings suggest that the assessment data can serve as a diagnostic basis for identifying differences in achievement across initial-ability groups and mathematical content domains. The results particularly indicate the need for closer examination of students whose achievement remains around the Proficient threshold and of content domains such as Lines and Angles that consistently showed comparatively lower performance. However, these implications concern instructional diagnosis rather than evidence of curriculum effectiveness. Future studies should employ psychometrically stronger and content-balanced assessments, equivalent pre- and post-measures, multiple schools, and longitudinal or comparative designs to examine learning development more rigorously.

5. CONCLUSION

This study examined the mathematics learning outcomes of Grade VII students within the Merdeka Curriculum by analyzing their Initial Mathematics Ability and end-of-semester summative assessment results across the assessed mathematical domains. The findings indicate moderate descriptive progress in students' mathematics performance, with most students performing at or near the expected proficiency level. Among the assessed

domains, geometry emerged as the weakest area, indicating a need for greater attention to students' understanding and application of geometric concepts.

These findings provide a descriptive picture of students' mathematics achievement and differences in performance across mathematical domains. However, the results do not demonstrate the effectiveness of the Merdeka Curriculum, because the study did not employ an experimental or quasi-experimental design with a comparison group. The observed changes may also reflect differences in assessment characteristics, students' initial abilities, and other instructional factors.

From a practical perspective, teachers can use domain-level assessment results to identify areas of difficulty, particularly geometry, and provide targeted learning support based on students' initial mathematical abilities. For future research, longitudinal or quasi-experimental studies using comparable pre- and post-assessment instruments and larger samples across multiple schools are needed to examine changes in mathematics achievement and the factors associated with them.

REFERENCES

- [1] E. Isfayani, "Analisis kesulitan belajar matematika materi bentuk aljabar pada siswa SMP kelas VII," *J. Pendidik. Mat. Malikussaleh*, vol. 3, no. 1, pp. 79–90, 2023, doi: 10.29103/jpmm.v3i1.11177.
- [2] N. Munzayanah, B. Usodo, L. Fitriana, and F. Nurhasanah, "Thinking processes in numeracy: Influencing factors, content, and varieties of thinking processes," *Mathline J. Mat. dan Pendidik. Mat.*, vol. 10, no. 4, pp. 843–858, 2025, doi: 10.31943/mathline.v10i4.1003.
- [3] M. A. Adeoye and H. A. Jimoh, "Problem-solving skills among 21st-century learners toward creativity and innovation ideas," *Think. Ski. Creat. J.*, vol. 6, no. 1, pp. 52–58, 2023, doi: 10.23887/tscj.v6i1.62708.
- [4] N. K. Mendrofa, "Computational thinking skills in 21st century mathematics learning," *JiIP J. Ilm. Ilmu Pendidik.*, vol. 7, no. 1, pp. 792–801, 2024, doi: 10.54371/jiip.v7i1.3780.
- [5] N. Sridana and K. Sarjana, "The implementation of mathematics learning in the context of 21st century skill competencies in junior high schools," in *Proc. 1st Annu. Conf. Educ. Soc. Sci. (ACCESS 2019), Adv. Soc. Sci., Educ. Humanit. Res.*, vol. 465, pp. 283–285, 2020, doi: 10.2991/assehr.k.200827.071.
- [6] V. Anand, "21st-century learning skills and NEP 2020," *Int. J. Multidiscip. Res.*, vol. 6, no. 6, pp. 1–10, 2024, doi: 10.36948/ijfmr.2024.v06i06.34134.
- [7] OECD, *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Paris, France: OECD Publishing, 2023, doi: 10.1787/53f23881-en.
- [8] T. T. Wijaya, W. Hidayat, N. Hermita, J. A. Alim, and C. A. Talib, "Exploring contributing factors to PISA 2022 mathematics achievement: Insights from Indonesian teachers," *Infinity J.*, vol. 13, no. 1, pp. 139–156, 2024, doi: 10.22460/infinity.v13i1.p139-156.
- [9] S. C. Earthadara and Kismiantini, "Path analysis of teachers' cognitive activation and students' mathematics achievement in PISA 2022 Indonesia," *J. Ris. Pendidik. Mat.*, vol. 12, no. 2, pp. 266–279, 2025, doi: 10.21831/jrpm.v12i2.84865.
- [10] A. S. Setiani, H. S. Lukman, and N. Agustiani, "Analisis kemampuan pemecahan masalah matematis siswa SMP berdasarkan indikator Krulik dan Rudnick ditinjau dari motivasi belajar," *AKSIOMA J. Progr. Stud. Pendidik. Mat.*, vol. 13, no. 2, pp. 656–670, 2024, doi: 10.24127/ajpm.v13i2.8612.
- [11] M. Elvi and N. A. R. Siregar, "Identifikasi kesalahan siswa menyelesaikan soal matematika garis dan sudut," *Pythagoras J. Progr. Stud. Pendidik. Mat.*, vol. 12, no. 2, pp. 128–139, 2023, doi: 10.33373/pythagoras.v12i2.5571.
- [12] A. Muzaki and Masjudin, "Analisis kemampuan literasi matematis siswa," *Mosharafa J. Pendidik. Mat.*, vol. 8, no. 3, pp. 493–502, 2019, doi: 10.31980/mosharafa.v8i3.584.
- [13] K. Nisa, N. Sridana, N. H. Salsabilla, and L. Hayati, "Deskripsi kemampuan pemecahan masalah siswa ditinjau kemampuan awal matematis," *J. Classr. Action Res.*, vol. 5, no. 3, pp. 17–24, 2023, doi: 10.29303/jcar.v5i3.4884.
- [14] Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, *Keputusan Kepala Badan Standar, Kurikulum, dan Asesmen Pendidikan Nomor 033/H/KR/2022 tentang Perubahan atas Keputusan Kepala BSKAP Nomor 008/H/KR/2022 tentang Capaian Pembelajaran pada Pendidikan Anak Usia*

- Dini, *Jenjang Pendidikan Dasar, dan Pendidikan Menengah pada Kurikulum Merdeka*. Jakarta, Indonesia, 2022.
- [15] M. N. Sholihat, U. Koswara, and D. Irawan, "Analisis pemahaman guru matematika SMP terhadap pembelajaran berdiferensiasi dan pemanfaatan media digital di Kabupaten Sumedang," *SIGMA J. Pendidik. Mat.*, vol. 16, no. 2, pp. 500–512, 2024, doi: 10.26618/sigma.v16i2.16329.
- [16] J. H. Lumbantoruan and M. R. R. Simorangkir, "Kesulitan guru dalam mengimplementasikan Kurikulum Merdeka Belajar materi matematika di Sekolah Menengah Pertama (SMP)," *AKSIOMA J. Progr. Stud. Pendidik. Mat.*, vol. 12, no. 1, pp. 1459–1473, 2023, doi: 10.24127/ajpm.v12i1.7082.
- [17] Badan Standar, Kurikulum, dan Asesmen Pendidikan, *Keputusan Kepala BSKAP Nomor 032/H/KR/2024 tentang Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah pada Kurikulum Merdeka*. Jakarta, Indonesia, 2024.
- [18] I. Zamjani et al., *Implementasi Kurikulum Merdeka dan Peningkatan Kemampuan Literasi dan Numerasi Peserta Didik*. Jakarta, Indonesia: Pusat Standar dan Kebijakan Pendidikan, Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2024.
- [19] D. Hardiany, "Pelaksanaan implementasi Kurikulum Merdeka pada pembelajaran matematika di SMP se-Kabupaten Kayong Utara," *Jiip J. Ilm. Ilmu Pendidik.*, vol. 7, no. 4, pp. 3924–3930, 2024, doi: 10.54371/jiip.v7i4.4266.
- [20] K. H. Agustina, M. A. N. Fahmi, Z. Nayyiroh, E. Yudianto, N. D. S. Lestari, and Sunardi, "Relevansi penerapan model pembelajaran terhadap pelajaran matematika berdasarkan Kurikulum Merdeka," *Aritmatika J. Ris. Pendidik. Mat.*, vol. 5, no. 1, pp. 26–38, 2024, doi: 10.35719/aritmatika.v5i1.293.
- [21] D. Ristayani, Wulandari, and E. Isfayani, "Pengembangan modul ajar Kurikulum Merdeka berbasis etnomatematika kain Ulos Sumatera Utara," *J. Penelit. Pembelajaran Mat. Sekol.*, vol. 8, no. 3, pp. 410–419, 2024, doi: 10.33369/jp2ms.8.3.410-419.
- [22] G. A. Bowen, "Document analysis as a qualitative research method," *Qual. Res. J.*, vol. 9, no. 2, pp. 27–40, 2009, doi: 10.3316/QRJ0902027.
- [23] U. B. Özkan, "Doküman inceleme yönteminde geçerlik ve güvenirlik: Eğitim bilimleri araştırmaları bağlamında kuramsal bir inceleme," *Buca Eğitim Fakültesi Dergisi*, no. 56, pp. 832–848, 2023, doi: 10.53444/deubefd.1258867.
- [24] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 6th ed. Thousand Oaks, CA, USA: SAGE Publications, 2023.
- [25] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung, Indonesia: Alfabeta, 2019.
- [26] K. E. Lestari and M. R. Yudhanegara, *Penelitian Pendidikan Matematika*. Bandung, Indonesia: Refika Aditama, 2018.
- [27] Riduwan, *Skala Pengukuran Variabel-Variabel Penelitian*. Bandung, Indonesia: Alfabeta.
- [28] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *Am. J. Phys.*, vol. 66, no. 1, pp. 64–74, 1998, doi: 10.1119/1.18809.
- [29] F. Gurmu, C. Tuge, and A. B. Hunde, "Effects of GeoGebra-assisted instructional methods on students' conceptual understanding of geometry," *Cogent Educ.*, vol. 11, no. 1, Art. no. 2379745, 2024, doi: 10.1080/2331186X.2024.2379745.
- [30] G. Sunzuma and U. Umbara, "Ethnomathematics-based technology in Indonesia: A systematic review," *Asian J. Math. Educ.*, vol. 4, no. 1, pp. 129–153, 2025, doi: 10.1177/27527263241305812.
-