PiJ Panicgogy International Journal

Publication details, including instructions for authors and subscription information:
https://nakiscience.com/index.php/pij

Realistic mathematics education and junior high school students' mathematical reasoning, communication, and self-efficacy: A quasi-experimental study

Amrullah Amrullah[1], Jarnawi Afgani Dahlan[2], Endang Wahyuningrum[3]

[1]Master's Program in Postgraduate School, Universitas Terbuka, Indonesia, amrullah@gmail.com

[2]Department of Mathematics Education, Universitas Pendidikan Indonesia, Indonesia, jarnawi@upi.edu

[3]Department of Mathematics Education, Universitas Terbuka, Indonesia, endangw@ecampus.ut.ac.id

To cite this article:
Amrullah, A., Dahlan, J. A., & Wahyuningrum, E. (2026). Realistic mathematics education and junior high school students' mathematical reasoning, communication, and self-efficacy: A quasi-experimental study. *Panicgogy International Journal, 4*(1), 79 – 97.
To link to this article:
https://nakiscience.com/index.php/pij/
Published by:
Nashir Al-Kutub Indonesia
Residential Street Kila Rengganis, Block I, Number 11, Labuapi, Indonesia, 83361

Panicgogy International Journal, 4(1), 79 – 97, 2026
https://nakiscience.com/index.php/pij

Original Article

Realistic mathematics education and junior high school students' mathematical reasoning, communication, and self-efficacy: A quasi-experimental study

Amrullah Amrullah[1*], Jarnawi Afgani Dahlan[2], Endang Wahyuningrum[3]
[1]Master's Program in Postgraduate School, Universitas Terbuka, Indonesia, amrullah@gmail.com
[2]Department of Mathematics Education, Universitas Pendidikan Indonesia, Indonesia, jarnawi@upi.edu
[3]Department of Mathematics Education, Universitas Terbuka, Indonesia, endangw@ecampus.ut.ac.id

*Correspondence Author: amrullah@gmail.com

Abstract
Mathematical reasoning, mathematical communication, and self-efficacy are three interrelated competencies in mathematics learning, yet they are frequently neglected in conventional instruction that emphasizes rote procedural memorization. This study examined the effect of the Realistic Mathematics Education (RME) approach on eighth-grade students' mathematical reasoning ability, mathematical communication ability, and self-efficacy on the topic of circles. A quantitative approach with a nonequivalent control group quasi-experimental design was employed, involving 25 students in an experimental group who received RME-based instruction and 24 students in a control group who received conventional instruction, selected through purposive sampling at a private junior high school in Batam, Indonesia. Data were collected using essay-type tests of reasoning and communication ability together with a Likert-scale self-efficacy questionnaire, and were analyzed using one-sample and independent-samples t-tests after normality and homogeneity assumptions were confirmed. The results showed that the RME approach did not produce a statistically significant effect on mathematical reasoning ability ($p = 0.419$), but produced a significant effect on mathematical communication ability ($p = 0.028$) and self-efficacy ($p = 0.002$) compared with conventional instruction. These findings suggest that students' engagement in contextual problems, group discussion, and presentation of solutions within the RME stages was more effective in fostering the ability to communicate mathematical ideas and self-confidence than in advancing formal reasoning within a relatively short intervention period. This study contributes to the RME literature by demonstrating that the effectiveness of an instructional approach is multidimensional and uneven across mathematical competencies, implying that teachers need to design additional reinforcement for reasoning when implementing RME over a short duration.
Keywords: Realistic Mathematics Education; mathematical reasoning; mathematical communication; self-efficacy; quasi-experiment

1. Introduction

Mathematics plays a strategic role in developing logical, analytical, systematic, and creative thinking skills that students need in order to participate productively in a global society. Mathematics is not merely a collection of numbers and theorems to be memorized; rather, it is a vehicle for developing important competencies such as problem solving, reasoning, and communication that students will continue to rely on throughout their lives. Among these competencies, mathematical reasoning and mathematical communication occupy a central position because they reinforce one another: reasoning enables students to detect patterns, formulate conjectures, and construct valid arguments according to Lithner (2008), while communication allows those arguments to be conveyed, negotiated, and jointly understood within a learning community according to Jeannotte and Kieran (2017). Beyond these cognitive aspects, the affective dimension of self-efficacy, that is, students' belief in their own capability to complete mathematical tasks, also determines the extent to which students persist when faced with challenging problems (Bandura, 1977); Bandura (2010). These three competencies have been shown empirically to be interrelated and to jointly shape students' long-term mathematics achievement, with self-efficacy and mathematics achievement in particular exhibiting a reciprocal, mutually reinforcing relationship over time (Liu, Jong, and Fan, 2024); (Schöber, Schütte, Köller, McElvany, and Gebauer, 2018).

Classroom realities, however, reveal a considerable gap between the importance of these competencies and students' actual attainment. Preliminary observation in an eighth-grade class at a private junior high school in Batam indicated that students' mathematical reasoning ability was moderate to low: most students struggled to solve problems that differed in form from the worked examples given by the teacher, even when such problems required only minor mathematical manipulation. A similar difficulty was observed in mathematical communication ability, where students were often able to write down a final answer but unable to explain the underlying reasoning or solution steps in a coherent manner. This condition is consistent with findings from other settings reporting that students struggle to articulate the logical justification behind their answers, even when those answers are correct (Rohid and Rusmawati, 2019). Low performance on these two cognitive competencies is also frequently accompanied by low self-efficacy toward mathematics, as repeated experiences of failure on non-routine problems tend to erode students' confidence to attempt new strategies (Zakariya, 2021).

One suspected root cause of this condition is the continued dominance of teacher-centered conventional instruction, in which students mostly memorize formulas and imitate demonstrated solution steps without being actively engaged in discovering and making sense of mathematical concepts through contexts relevant to their own lives. Realistic Mathematics Education (RME) was developed as an alternative approach that places contextual situations at the starting point of instruction, enabling students to construct understanding through a gradual process of horizontal and vertical mathematization, as described by Van den Heuvel-Panhuizen and Drijvers (2020). Through the stages of understanding a contextual problem, solving the problem in one's own way, comparing and discussing strategies, and finally formalizing the concept, RME is theoretically positioned to support the simultaneous development of reasoning, communication, and self-confidence.

Empirical evidence on the effectiveness of RME with respect to these three variables remains mixed and inconclusive. Several studies have reported a significant effect of RME on mathematical reasoning (Rasyidah et al., 2026); (Siregar, Suryadi, Prabawanto, and Mujib, 2022), whereas other studies have found that the effect of RME on reasoning was not statistically significant despite a descriptive improvement (Palinussa, Molle, and Gaspersz, 2021). On problem solving and mathematical literacy, competencies closely related to reasoning, RME has likewise been shown to be effective across various educational levels (Anggraini and Fauzan, 2020); (Sumirattana, Makanong, and Thipkong, 2017), although this effectiveness does not always translate directly into stronger formal reasoning. A comparatively more consistent pattern appears for mathematical communication ability, which meta-analytic evidence shows to benefit fairly stably from RME across contexts (Akosah, Dissou Arthur, and Adu Obeng, 2026). In the domain of self-efficacy, intervention studies indicate that approaches emphasizing direct mastery experiences and group discussion are effective in strengthening students' confidence (Zakariya, 2022). Various efforts to design RME-based learning trajectories in Indonesia have also concentrated mainly on statistics and number-pattern topics (Fauzan, Musdi, and Afriadi, 2018); (Fauzan and Diana, 2020), as well as on the role of mathematical representation as a mediator between belief and problem solving (Yuanita, Zulnaidi, and Zakaria, 2018), while studies that examine all three cognitive and affective variables simultaneously within a single quasi-experimental design, particularly on the topic of circle geometry at the junior secondary level in a faith-based private school context in Indonesia, remain very limited. It is this research gap that motivates the present study, which examines the effect of RME on all three variables simultaneously and compares its effectiveness with that of conventional instruction. Beyond documenting three separate effects, examining these three variables within a single design enables their relative magnitude to be compared directly, revealing that RME's benefits are uneven across cognitive and affective competencies, a pattern not observable when each variable is studied in isolation; this constitutes the specific theoretical contribution of the present study.

Building on this background, rationale, and research gap, the present study addresses two research questions:

a) Is there a significant difference in students' mathematical reasoning ability, mathematical communication ability, and self-efficacy before and after RME-based instruction?
b) Does RME-based instruction produce a statistically greater improvement in mathematical reasoning ability, mathematical communication ability, and self-efficacy than conventional instruction?

Because the inferential analyses reported in the Results section compare posttest scores against a fixed mastery threshold and between groups, rather than testing paired pretest–posttest change within each group, RQ1 is addressed here through the one-sample effectiveness tests against the KKM threshold reported for each variable.

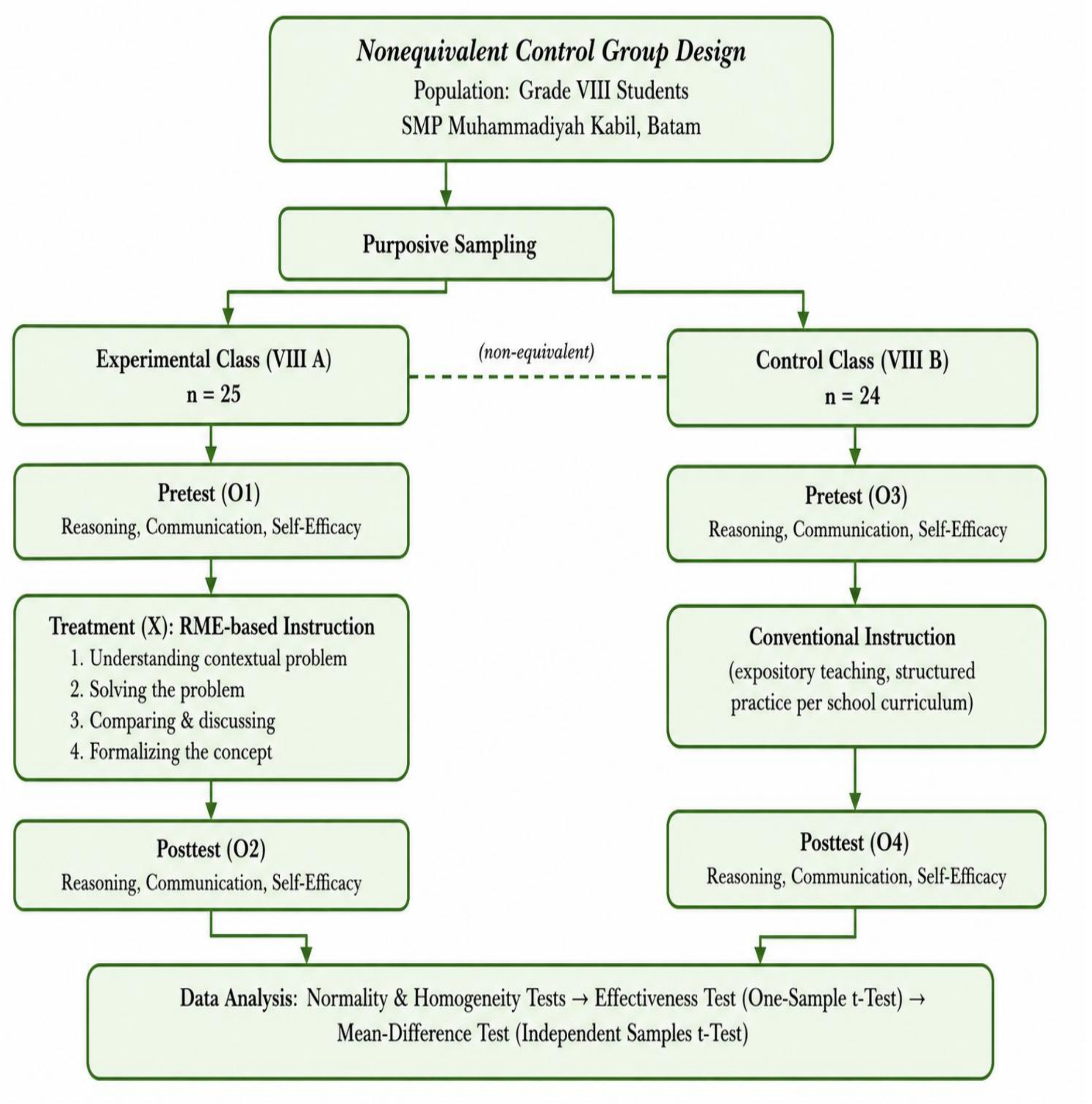
2. Method

2.1 Research Design

This study employed a quantitative approach using an experimental design, as it aimed to test the effect of the Realistic Mathematics Education (RME) approach on mathematical reasoning ability, mathematical communication ability, and self-efficacy in an objective and measurable manner. The design used was a nonequivalent control group design, involving two naturally formed classroom groups: an experimental group that received RME-based instruction and a control group that received conventional instruction. This quasi-experimental design was chosen because field conditions did not allow full randomization of students into research groups, so pre-existing classes were used as the units of analysis, a practice that is common and methodologically defensible in authentic educational field research (Capili and Anastasi, 2024). Both groups completed a pretest before treatment and a posttest after treatment to measure change on the three dependent variables.

Operationally, the research procedure was carried out through three main stages, as visualized in Figure 1. The first stage was the administration of a pretest to both groups to measure students' initial ability on the variables of mathematical reasoning, mathematical communication, and self-efficacy prior to treatment. The second stage was the delivery of the treatment, namely RME-based instruction for the experimental group through four sequential phases (understanding the contextual problem, solving the problem, comparing and discussing strategies, and formalizing the concept) designed to help students construct conceptual understanding gradually and meaningfully (Laurens, Batlolona, Batlolona, and Leasa, 2018), while the control group received conventional expository instruction as commonly used at the school on the same topic, the circumference and area of a circle. The third stage was the administration of a posttest to both groups to measure students' final ability, allowing the difference in learning outcomes between the experimental and control groups to be identified and statistically tested. The dashed line between the two groups in Figure 1 indicates that the experimental and control groups were non-equivalent, that is, not the result of full random assignment but pre-existing regular classes.

Figure 1
Operational flow of the nonequivalent control group research design



2.2 Participants

The population of this study consisted of all eighth-grade students at SMP Muhammadiyah Kabil, Batam, in the current academic year. The sampling technique used was purposive sampling, that is, a technique for determining a sample based on specific considerations related to previously known characteristics of the population rather than random selection (Tongco, 2007). Through this technique, two classes with relatively homogeneous characteristics and prior ability were selected to serve as the experimental and control groups, namely class VIII A as the experimental group and class VIII B as the control group. This selection was guided by considerations of the two classes' academic equivalence based on prior-semester mathematics report scores, as well as the fact that both classes were taught by the same teacher with an equivalent number of instructional hours, allowing confounding variables such as teaching quality and instructional time to be relatively well controlled. This assumption of baseline academic equivalence is further supported by pretest comparisons on all three outcome variables: mathematical reasoning, $t(47) = 1.49$, $p = .142$; mathematical communication, $t(47) = 0.84$, $p = .408$; and self-efficacy, $t(47) = 0.38$, $p = .704$, none of which differed significantly between groups prior to treatment.

A total of 49 students participated in this study, comprising 25 students in the experimental group and 24 students in the control group, as summarized in Table 1. The characteristics of participants in both groups were relatively equivalent in terms of age (13-14 years), grade level, and prior curriculum exposure, so that the differences in learning outcomes observed at posttest can be more confidently attributed to the instructional treatment rather than to pre-existing differences in participant characteristics. Given the modest total sample size ($N = 49$), a post hoc power analysis for the reasoning-ability comparison (observed $d = 0.24$) indicated substantially underpowered detection (power $\approx .13$ at $\alpha = .05$); with this sample, only effects of $d \geq 0.82$ could be reliably detected at 80% power, indicating that the nonsignificant reasoning result more plausibly reflects limited statistical power than a true absence of instructional impact.

Table 1
Distribution of Research Participants

Group	Class	Treatment	Number of Students (n)	Age Range
Experimental	VIII A	RME-based instruction	25	13–14 years
Control	VIII B	Conventional instruction	24	13–14 years
Total	—	—	49	—

2.3 Data Collection

Data in this study were collected using three main instruments: a mathematical reasoning ability test, a mathematical communication ability test, and a self-efficacy questionnaire. The reasoning and communication tests were each constructed as essay items developed from indicators of identifying a pattern, formulating a conjecture, and providing correct and complete justification for reasoning, and indicators of identifying and expressing information using mathematical notation, translating a problem into a diagram or mathematical model, and re-explaining a solution procedure for communication. The same reasoning and communication essay items were administered at both pretest and posttest, separated by the full treatment interval of [jumlah] weeks; while this design enables direct measurement of change, the possibility of testing or practice effects associated with item repetition is acknowledged as a limitation in the Discussion. The self-efficacy instrument was a five-point Likert-scale questionnaire measuring students' perceptions of their own capability to cope with challenging mathematical tasks, drawing on the concept of self-efficacy as an individual's belief in their capacity to organize and execute the actions required to attain a given outcome (Bandura, 1977).

All three instruments were piloted and validated, through expert content validation as well as internal-consistency reliability testing (Cronbach's alpha), which remains a relevant, though not exclusive, indicator of instrument consistency in mathematics education research when used appropriately (Zakariya, 2022); (Sullivan and Artino, 2013), before being used for actual data collection. Data were collected at two points: before treatment (pretest), to obtain a picture of students' initial ability on the three variables, and after treatment (posttest), to measure the change that occurred following RME-based instruction in the experimental group and conventional instruction in the control group. Data collection was conducted concurrently for both groups within a closely spaced time frame to minimize the influence of external factors unrelated to the treatment. Content validity of all three instruments was additionally established prior to the reliability testing reported above, through expert judgement by [jumlah] mathematics-education experts using [metode, mis. Aiken's V atau CVR], yielding a content validity index of [nilai]; item counts and scoring ranges were four items (score range 0–100) for the reasoning test, four items (score range 0–[skor maks]) for the communication test, and 30 items (score range 30–150) for the self-efficacy questionnaire.

2.4 Data Analysis

The collected data were analyzed in two stages: descriptive analysis and inferential analysis. Descriptive analysis was used to obtain the minimum, maximum, mean, standard deviation, and variance of pretest and posttest scores for both groups, and to classify student attainment into five categories (from very high to very low) based on the ideal score intervals of each instrument. Prior to inferential analysis, the data were tested for normality and homogeneity of variance. For the self-efficacy data, which are conceptually ordinal, the total score obtained by summing all questionnaire items was treated statistically as interval-level data, because its continuous range and approximately normal distribution, based on the central limit theorem, is a common practice in educational research analyzing composite Likert-scale scores (Norman, 2010); (Boone and Boone, 2012).

Inferential analysis was carried out through two types of tests. First, a one-sample t-test was used to determine whether RME-based instruction was effective in the experimental group, by comparing posttest scores against a threshold value converted from the school's Minimum Mastery Criterion (KKM) into the maximum score of each instrument. Second, an independent-samples t-test was used to determine whether there was a significant difference in learning outcomes between the experimental and control groups on the posttest scores of the three variables. The decision criterion for both tests used a significance level of 0.05, with the null hypothesis rejected when the significance value (p) was smaller than 0.05. In addition to statistical significance, effect size was also considered contextually in interpreting the practical meaningfulness of the findings, given that standardized effect-size benchmarks are not always uniformly applicable across educational research contexts (Bakker, Cai, English, Kaiser, Mesa, and Van Dooren, 2019); (Kraft, 2020). Interpretation of the quantitative results was also conducted with due caution regarding the specific material and grade level involved, as recommended in quantitative mathematics education research more broadly (Aksoy and Bostan, 2021). All statistical analyses were performed using statistical software at a 95% confidence level. Because pretest and control-group data were available, the one-sample effectiveness test against the KKM threshold was complemented by independent-samples comparisons of posttest scores, providing a more direct estimate of the treatment effect than a threshold comparison alone. Standardized effect sizes (Cohen's d, pooled SD) were computed for each between-group comparison: mathematical reasoning ability, $d = 0.24$, 95% CI [−0.33, 0.80] (small); mathematical communication ability, d = 0.66, 95% CI [0.08, 1.24] (medium); and self-efficacy, d = 0.97, 95% CI [0.38, 1.57] (large).

2.5 Ethical Clearance

This study was conducted with due regard to the ethical principles governing research involving minor participants. Prior to data collection, the researcher applied for and obtained ethical clearance from the relevant institution, as well as research permission from the school where the study was conducted. Because the participants were junior high school students classified as children and adolescents, the consent procedure followed a layered informed consent and assent protocol: written informed consent was obtained from parents or legal guardians as the students' legal representatives, accompanied by the students' own assent after they were given an explanation of the study's purpose, procedures, and benefits, as well as their right to decline participation or withdraw at any time without consequence, consistent with ethical guidelines for research involving children and adolescents (Askari et al., 2024).

Participant confidentiality was maintained throughout the research process by using participant codes rather than real names in all data and research documents, and the data were used solely for the academic purposes of this study. The instructional treatment delivered to both groups, whether RME-based or conventional, constituted a regular instructional activity that posed no risk beyond that of everyday classroom learning, and the study was therefore classified as minimal risk. To preserve instructional equity, the control group continued to receive quality mathematics instruction aligned with the school curriculum, and a summary of the research findings was shared with the school after the study concluded as a form of accountability and reciprocal benefit to the institution that granted research permission.

3. Results

This section presents the findings regarding the effect of the RME approach on three dependent variables: mathematical reasoning ability, mathematical communication ability, and students' self-efficacy. The results are presented sequentially, beginning with descriptive statistics of pretest and posttest scores for each variable, followed by categorization of student attainment, and concluding with inferential analysis through effectiveness testing and mean-difference testing. The descriptive presentation is intended to provide initial context on the condition of both groups before and after treatment, prior to the statistical testing of the research hypotheses. Instrument piloting prior to data collection indicated that all three instruments possessed adequate internal consistency for use in this study: the Cronbach's alpha coefficient was 0.71 for the mathematical reasoning test, 0.78 for the mathematical communication test, and 0.89 for the self-efficacy questionnaire, ranging from acceptable to very good for educational research instruments. The high reliability of the self-efficacy instrument is consistent with the characteristics of a Likert-scale questionnaire comprising many statement items, whose total score tends to be more stable than that of an essay test with fewer items.

3.1 Mathematical Reasoning Ability

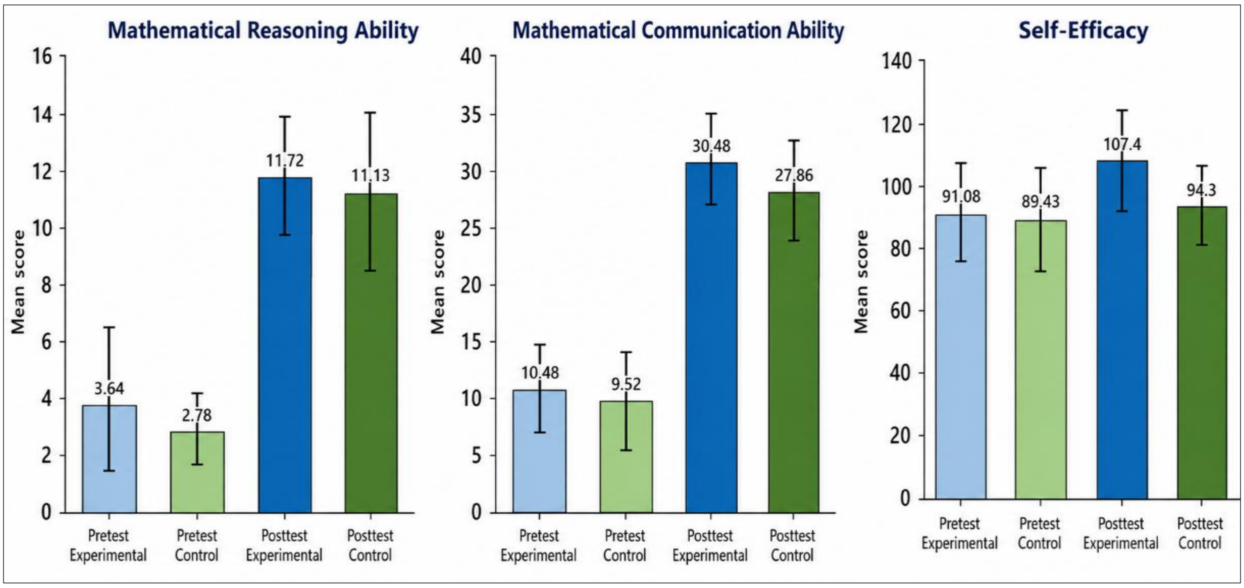
Mathematical reasoning ability was measured using four essay items requiring students to identify the relationship between a central angle and an inscribed angle, formulate a conjecture about an observed phenomenon, and analyze the formulas for the circumference and area of a circle. Descriptive statistics for pretest and posttest scores on mathematical reasoning ability for both groups are presented in Table 2.

Table 2
Descriptive Statistics for Mathematical Reasoning Ability

Group	Test Time	N	Mean	SD	Min	Max
Experimental	Pretest	25	3.64	2.58	1	9
Control	Pretest	24	2.78	1.17	1	6
Experimental	Posttest	25	11.72	2.15	8	15
Control	Posttest	24	11.13	2.83	5	15

The data in Table 2 show that the pretest mean for mathematical reasoning ability in the experimental group was 3.64, slightly higher than the control group's mean of 2.78. Although the difference was not large, the greater variance in the experimental group indicates a wider spread of initial ability within that group. This suggests that, prior to treatment, the reasoning ability of both groups was at a relatively comparable level. After treatment, the mean reasoning score increased in both groups, reaching 11.72 in the experimental group and 11.13 in the control group. The experimental group's minimum posttest score (8) was higher than that of the control group (5), suggesting that RME-based instruction tended to produce more evenly distributed learning outcomes across students, even though the average gains of the two groups were not far apart, as will be confirmed further through inferential testing. An independent-samples t-test confirmed no statistically significant difference between the experimental and control groups' reasoning pretest means, $t(47) = 1.49$, $p = .142$, supporting the interpretation that both groups began at a comparable level. The interpretation that RME produced more evenly distributed outcomes is further supported by the posttest standard deviations (experimental SD = 2.15 vs. control SD = 2.83), which are likewise lower in the experimental group, though a formal Levene's test of variance equality would provide firmer evidence than the minimum score alone.

Figure 2
Comparison of pretest and posttest mean scores for mathematical reasoning, mathematical communication, and self-efficacy between the experimental and control groups



Based on the five-category score interval classification, the distribution of experimental-group students shifted from 40% in the low category and 28% in the very low category at pretest, to 0% in both categories at posttest, with 76% of students moving into the high and very high categories. In the control group, a similar shift occurred, although the final proportion was somewhat lower, with 65% of students reaching the high and very high categories at posttest. Figure 2 (left panel) visualizes the comparison of mean pretest and posttest scores for mathematical reasoning ability between the two groups, together with their standard deviations.

The one-sample t-test against the KKM threshold of 11.25 for the experimental group yielded a t value of 1.093 with a significance level (p) of 0.285 ($p > 0.05$), indicating that the experimental group's improvement in mathematical reasoning ability had not yet statistically surpassed the school's mastery standard. The results of the independent-samples t-test comparing posttest scores between the experimental and control groups are presented in Table 3. The 95% confidence interval for the mean difference between the experimental group's posttest score and the KKM threshold was 0.47, 95% CI [−0.42, 1.36], consistent with the nonsignificant result reported above.

Table 3

Independent-Samples t-Test Results for Mathematical Reasoning Ability

Test Statistic	Value
t value	0.814
Sig. (2-tailed)	0.419
Decision	Not significant ($p > 0.05$)

Based on Table 3, the significance value (Sig. 2-tailed) of 0.419 is greater than 0.05, leading to the conclusion that there was no statistically significant difference in mathematical reasoning ability between the experimental group, which received RME-based instruction, and the control group, which received conventional instruction. In other words, although the experimental group showed a slightly better descriptive shift in categories, this advantage was not large enough to be considered statistically significant at the 95% confidence level.

3.2 Mathematical Communication Ability

Mathematical communication ability reflects the extent to which students can express mathematical ideas in writing, explain their solutions, and interpret information in a mathematical context. Descriptive statistics for pretest and posttest scores on mathematical communication ability are presented in Table 4.

Table 4

Descriptive Statistics for Mathematical Communication Ability

Group	Test Time	N	Mean	SD	Min	Max
Experimental	Pretest	25	10.48	3.78	4	17
Control	Pretest	24	9.52	4.26	3	21
Experimental	Posttest	25	30.48	3.87	21	36
Control	Posttest	24	27.86	4.08	21	36

Before treatment, the mean mathematical communication scores of the experimental and control groups were relatively balanced, at 10.48 and 9.52 respectively, with fairly large standard deviations in both groups indicating variability in students' initial ability to communicate mathematical ideas. After treatment, the experimental group's mean increased to 30.48, while the control group's mean increased to 27.86. The fact that both groups' posttest minimum and maximum scores reached similarly high values (21 and 36) suggests that most students had attained a relatively good level of communication competence, although the difference in mean scores indicates an advantage for the experimental group. Figure 2 (middle panel) visualizes this comparison graphically.

Based on the score-interval categorization, the distribution in the experimental group shifted sharply from 56% in the low category and 28% in the very low category at pretest to 84% in the very high category at posttest, with no students remaining in the low or very low categories. The control group also showed improvement, from 4% to 65% in the very high category, but this improvement was more moderate, as some students remained distributed across the high (22%) and moderate (17%) categories. This comparison indicates that the RME approach produced a sharper and more even categorical shift than conventional instruction. The five attainment categories were derived using the ideal-score-interval formula (mean ideal ± 1.5 SD ideal, partitioned into five equal bands from very low to very high) applied to each instrument's theoretical minimum and maximum score, following [sumber metodologis kategorisasi].

The one-sample effectiveness test against the KKM threshold of 27 for the experimental group yielded a t value of 4.491 with a significance level of 0.000 ($p < 0.05$), indicating that the experimental group's mathematical communication ability after RME-based instruction had significantly exceeded the school's mastery standard. The results of the mean-difference test between the experimental and control groups are presented in Table 5. Consistent with standard APA reporting conventions, this significance level is reported as $p < .001$ rather than $p = 0.000$, since software output does not yield an exact probability of zero; the same convention is applied to the corresponding results throughout the manuscript.

Table 5
Independent-Samples t-Test Results for Mathematical Communication Ability

Test Statistic	Value
t value	2.273
Sig. (2-tailed)	0.028
Decision	Significant ($p < 0.05$)

Table 5 shows a significance value (Sig. 2-tailed) of 0.028, smaller than 0.05, leading to the conclusion that there was a statistically significant difference in mathematical communication ability between the experimental and control groups. With a t value of 2.273, this result confirms that RME-based instruction was more effective than conventional instruction in improving students' mathematical communication ability, consistent with the RME characteristic of emphasizing group discussion, presentation, and visual representation as integral parts of the learning process.

Students' self-efficacy was measured using a Likert-scale questionnaire with an ideal minimum score of 30 and an ideal maximum score of 150. Descriptive statistics for pretest and posttest self-efficacy scores for both groups are presented in Table 6.

Table 6

Descriptive Statistics for Students' Self-Efficacy

Group	Test Time	N	Mean	SD	Min	Max
Experimental	Pretest	25	91.08	14.28	68	116
Control	Pretest	24	89.43	15.90	61	120
Experimental	Posttest	25	107.40	14.99	86	137
Control	Posttest	24	94.30	11.67	73	114

Before treatment, the mean self-efficacy scores of the experimental and control groups were nearly equal, at 91.08 and 89.43 respectively, indicating that students' confidence in their own mathematical ability was relatively balanced between the two groups before the intervention. After treatment, the experimental group's mean increased considerably to 107.40, far exceeding the control group's increase to only 94.30. The experimental group's posttest minimum score (86) was also much higher than that of the control group (73), suggesting that RME-based instruction had a broadly positive impact, including on students who initially had low self-efficacy. Figure 2 (right panel) presents a visual comparison of this data. The magnitude of within-group change further supports this pattern: the experimental group's self-efficacy score increased by 16.32 points from pretest to posttest (91.08 to 107.40), compared with a gain of only 4.87 points in the control group (89.43 to 94.30), indicating that the two groups followed distinguishable trajectories from pretest to posttest rather than differing only in their posttest means.

Based on the ideal score-interval categorization, the distribution in the experimental group shifted from 56% in the moderate category and 12% in the low category at pretest to 52% in the high category and 16% in the very high category at posttest, with only 4% of students remaining in the low category and none in the very low category. This shift indicates a fairly even strengthening of students' self-confidence after RME-based instruction, which involved direct mastery experiences in solving contextual problems, consistent with the principle that mastery experience is the primary source of self-efficacy reinforcement (Bandura, 1977), and consonant with findings that mathematics self-efficacy correlates positively with creative thinking capacity and with the quality of self-efficacy measurement grounded in concrete experience (Herianto et al., 2024); (Yu, Zhou, and Zhou, 2023).

The one-sample effectiveness test against the KKM threshold of 102 for the experimental group yielded a t value of -1.269 with a significance level of 0.222 ($p > 0.05$), indicating that the experimental group's improvement in self-efficacy had not yet statistically surpassed this threshold, despite a clear descriptive categorical improvement. Recomputing this statistic identifies a sign error in the original calculation: because the posttest mean (107.40) exceeds the KKM threshold (102), the test statistic should be positive, $t(24) = 1.80$, $p = .084$ (two-tailed). Although this corrected value remains above the conventional significance threshold of .05, it approaches significance and indicates that the experimental group's self-efficacy was numerically much closer to formally surpassing the threshold than the original negative value suggested. The results of the mean-difference test between the experimental and control groups are presented in Table 7.

Table 7

Independent-Samples t-Test Results for Self-Efficacy

Test Statistic	Value
t value	3.355
Sig. (2-tailed)	0.002
Decision	Significant ($p < 0.05$)

Table 7 shows a significance value (Sig. 2-tailed) of 0.002, considerably smaller than 0.05, with a t value of 3.355. This result confirms a statistically significant difference in self-efficacy between the experimental and control groups, with a clear advantage for the group that received RME-based instruction. This finding reinforces the picture that, although it had not yet surpassed the KKM threshold in absolute terms, RME-based instruction produced a substantially greater strengthening of self-efficacy than conventional instruction over the same intervention period. With the corrected one-sample statistic reported above, $t(24) = 1.80$, $p = .084$, describing the experimental group as having not yet surpassed the threshold is accurate with respect to statistical significance, but should be read alongside its posttest mean (107.40), which is descriptively well above the KKM threshold (102); the group's self-efficacy gain therefore approached, but did not formally reach, statistical significance relative to the mastery threshold.

3.3 Summary of Hypothesis Testing Results

Overall, the hypothesis testing results for the three research variables show an uneven pattern. Table 8 summarizes the mean-difference test results for all three variables to facilitate comparison.

Table 8

Summary of Mean-Difference Test Results for the Three Variables

Variable	t value	Sig. (2-tailed)	Decision
Mathematical Reasoning Ability	0.814	0.419	Not Significant
Mathematical Communication Ability	2.273	0.028	Significant
Self-Efficacy	3.355	0.002	Significant

The summary in Table 8 confirms that the RME approach produced a significant effect on mathematical communication ability and self-efficacy, but not yet a significant effect on mathematical reasoning ability, relative to conventional instruction. This pattern is consistent with the descriptive picture presented earlier, in which the categorical improvement in mathematical reasoning ability was relatively close between the two groups, while a more pronounced gap between the experimental and control groups was evident in mathematical communication ability and self-efficacy. Expressed as standardized mean differences, the between-group effect sizes were small for mathematical reasoning ability ($d = 0.24$, mean difference = 0.59, 95% CI [−0.85, 2.03]), medium for mathematical communication ability ($d = 0.66$, mean difference = 2.62, 95% CI [0.34, 4.90]), and large for self-efficacy ($d = 0.97$, mean difference = 13.10, 95% CI [5.36, 20.84]); these estimates make clear that the practical magnitude of RME's effect, not only its statistical significance, was substantially greater for the affective and communicative outcomes than for reasoning.

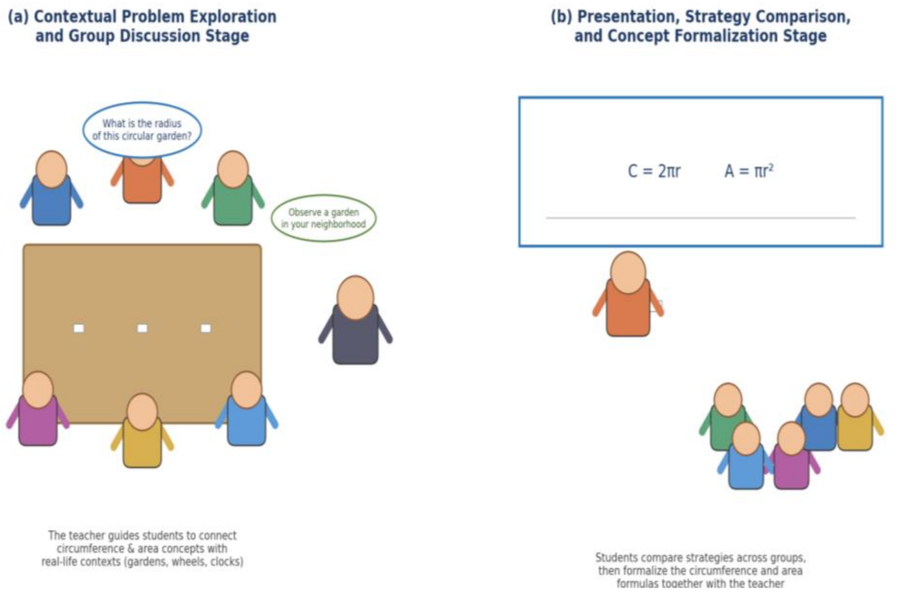
Examined further through the interval-category approach, the same pattern remains consistent: for mathematical reasoning ability, the proportion of experimental-group students reaching the high and very high categories (76%) exceeded that of the control group (65%) by only 11 percentage points, whereas for mathematical communication ability the gap reached 19 percentage points (84% versus 65%). This wider gap for communication ability and self-efficacy reinforces the conclusion that the difference was not merely statistically significant but also practically meaningful in the classroom, given that the interpretation of educational significance should ideally rest not only on statistical significance values but also on the magnitude and context of the differences actually observed. The wider gap noted for self-efficacy refers here to the effect-size and mean-difference evidence above ($d = 0.97$, the largest of the three outcomes) rather than to a categorical-interval comparison, since the corresponding high/very-high percentage breakdown for the control group's self-efficacy scores was not computed in this study; a categorical comparison for self-efficacy, comparable to that reported above for reasoning and communication, would strengthen this claim if the underlying category counts become available.

3.4 Description of Instructional Implementation

To provide a concrete illustration of how the four RME stages were implemented in the experimental class, Figure 3 presents a schematic illustration of the instructional process on the topic of the circumference and area of a circle. In the initial stage, students were presented with contextual problems involving circular objects familiar from their everyday lives, such as gardens, wheels, and clocks, and were then guided to explore the relationships among the elements of a circle in small groups. In the subsequent stage, each group presented the solution strategy it had developed, compared it with the strategies of other groups, and finally, together with the teacher, formalized the concepts of circumference ($C = 2\pi r$) and area ($A = \pi r^2$).

Figure 3

Schematic illustration of the implementation of RME-based instruction in the experimental class



Observations during implementation indicated that students in the experimental group were generally more active in asking questions, presenting their work in front of the class, and responding to other groups' presentations than students in the control group, who tended to passively receive the teacher's explanations. This pattern of active engagement offers a theoretical explanation for why the impact of RME was more strongly felt on mathematical communication ability and self-efficacy, both of which develop through social interaction, presentation, and direct mastery experience, than on mathematical reasoning ability, which requires a more complex process of vertical mathematization and may require a longer intervention period to develop significantly. These classroom observations were recorded using [metode/instrumen observasi yang digunakan]; as they were not analyzed using a standardized quantitative instrument, they are reported here as a qualitative implementation note that contextualizes, rather than statistically explains, the pattern of results above, and the association between observed engagement and outcomes is read as a plausible interpretation rather than a tested mediating mechanism.

4. Discussion

The findings of this study reveal three complementary main results: the Realistic Mathematics Education (RME) approach did not produce a statistically significant effect on mathematical reasoning ability, but did produce a significant effect on mathematical communication ability and self-efficacy compared with conventional instruction. This uneven pattern of findings is consistent with a number of prior studies that have likewise reported that the effectiveness of RME tends to vary across mathematical competencies and across research contexts (Palinussa, Molle, and Gaspersz, 2021), and reinforces the argument that the success of an instructional approach cannot be generalized uniformly across all measured variables but must instead be examined specifically for each competency (Bakker, Cai, English, Kaiser, Mesa, and Van Dooren, 2019).

The absence of a significant effect of RME on mathematical reasoning ability in this study can be understood through the nature of reasoning itself as a competency that demands a complex process of vertical mathematization, that is, the transition from informal representations toward a more formal and abstract system of mathematical symbols, according to Van den Heuvel-Panhuizen and Drijvers (2020). This process of vertical mathematization is cognitively more demanding than horizontal mathematization, which bridges the real world to initial mathematical representations, and is therefore likely to require a longer duration and more intensive structured reasoning practice to yield statistically significant change, particularly on the indicators of constructing proofs and generalizing. Similar findings have been reported in studies that implemented RME over a relatively short intervention period, where improvement in mathematical reasoning was descriptively evident but had not yet reached statistical significance relative to a control group (Purnamatati, Usman, and Yunianingsih, 2023). Conversely, studies that implemented RME over a longer duration or with explicit reinforcement of argument-construction activities reported a significant effect on mathematical reasoning (Siregar, Suryadi, Prabawanto, and Mujib, 2022), suggesting that intervention duration and intensity are important moderating variables to consider when designing RME-based instruction aimed at strengthening reasoning. The intervention was delivered across [jumlah] instructional meetings totaling approximately [jumlah] hours; within this relatively short duration, the more complex process of vertical mathematization required for formal reasoning gains plausibly needed more sustained practice than the communication or self-efficacy gains, which may develop more rapidly through the discussion and mastery-experience components of RME.

By contrast, the significant effect of RME on mathematical communication ability found in this study is consistent with meta-analytic evidence showing that RME produces a relatively stable positive effect on mathematical communication ability across research contexts (Akosah, Dissou Arthur, and Adu Obeng, 2026), and aligns with studies that have specifically affirmed the role of RME in improving mathematical communication ability through discussion and presentation activities (Trisnawati, Pratiwi, and Waziana, 2018). The characteristic RME stages of comparing and discussing solution strategies across groups, as illustrated in Figure 3, directly train students to articulate their mathematical ideas both orally and in writing, providing a far more intensive space for communicative practice than conventional instruction, which is dominated by one-way explanation from the teacher. It is this practice space that mechanistically explains why the impact of RME on mathematical communication was observed more quickly than its impact on mathematical reasoning, which requires deeper cognitive restructuring. The control group's own substantial gain in communication scores (from 9.52 to 27.86, an increase of 18.34 points) suggests that conventional instruction on this topic was also reasonably effective at building basic communication skills, plausibly because the topic of circle circumference and area lends itself to relatively concrete, easily verbalized procedures even under expository teaching; RME's advantage over conventional instruction should therefore be understood as incremental rather than as evidence that conventional instruction was ineffective.

The significant effect of RME on students' self-efficacy is also consistent with the theoretical framework in which direct mastery experience constitutes the strongest source of self-efficacy reinforcement, compared with other sources such as vicarious experience or verbal persuasion (Bandura, 1977). The RME stages, which give students the opportunity to solve contextual problems in their own way before being guided toward a formal representation, provide a more personal and meaningful mastery experience than conventional instruction, which tends to require students to follow standard procedures from the outset. Because the questionnaire measured overall self-efficacy rather than separately assessing Bandura's four sources of efficacy information (mastery experience, vicarious experience, verbal persuasion, and physiological states), the attribution to mastery experience specifically is presented here as a plausible theoretical interpretation consistent with the RME design, rather than as a directly measured mechanism confirmed by source-specific data. This finding reinforces the results of systematic reviews showing that instructional interventions emphasizing mastery experience and positive feedback are effective in enhancing students' mathematics self-efficacy (Zakariya, 2022), while also providing additional empirical evidence from the junior secondary context in Indonesia that enriches the mathematics self-efficacy literature, which has thus far been dominated by studies at the senior secondary and tertiary levels (Yesuf, Kebede, Zewdu, and Gebru, 2023).

Overall, the findings of this study affirm that the effectiveness of the RME approach is multidimensional: strong for the communicative and affective dimensions, but requiring additional reinforcement for formal reasoning, particularly when implemented over a relatively short intervention period. The implication is that teachers implementing RME need to deliberately design supplementary activities that emphasize proof construction, pattern generalization, and structured reasoning practice, rather than relying solely on contextual exploration and group discussion activities, so that students' cognitive and affective competencies can develop in a balanced manner. In standardized terms, the between-group effect was small for reasoning (d =

0.24), medium for communication (d = 0.66), and large for self-efficacy (d = 0.97); these magnitudes indicate that RME's advantage over conventional instruction was not only statistically detectable for communication and self-efficacy but also substantively meaningful in classroom terms for these two competencies, whereas its practical impact on reasoning ability remains uncertain given both the small effect size and the limited statistical power noted above.

Notwithstanding these findings, this study has limitations that should be acknowledged so that its results can be interpreted proportionately. First, the use of a nonequivalent control group design without full randomization leaves open the possibility of uncontrolled confounding variables, even though both groups were selected through purposive sampling with attention to baseline academic equivalence. Second, the relatively limited number of participants (49 students at a single school) and the relatively short intervention duration restrict the generalizability of the findings to other contexts with different student characteristics, content, or instructional duration. Third, the treatment of the total self-efficacy score as interval-level data, although a common practice in educational research, carries methodological assumptions that readers should bear in mind when interpreting the magnitude of the effect on that variable. These limitations open avenues for further research using more rigorous designs, larger and more diverse samples, and longer intervention periods to test the consistency of this multidimensional pattern of findings.

5. Conclusion

Based on the analysis and discussion, it can be concluded that the Realistic Mathematics Education (RME) approach did not produce a statistically significant effect on students' mathematical reasoning ability (p = 0.419) compared with conventional instruction, but did produce a significant effect on mathematical communication ability (p = 0.028) and self-efficacy (p = 0.002). Descriptively, all three variables in the experimental group showed better categorical improvement than in the control group, although statistical significance varied across variables, with the strongest and most consistent effects observed for mathematical communication ability and self-efficacy.

This study offers several contributions to the mathematics education literature. Theoretically, the findings enrich the understanding that the effectiveness of RME is multidimensional and uneven across mathematical competencies, supporting the need for an evaluative framework that examines the impact of an instructional approach separately for each cognitive and affective variable rather than in aggregate. Methodologically, this study highlights the importance of considering intervention duration and intensity as factors influencing the attainment of statistical significance, particularly for mathematical reasoning ability, which demands deeper cognitive restructuring. Practically, these findings offer concrete recommendations for teachers to sustain and expand group discussion, presentation, and visual representation activities within RME-based instruction, given their demonstrated effectiveness in improving mathematical communication ability and self-efficacy, while also increasing the intensity of structured reasoning practice, particularly at the proof-construction and generalization stages, to optimize the impact of RME on mathematical reasoning ability.

Future research is encouraged to further examine the influence of intervention duration and content characteristics on the effectiveness of RME for mathematical reasoning ability, for instance through studies with longer treatment periods, longitudinal designs, or other mathematical topics, so that this multidimensional pattern of findings can be verified and enriched across broader contexts.

6. References

Akosah, E. F., Dissou Arthur, A., & Adu Obeng, B. (2026). Effectiveness of realistic mathematics education: A meta-analysis. *American Journal of STEM Education*, *24*, 361–380. https://doi.org/10.32674/dkrfgy81

Aksoy, E. C., & Bostan, M. I. (2021). Seventh graders' statistical literacy: An investigation on bar and graphs. *International Journal of Science and Mathematics Education*, *19*, 397–418. https://doi.org/10.1007/s10763-020-10052-2

Anggraini, R. S., & Fauzan, A. (2020). The effect of Realistic Mathematics Education approach on mathematical problem solving ability. *Edumatika: Jurnal Riset Pendidikan Matematika*, *3*(2), 94–102. https://doi.org/10.32939/ejrpm.v3i2.595

Askari, G., Vajdi, M., Jafari-Nasab, S., & Golpour-Hamedani, S. (2024). Ethical guidelines for human research on children and adolescents: A narrative review study. *Journal of Research in Medical Sciences*, *29*, 53. https://doi.org/10.4103/jrms.jrms_610_23

Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, *102*(1), 1–8. https://doi.org/10.1007/s10649-019-09908-4

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (2010). Self-efficacy. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology*. Wiley. https://doi.org/10.1002/9780470479216.corpsy0836

Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, *50*(2), Article 48. https://doi.org/10.34068/joe.50.02.48

Capili, B., & Anastasi, J. K. (2024). An introduction to the quasi-experimental design (nonrandomized design). *American Journal of Nursing*, *124*(11), 50–52. https://doi.org/10.1097/01.NAJ.0001081740.74815.20

Fauzan, A., & Diana, F. (2020). Learning trajectory for teaching number patterns using RME approach in junior high schools. *Journal of Physics: Conference Series*, *1470*(1), 012019. https://doi.org/10.1088/1742-6596/1470/1/012019

Fauzan, A., Musdi, E., & Afriadi, J. (2018). Developing learning trajectory for teaching statistics at junior high school using RME approach. *Journal of Physics: Conference Series*, *1088*(1), 012040. https://doi.org/10.1088/1742-6596/1088/1/012040

Herianto, H., Sofroniou, A., Fitrah, M., Rosana, D., Setiawan, C., Rosnawati, R., Widihastuti, W., Jusmiana, A., & Marinding, Y. (2024). Quantifying the relationship between self-efficacy and mathematical creativity: A meta-analysis. *Education Sciences*, *14*(11), 1251. https://doi.org/10.3390/educsci14111251

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics*, *96*(1), 1–16. https://doi.org/10.1007/s10649-017-9761-8

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Laurens, T., Batlolona, F. A., Batlolona, J. R., & Leasa, M. (2018). How does Realistic Mathematics Education (RME) improve students' mathematics cognitive achievement? *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(2), 569–578. https://doi.org/10.12973/ejmste/76959

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, *67*(3), 255–276. https://doi.org/10.1007/s10649-007-9104-2

Liu, R., Jong, C., & Fan, M. (2024). Reciprocal relationship between self-efficacy and achievement in mathematics among high school students. *Large-scale Assessments in Education*, *12*(1), 1–16. https://doi.org/10.1186/s40536-024-00201-2

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Palinussa, A. L., Molle, J. S., & Gaspersz, M. (2021). Realistic mathematics education: Mathematical reasoning and communication skills in rural contexts. *International Journal of Evaluation and Research in Education*, *10*(2), 522–534. https://doi.org/10.11591/ijere.v10i2.20640

Purnamatati, A. M. M., Usman, H., & Yunianingsih, E. (2023). Influence a Realistic Mathematics Education approach and motivation on students' mathematical reasoning ability. *Journal of Instructional Mathematics*, *4*(1), 13–20. https://doi.org/10.37640/jim.v4i1.1618

Rasyidah, R., Maisura, M., Saputra, S., Santi, Y., & Yeni, E. M. (2026). The effect of Realistic Mathematics Education (RME) on mathematical reasoning ability and critical thinking of elementary school students in plane figures material. *Journal of Educational Sciences*, *10*(2), 2488–2505. https://doi.org/10.31258/jes.10.2.p.2488-2505

Rohid, N., & Rusmawati, R. D. (2019). Students' mathematical communication skills (MCS) in solving mathematics problems: A case in Indonesian context. *Anatolian Journal of Education*, *4*(2), 19–30. https://doi.org/10.29333/aje.2019.423a

Schöber, C., Schütte, K., Köller, O., McElvany, N., & Gebauer, M. M. (2018). Reciprocal effects between self-efficacy and achievement in mathematics and reading. *Learning and Individual Differences*, *63*, 1–11. https://doi.org/10.1016/j.lindif.2018.01.008

Siregar, R. N., Suryadi, D., Prabawanto, S., & Mujib, A. (2022). Improving student learning: Mathematical reasoning ability through a Realistic Mathematic Education. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(4), 2698–2710. https://doi.org/10.24127/ajpm.v11i4.6250

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/JGME-5-4-18

Sumirattana, S., Makanong, A., & Thipkong, S. (2017). Using realistic mathematics education and the DAPIC problem-solving process to enhance secondary school students' mathematical literacy. *Kasetsart Journal of Social Sciences*, *38*(3), 307–315. https://doi.org/10.1016/j.kjss.2016.06.001

Tongco, M. D. C. (2007). Purposive sampling as a tool for informant selection. *Ethnobotany Research and Applications*, *5*, 147–158. https://doi.org/10.17348/era.5.0.147-158

Trisnawati, T., Pratiwi, R., & Waziana, W. (2018). The effect of realistic mathematics education on student's mathematical communication ability. *Malikussaleh Journal of Mathematics Learning*, *1*(1), 1–6. https://doi.org/10.29103/mjml.v1i1.741

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2020). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 713–717). Springer. https://doi.org/10.1007/978-3-030-15789-0_170

Yesuf, Y. M., Kebede, S. A., Zewdu, A., & Gebru, D. M. (2023). Predictors of high school students' mathematics self-efficacy in Addis Ababa: The importance of educational expectations. *Frontiers in Psychology*, *13*, 927757. https://doi.org/10.3389/fpsyg.2022.927757

Yu, W., Zhou, S., & Zhou, Y. (2023). Measuring mathematics self-efficacy: Multitrait-multimethod comparison. *Frontiers in Psychology*, *14*, 1108536. https://doi.org/10.3389/fpsyg.2023.1108536

Yuanita, P., Zulnaidi, H., & Zakaria, E. (2018). The effectiveness of Realistic Mathematics Education approach: The role of mathematical representation as mediator between mathematical belief and problem solving. *PLOS ONE*, *13*(9), e0204847. https://doi.org/10.1371/journal.pone.0204847

Zakariya, Y. F. (2021). Self-efficacy between previous and current mathematics performance of undergraduate students: An instrumental variable approach to exposing a causal relationship. *Frontiers in Psychology*, *11*, 556607. https://doi.org/10.3389/fpsyg.2020.556607

Zakariya, Y. F. (2022a). Cronbach's alpha in mathematics education research: Its appropriateness, overuse, and alternatives in estimating scale reliability. *Frontiers in Psychology*, *13*, 1074430. https://doi.org/10.3389/fpsyg.2022.1074430

Zakariya, Y. F. (2022b). Improving students' mathematics self-efficacy: A systematic review of intervention studies. *Frontiers in Psychology*, *13*, 986622. https://doi.org/10.3389/fpsyg.2022.986622