Daily Retrieval Practice and Grade 8 Achievement in Polynomials and Rational Expressions

An Action Research

[Copy for blind review: the researchers' names, positions, and school are removed. The final copy lists them here.]

Conducted June to August 2026

Abstract

This action research tested whether five minutes of daily retrieval practice at the start of each Mathematics period would raise Grade 8 learners' achievement in special products, factoring, and rational algebraic expressions. Two intact sections with the same teacher took part in Term 1 of School Year 2026–2027. A coin toss gave one section daily recall quizzes and the other an equally long teacher-led review. The study used a quasi-experimental pretest–posttest design with a nonequivalent comparison group and a focus group discussion. Of 83 enrolled learners, 75 with parental consent and both tests were analyzed (38 and 37). A 30-item test, validated by eight experts (content validity index = .94) and piloted with 40 Grade 9 learners (KR-20 = .87), was given before and after the six weeks. Gains were compared with Welch's t test and checked with an analysis of covariance. Reaching a 75% target was compared with a chi-square test, and the focus group was analyzed thematically. Both sections started level (9.18 and 9.14 of 30). The retrieval practice section gained 9.92 points and the comparison section 6.86, $t(72.69) = 2.54$, $p = .013$, Hedges' $g = 0.58$, 95% CI [0.12, 1.04]. The share reaching the target did not differ significantly (28.9% and 13.5%), $x^2(1, N = 75) = 2.66$, $p = .103$. Learners said the quizzes revealed what they had forgotten, without the fear of a grade. Daily retrieval practice is a low-cost routine worth extending, with a delayed test to check retention.

Keywords: retrieval practice, low-stakes quizzing, mathematics achievement, quasi-experimental design, action research

Acknowledgment

The researchers thank the school principal, who approved the study, fitted the pilot test into the second week of lessons, and lent the school's printer; the Schools Division Research Committee, whose reviewers asked for the comparison section's review time to be made equal to the quizzes' five minutes; the eight members of the validation panel, who rated all 36 draft items and explained each rating they withheld; the Grade 9 adviser and learners who took the pilot test; and the parents and learners of the two Grade 8 sections. The study was funded by the Basic Education Research Fund (BERF) for 2026.

Daily Retrieval Practice and Grade 8 Achievement in Polynomials and Rational Expressions

Context and Rationale

Algebra in Grade 8 builds on itself. A learner who cannot recognize the square of a binomial in week 3 cannot factor a perfect square trinomial in week 5, and a learner who cannot factor cannot simplify a rational algebraic expression in week 7. The Grade 8 Mathematics budget of work places these competencies in weeks 3 to 8 of Term 1, one after another.

In our classes, learners forget the earlier lessons while the next ones are taught. In School Year 2025–2026, our Grade 8 learners answered 41% of the items on special products and factoring correctly in the quarter's second summative test, although most had answered similar items correctly in class a few weeks earlier. When asked to factor $x^2 + 6x + 9$ during a lesson on rational expressions, most of a class could not say where to begin. Reteaching took days we did not have.

Research on memory offers a simple remedy. Recalling information from memory strengthens it more than reading it again, an effect known as the testing effect or retrieval practice (Roediger & Karpicke, 2006). In classrooms, short low-stakes quizzes improve later test scores. A meta-analysis of 222 classroom studies with 48,478 students found a medium effect, $g = 0.499$ (Yang et al., 2021), and a review of 50 classroom experiments found medium or large benefits in 57% of the effects measured (Agarwal et al., 2021). In Cebu City, Grade 7 learners who took daily practice tests on fractions for eight weeks gained more than learners given conventional review (Alenton, 2022).

Reviews of study techniques rate practice testing among the most useful ones for learners (Dunlosky et al., 2013), and beginning a lesson with a short review of previous learning is the first of Rosenshine's (2012) principles of instruction.

DepEd's assessment policy leaves room for such quizzes. Formative assessment "shall be used to monitor learning progress but shall not be used for grade computation" (Department of Education [DepEd], 2026b, para. 15), and teachers are to build in self-assessment and progress tracking (para. 17). A recall quiz that learners check themselves and that never enters the class record fits both.

Most of the evidence above comes from classes unlike ours: only 6% of the classroom experiments Agarwal et al. (2021) reviewed were conducted outside Western, educated, industrialized, rich, and democratic countries. Whether five minutes of recall a day helps in a Philippine public school, with about 40 learners in a section and quizzes on quarter sheets of paper, is a question for our own classrooms. This action research tested whether daily retrieval practice would raise our Grade 8 learners' achievement in polynomials and rational expressions, and how the learners experienced it. It compared a section that answered the daily recall quizzes with a section that received an equally long teacher-led review. The study addresses the Teaching and Learning theme of the Basic Education Research Agenda (DepEd, 2016).

Innovation, Intervention, and Strategy

The innovation, which we called Five-Minute Recall, replaced the usual review at the start of the Mathematics period with three questions that

learners answered from memory. Its design follows two findings from the research above: recalling is stronger practice than rereading, and recall spread over days helps more than recall crowded into one review.

The Intervention

Every Mathematics period of the retrieval practice section began with the routine in Table 1. Learners answered on a quarter sheet of paper, without notes, and checked their own answers. The teacher did not collect, score, or record the sheets, so the quizzes counted toward no grade.

Table 1

The Five-Minute Recall Routine

Step	What happens	Time
1. Questions	The teacher shows three questions on the board: one from the previous lesson, one from the previous week, and one from an earlier lesson of the unit.	0:30
2. Recall	Learners answer on a quarter sheet without notes or help.	3:00
3. Check	The teacher shows the answers and one worked solution; learners mark their own sheets.	1:00
4. Note	Learners circle an item they missed and write what to ask during the lesson.	0:30

Note. The teacher wrote 84 recall questions for the 28 periods held. None was taken from the unit test.

The comparison section received the same lessons, worksheets, homework, and summative tests from the same teacher. Its periods began with five minutes of teacher-led review: the teacher summarized the previous lesson on the board and answered learners' questions. The two sections therefore spent the same time on review; the difference lay in who did the remembering. After the posttest, the comparison section began the same recall routine for the rest of the term.

Strategy and Evaluation Plan

The routine ran for six weeks, from June 29 to August 7, 2026 (weeks 3 to 8 of Term 1 in the budget of work, counted from the start of the Instructional Block on June 15), with a pretest on June 26 and a posttest on August 10. A second researcher observed one period of each section every week with a five-item fidelity checklist: the review began within two minutes of the bell, it lasted about five minutes, it had three questions spread across earlier lessons (retrieval section only), learners worked without notes (retrieval section only), and no score was recorded. We decided in advance to count the routine successful if the retrieval practice section's mean gain was significantly higher than the comparison section's at the .05 level, and if learners described it as helpful in the focus group.

Action Research Questions

The study answered four questions:

1. What were the pretest and posttest scores of the learners in the retrieval practice section and in the comparison section?
2. Is there a significant difference between the mean gains of the two sections from the pretest to the posttest?
3. Is reaching the 75% target on the posttest (at least 23 of 30 items) associated with section?
4. How did the learners in the retrieval practice section describe the daily recall quizzes?

Two null hypotheses were tested at the .05 level of significance:

H_{01}: There is no significant difference between the mean gains of the retrieval practice section and the comparison section.

H_{02}: There is no significant association between section and reaching the 75% target on the posttest.

Action Research Methods

Research Design

The study used a quasi-experimental pretest–posttest design with a nonequivalent comparison group (Shadish et al., 2002), followed by a focus group discussion to explain the results, an explanatory sequential mixed methods design (Creswell & Plano Clark, 2018). The learners could not be assigned to conditions one by one, because a class is taught as a section. The two Grade 8 sections of the researcher-teacher were assigned to conditions by a coin toss on June 24, 2026, witnessed by the school research coordinator. With only two sections, a coin toss decides which section gets the routine but cannot make the sections equal; the pretest was used to check how similar they were, and the analysis of covariance to adjust for the difference that remained.

Participants and Other Sources of Data and Information

The participants were the learners of two Grade 8 sections of a public secondary school, both taught Mathematics by the first researcher in the morning, at 7:30 a.m. (retrieval practice section) and 9:15 a.m. (comparison section). All 83 enrolled learners were invited, so no sampling was done. Table 2 shows how many took part.

Table 2

Learners Enrolled, Excluded, and Analyzed, by Section

Learners	Retrieval practice section	Comparison section	Total
Enrolled in the two sections	42	41	83
No parental consent	1	2	3
Transferred to another school in July	1	1	2
Enrolled after the pretest	0	1	1
Absent from the posttest and its make-up	2	0	2
Analyzed	38	37	75

Note. Learners who were not analyzed took part in all lessons and reviews; only their scores were left out.

The analyzed learners were 19 girls and 19 boys in the retrieval practice section and 19 girls and 18 boys in the comparison section.

Sample Size

Because whole sections took part, the number of learners was fixed. A sensitivity power analysis in G*Power (Faul et al., 2007) showed that 40 learners per section, the size planned, give a power of .80 to detect a difference of d = 0.63 or larger at α = .05, two-tailed; with the 38 and 37 learners analyzed, the smallest such difference is d = 0.66. A medium effect of d = 0.50 would need 64 learners per section (Cohen, 1992), so a smaller real effect could go undetected in this study.

Other Sources

Eight experts rated the draft test: the school's Mathematics department head, two Master Teachers in Mathematics of the school who were not part of the research team, the division's Education Program Supervisor in Mathematics, two Grade 8 Mathematics teachers of other schools, a Master Teacher in Mathematics of another school, and a mathematics education

faculty member of a state university. Forty Grade 9 learners of the same school, who had studied the topics in Grade 8, took the pilot test. For the focus group, we chose eight learners of the retrieval practice section by purposive sampling: three whose gains were in the top third of the section, three from the middle third, and two from the bottom third, four girls and four boys. One focus group was held to help explain the test results, not to reach saturation, which studies of similar participants reach within four to eight focus groups (Hennink & Kaiser, 2022).

Data Gathering Methods

The Unit Test

The researchers wrote a 30-item multiple-choice test, each item with four options, from a table of specifications (Table 3). The number of items for each competency equals the number of class days planned for it, so the test weighs each competency as instruction did.

Table 3

Table of Specifications of the Unit Test

Competency (Grade 8 Mathematics, Term 1)	Days	Items	Item numbers
Multiply simple monomials and binomials with simple binomials and multinomials	4	4	1–4
Use special product patterns to multiply binomials	5	5	5–9
Completely factor different types of polynomials	7	7	10–16
Solve problems involving special products and factors of polynomials	4	4	17–20
Simplify rational algebraic expressions	4	4	21–24
Perform operations on rational algebraic expressions	4	4	25–28
Solve problems involving simple rational algebraic equations	2	2	29–30

Competency (Grade 8 Mathematics, Term 1)	Days	Items	Item numbers
Total	30	30	

Note. Competencies shortened from the Grade 8 Mathematics budget of work, weeks 3 to 8 of Term 1. The factoring competency lists polynomials with a common monomial factor, the difference of two squares, and quadratic trinomials, including perfect square trinomials.

Content validity was checked with Lawshe's (1975) content validity ratio (CVR). Each of the eight experts rated every one of 36 draft items as "essential," "useful but not essential," or "not necessary" to the competency it was written for. For each item, CVR = $(n_e - N/2) / (N/2)$, where n_e is the number of experts who rated the item essential and N is the number of experts. With eight experts the critical value is .75 in both Lawshe's (1975) table and Ayre and Scally's (2014) exact recalculation, so an item was kept only if at least seven experts rated it essential. Twenty-five items were rated essential by all eight experts (CVR = 1.00) and seven by seven experts (CVR = .75). Four were dropped: one tested a law of exponents rather than multiplying polynomials (CVR = .00), one repeated another item with different numbers (CVR = .25), one was a long word problem in which reading decided the answer (CVR = .50), and one required factoring a sum of two cubes, which the competencies do not include (CVR = −.25). The content validity index, the mean CVR of the 30 items in the final test, is .94.

The 32 remaining items were piloted on June 22, 2026, with 40 Grade 9 learners. For each item we computed the difficulty index p, the proportion of learners who answered correctly, and the discrimination index D, the proportion correct among the 11 highest scorers (the top 27%) minus the proportion correct among the 11 lowest (ties at the cutoffs broken by the order of the list). Items with D below .20 were to be revised or removed (Ebel &

Frisbie, 1991). Two were removed. One asked for "a factorization" of $2x^2 - 18$ and listed both $2(x^2 - 9)$ and $2(x + 3)(x - 3)$, so it had two defensible answers; strong and weak learners split between them ($p = .50$, $D = .09$). All 40 learners answered the other correctly ($p = 1.00$, $D = .00$). The 30 items kept had difficulty indices from .30 to .88 and discrimination indices from .27 to .73. Their reliability in the pilot, by the Kuder–Richardson Formula 20 (KR-20; Kuder & Richardson, 1937), was .87; in this study it was .81 on the posttest.

The same 30 items, in a different order, were given as the pretest on June 26 and the posttest on August 10, 2026, in one 45-minute period. Learners absent from the posttest took it on August 12. The first researcher scored the answer sheets with a key, and the second researcher rechecked every sheet; three scoring errors were corrected.

Implementation Records

The first researcher kept a daily log of the periods held and each learner's attendance in both sections. The second researcher observed six periods of each section with the fidelity checklist described above.

Focus Group Discussion

The focus group met on August 13, 2026, after classes, for 45 minutes. The third researcher, who did not teach the learners, facilitated it in Filipino and English with a guide of five questions (Appendix C), so that learners could speak freely about their teacher's routine. It was audio-recorded with the permission of the learners and their parents, transcribed word for word, and translated into English where learners spoke Filipino.

Setting and Data Storage

The tests and the focus group took place in the learners' classrooms. Each learner was given a code (L01 to L83) after the class lists were shuffled, and only the codes appear in the data set and this report. The key linking codes to names was kept by the first researcher in a password-protected file, apart from the data set. Answer sheets and consent forms were kept in a locked cabinet at the school. The audio recording was deleted after the transcript was checked against it in September 2026.

Ethical Considerations

The Schools Division Research Committee approved the proposal before data collection; its reference number is withheld for blind review. Parents received a letter in English and Filipino that explained the purpose, the procedures, the risks and benefits, and the right to withdraw, and signed a consent form; learners gave their assent after the same explanation in class. Taking part was voluntary, and a learner who declined lost nothing: every learner received the same lessons, reviews, and tests, and only the scores of consenting learners were analyzed. The risks were minimal: the time of two tests and, for eight learners, a 45-minute focus group. Because the comparison section could benefit from the routine too, it received the recall quizzes after the posttest. The researchers declared their use of an artificial intelligence (AI) tool in Appendix D (DepEd, 2026a).

Data Analysis

Scores were summarized with the mean (*M*), standard deviation (*SD*), median, range, and mean percentage score (MPS). For Question 2, each

learner's gain (posttest minus pretest) was computed, and the normality of the gains in each section was checked with the Shapiro–Wilk test (Shapiro & Wilk, 1965). The mean gains were compared with Welch's *t* test, which does not assume equal variances (Delacre et al., 2017), and the effect size was reported as Hedges' *g* with its 95% confidence interval (Lakens, 2013). Because the sections were not randomly formed, the result was checked with an analysis of covariance (ANCOVA) of the posttest scores with the pretest as covariate (Dimitrov & Rumrill, 2003), and with the Mann–Whitney *U* test. For Question 1, the relationship between pretest and posttest scores was measured with Spearman's rank correlation, because the pretest scores were skewed. For Question 3, a chi-square test of independence compared the number of learners in each section who reached 23 of 30 items, with phi (φ) as the effect size. The *t* test, the chi-square test, and the ANCOVA (as a multiple regression) were run in Microsoft Excel's Analysis ToolPak and its built-in functions; the Shapiro–Wilk tests were run in the free program jamovi (The jamovi project, 2026) under Exploration, Descriptives, split by section, and the Mann–Whitney tests in its Independent Samples T-Test, which reports the smaller of the two *U* values. For Question 4, two researchers coded the transcript separately and agreed on themes through the six phases of thematic analysis (Braun & Clarke, 2006).

Discussion of Results and Reflection

Implementation

Twenty-eight of the 30 planned periods were held; two days were lost to class suspensions for heavy rain in July. The learners analyzed attended a

mean of 25.8 periods in the retrieval practice section (range 23–28) and 25.9 in the comparison section (range 23–28). Eleven of the 12 observed periods met every checklist item that applied to their section; in the other, the recall took nine minutes because the teacher explained a missed item at length.

Question 1: Scores Before and After

Both sections began the unit at about the same level and both improved, the retrieval practice section more (Table 4 and Figure 1).

Table 4

Pretest, Posttest, and Gain Scores by Section

Score	Section	*n*	*M*	*SD*	*Mdn*	Range	MPS
Pretest	Retrieval practice	38	9.18	3.35	9.0	3–21	30.6
Pretest	Comparison	37	9.14	4.17	9.0	4–22	30.5
Posttest	Retrieval practice	38	19.11	4.91	19.0	9–28	63.7
Posttest	Comparison	37	16.00	5.54	15.0	5–26	53.3
Gain	Retrieval practice	38	9.92	5.11	10.5	1 to 19	33.1
Gain	Comparison	37	6.86	5.31	7.0	−5 to 18	22.9

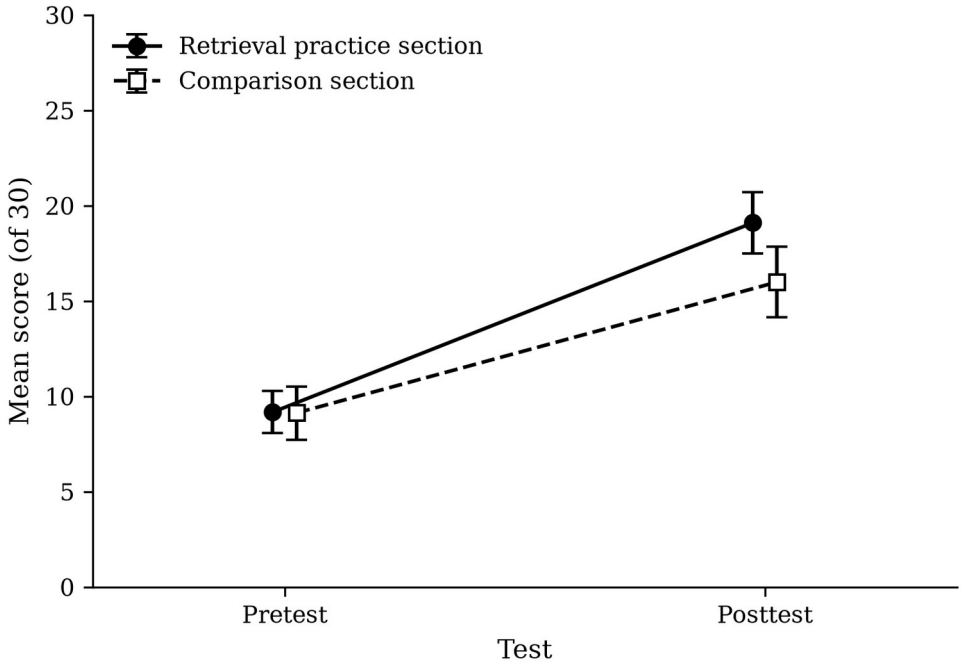
Note. Scores are numbers of correct answers out of 30. MPS = mean percentage score; for gains, the mean gain as a percentage of 30.

The pretest means differed by only 0.05 points, $t(68.97) = 0.06$, $p = .955$, 95% CI [−1.70, 1.79]. Pretest scores were skewed in both sections, as scores near the guessing floor usually are (Shapiro–Wilk $p = .026$ and $p < .001$), and the Mann–Whitney test gave the same result, $U = 636.5$, $p = .482$. A nonsignificant pretest difference does not prove the sections equal, but the interval suggests that any difference at the start was less than two items. Both pretest means were less than two items above 7.5, the score expected from guessing alone on 30 four-option items, as is usual before a unit is taught. For the same reason the pretest was less reliable (KR-20 = .59) than the

posttest. Pretest and posttest scores were moderately related, $r_s(73) = .31$, $p = .007$, measured with Spearman's rho because the pretest was skewed.

Figure 1

Mean Scores on the Unit Test Before and After the Intervention, by Section



Note. Error bars show 95% confidence intervals of the means.

Question 2: Difference in Gains

The gains were close to normally distributed in both sections, Shapiro–Wilk $W = .96$, $p = .174$ (retrieval practice) and $W = .99$, $p = .951$ (comparison), so the t test was appropriate. Table 5 gives Excel's output for Welch's t test.

Table 5

Welch's t Test of the Gains: Excel's Analysis ToolPak Output

	Retrieval practice	Comparison
Mean	9.9211	6.8649
Variance	26.1287	28.2312
Observations	38	37
Hypothesized Mean Difference	0	

	Retrieval practice	Comparison
df	73	
t Stat	2.5375	
P(T<=t) one-tail	0.0066	
t Critical one-tail	1.6660	
P(T<=t) two-tail	0.0133	
t Critical two-tail	1.9930	

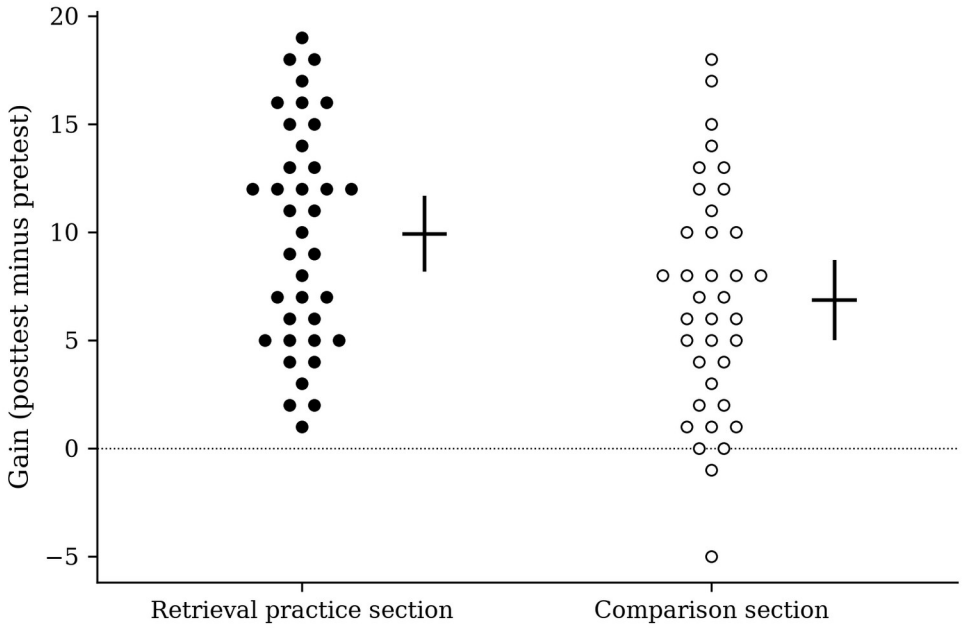
Note. Output of "t-Test: Two-Sample Assuming Unequal Variances," rounded to four decimals. The ToolPak rounds the degrees of freedom to the nearest whole number; the exact value, used in the text, is 72.69.

The retrieval practice section gained 3.06 points more than the comparison section, $t(72.69) = 2.54$, $p = .013$, 95% CI [0.66, 5.46], Hedges' $g = 0.58$, 95% CI [0.12, 1.04]. H_{01} is rejected. The effect is medium by Cohen's (1992) guide, but its interval is wide: the data are consistent with an effect as small as 0.12 or as large as 1.04.

The two checks agree. Adjusted for pretest scores, the posttest means were 19.09 for the retrieval practice section and 16.01 for the comparison section, $F(1, 72) = 7.41$, $p = .008$, partial $\eta^2 = .09$; the slopes of the posttest on the pretest did not differ significantly between the sections (test of homogeneity of slopes, $F(1, 71) = 0.27$, $p = .608$), so the ANCOVA's assumption of parallel slopes was not contradicted. The Mann–Whitney test, which does not assume normality, gave the same conclusion, $U = 486$, $p = .022$. Figure 2 shows that the gains of the two sections overlap a great deal: the routine moved the section's center, not every learner.

Figure 2

Gains From Pretest to Posttest, by Section



Note. Each dot is one learner. Bars show the mean and its 95% confidence interval.

Both sections improved significantly within themselves, retrieval practice $t(37) = 11.96$, $p < .001$, $dz = 1.94$, and comparison $t(36) = 7.86$, $p < .001$, $dz = 1.29$, because both were taught the unit. These paired tests show that learners learned; only the comparison between sections shows what the routine added.

Question 3: Reaching the 75% Target

Twice the share of the retrieval practice section answered at least 23 of the 30 items correctly (Table 6). The association was not significant, $x^2(1, N = 75) = 2.66$, $p = .103$, $\varphi = .19$, so H_{02} is not rejected. Every expected frequency was above 5 (the smallest was 7.89), so the chi-square test was appropriate.

Table 6

Learners Reaching the 75% Target on the Posttest, by Section

Section	Reached the target	Did not reach it	Total
Retrieval practice	11 (28.9%)	27 (71.1%)	38
Comparison	5 (13.5%)	32 (86.5%)	37
Total	16 (21.3%)	59 (78.7%)	75

Note. The target is 23 of 30 items (75%).

A nonsignificant result does not show that the routine made no difference to reaching the target. If the routine made no difference, a gap this large would still arise by chance about one time in ten among 75 learners. Sorting learners into "reached" and "did not reach" also discards most of the information in their scores: a learner who rose from 9 to 22 counts the same as one who stayed at 9. Most of the retrieval practice section's added gain fell below the cutoff, which is why the mean gains differed significantly and the proportions did not.

Question 4: The Learners' Experience

Four themes emerged from the focus group (Table 7). Participants are identified by number, with the third of the section their gain fell in.

Table 7

Themes From the Focus Group Discussion (n = 8)

Theme	Participants	Sample response
Finding out what I had forgotten	7	"I thought I already knew it, but when the question came, I went blank. That's when I found out I still had to study it." (P2, top third)
Mistakes without a grade	6	"It's not graded, so I don't get nervous even when I'm wrong." (P6, middle third)

Theme	Participants	Sample response
Old lessons came back	5	"On the posttest there was a perfect square trinomial, and I saw the pattern right away because we answered one like it in the recall many times." (P1, top third)
Hard to keep up after an absence	3	"When I was absent the day before, I couldn't answer the first question." (P7, bottom third)

Note. Responses in Filipino are translated into English.

The first theme was the most common. Learners described the quiz as the moment they learned what they had not yet learned: "Akala ko alam ko na, pero pagdating ng tanong, blangko ako" (P2). Four said they asked about a missed item during the lesson that followed, which the routine's fourth step invited.

The second theme explains why learners took the quizzes seriously although nothing was recorded. Six said they were not afraid of being wrong: "Hindi naman graded, kaya hindi ako kinakabahan kahit mali" (P6). Two added that checking their own sheets felt fair because no classmate saw their score.

The third theme links the routine to the test. Five learners said that questions from earlier lessons kept the patterns of special products and factoring fresh, and three recalled a specific posttest item that they recognized from the recall.

The fourth theme is the routine's weakness. Three learners, two from the bottom third, said they could not answer the question on the previous lesson after an absence, and two said five minutes was too short for items with fractions: "I was still copying when the answers were shown" (P8, bottom third).

Reflection

The retrieval practice section gained about three points more than the comparison section on a 30-item test, a medium effect that held when pretest scores were taken into account and when normality was not assumed. The size of the effect, $g = 0.58$, is close to the averages that reviews of classroom quizzing report: $g = 0.499$ across 222 classroom studies (Yang et al., 2021) and $g = 0.67$ across the classroom studies in Adesope et al. (2017). The focus group suggests why it worked for our learners: the quizzes showed them gaps while there was still time to ask, and the absence of a grade made it safe to be wrong.

We cannot say that the routine alone caused the difference. The sections were intact, and a coin toss between two sections does not balance them the way random assignment of learners would. Although the pretest means were nearly equal, the sections met at different hours, and the retrieval practice section met first in the morning. The same teacher taught both sections and knew which one was using the routine, so the teacher's enthusiasm may have differed between them. Learners of the two sections talk to one another; if comparison learners heard about the quizzes, the difference would be smaller, not larger. Losses were small and even, four learners in each section.

The test measured learning right after the unit. The strongest claim for retrieval practice is that it slows forgetting, and a delayed test four or more weeks after the unit would show whether the advantage lasts. The pretest's low reliability matters little for the gain comparison, since both sections

started near chance, but it limits what the pretest can tell about individual learners.

The 75% target was reached by more learners of the retrieval practice section, but the difference was not significant. A larger study, or a test given again later, would show whether the routine moves more learners past such a target. Learners who were often absent benefited least, because the first question depended on the previous lesson. In the next cycle we will give absent learners the previous day's worked solution on their return, and we will allow seven minutes when an item involves rational expressions.

Daily retrieval practice cost almost nothing: a quarter sheet of paper per learner per day and the teacher's time to write three questions. It took no time from the lesson, because it replaced a review of the same length. We recommend it to the school's Mathematics teachers, with the adjustments above and a delayed test in the next study.

Action Plan

Table 8

Action Plan

Activity	Objective	Persons involved	Timeline	Resources	Success indicator
Present the results at the school's learning action cell (LAC) session for Mathematics teachers	Share the findings and the routine	Research team, Mathematics department, school principal	November 2026	Report, slides, sample recall questions	Decision on adoption recorded in the LAC minutes
Extend Five-Minute Recall to all Grade 8 sections	Give every Grade 8 learner the routine, with the adjustments for absent learners	Grade 8 Mathematics teachers	Term 3, SY 2026–2027	Bond paper, printed worked solutions	Routine held in at least 90% of periods (daily logs)
Build a shared bank of recall questions	Save teachers' time	Grade 8 Mathematics teachers, department head	December 2026 to January 2027	Shared folder	Three questions for every lesson of Term 3
Brief parents at the parent–teacher conference	Explain the ungraded quizzes and ask parents to ask their children about them	Section advisers	End-of-Term Block of Term 2, December 2026	One-page brief (100 copies)	Brief given to every parent who attends
Propose a follow-up study with a delayed test	Find out whether the advantage lasts	Research team	Proposal for the 2027 BERF call	BERF grant	Proposal submitted
Present at the division research forum	Disseminate the results to other schools	Research team	As scheduled by the division	Report, slides	Presentation delivered

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review, 33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Alenton, J. C. (2022). Daily practice test: Effects on mathematics performance in solving the fundamental operations on fractions. *IJERI: International Journal of Educational Research and Innovation*, (18), 47–61. https://doi.org/10.46661/ijeri.4519

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development, 47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology, 30*(1), 92–101. https://doi.org/10.5334/irsp.82

Department of Education. (2016). *Adoption of the basic education research agenda* (DepEd Order No. 39, s. 2016). https://www.deped.gov.ph/wp-content/uploads/2016/06/DO_s2016_039.pdf

Department of Education. (2017). *Research management guidelines* (DepEd Order No. 16, s. 2017). https://www.deped.gov.ph/wp-content/uploads/2017/03/DO_s2017_016.pdf

Department of Education. (2021). *Amendment to DepEd Order No. 016, s. 2017* (DepEd Order No. 026, s. 2021). https://www.deped.gov.ph/wp-content/uploads/2021/06/DO_s2021_026.pdf

Department of Education. (2026a). *Foundational guidelines on artificial intelligence (AI) in basic education* (DepEd Order No. 003, s. 2026). https://www.deped.gov.ph/2026/02/20/february-20-2026-do-003-s-2026-foundational-guidelines-on-artificial-intelligence-ai-in-basic-education/

Department of Education. (2026b). *Revised guidelines on classroom assessment, grading system, and awards and recognition for the K to 12 Basic Education Program* (DepEd Order No. 015, s. 2026). https://www.deped.gov.ph/2026/06/04/june-4-2026-do-015-

s-2026-revised-guidelines-on-classroom-assessment-grading-system-and-awards-and-recognition-for-the-k-to-12-basic-education-program/

Dimitrov, D. M., & Rumrill, P. D., Jr. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159–165. https://doi.org/10.3233/WOR-2003-00285

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine, 292*, Article 114523. https://doi.org/10.1016/j.socscimed.2021.114523

The jamovi project. (2026). *jamovi* (Version 2.7) [Computer software]. https://www.jamovi.org

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151–160. https://doi.org/10.1007/BF02288391

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*(1), 12–19, 39. https://www.aft.org/ae/spring2012/rosenshine

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic

review. *Psychological Bulletin, 147*(4), 399–435.

https://doi.org/10.1037/bul0000309

Financial Report

The study was funded by a BERF grant of PhP 7,540.00 for school-based action research. Under DepEd Order No. 16, s. 2017, as amended by DepEd Order No. 026, s. 2021, 80% (PhP 6,032.00) was released on submission of the work plan and the signed memorandum of agreement, and the last 20% (PhP 1,508.00) is released when the research committee accepts the final report (DepEd, 2017, 2021).

Table 9

Statement of Expenses

Date (2026)	Particulars	Receipt No.	Category	Budget (PhP)	Actual (PhP)
May 22	Photocopying of rating forms, consent letters, and assent forms (190 pages)	OR 0412	Reproduction	300.00	285.00
May 25 to June 5	Transportation to deliver and collect rating forms (4 trips)	Acknowledgment receipts	Domestic travel	600.00	540.00
May 25 to June 5	Mobile load for coordinating with the panel	OR 7718	Communication	500.00	500.00
June 19	Photocopying of test booklets and answer sheets (950 pages)	OR 0436	Reproduction	1,500.00	1,425.00
June 19	Bond paper, A4, 3 reams	OR 2290	Supplies and materials	840.00	855.00
June 19	Printer ink, 4 bottles	OR 2290	Supplies and materials	1,400.00	1,560.00
August 13	Snacks for the focus group discussion (10 persons)	OR 1157	Food	800.00	800.00
October 26	Printing and binding of the final report (4 copies)	OR 0598	Reproduction	1,400.00	1,400.00
October 27	One-page research brief for parents (100 copies)	OR 0599	Dissemination	200.00	200.00
	Total			7,540.00	7,565.00

Note. Receipt numbers are invented for this sample. The PhP 25.00 by which expenses exceeded the grant, from a rise in the price of ink, was paid by the researchers.

Appendices

Appendix A. Data Set

The executive data set is the Excel workbook submitted with this report: each learner's code, section, sex, attendance, and pretest and posttest scores; the item responses; the experts' ratings; the pilot responses; and every computation in this report that Excel can make, as a live formula. The Shapiro –Wilk and Mann–Whitney tests come from jamovi, and the power analysis from G*Power.

Appendix B. Consent and Assent Forms

Kept at the school with the signed forms, and left out of the archived report.

Appendix C. Focus Group Discussion Guide

1. What happens in the first five minutes of our Mathematics period?
2. How do you feel when you answer the recall questions?
3. What do you do when you get a question wrong?
4. Did the recall help you in the test? How?
5. What would you change about the recall?

Appendix D. Declaration of AI Use

We, the researchers, declare that we used [name of the AI tool] only to check the grammar of the English text of this report, for the purpose of finalizing the research study. The prompt used was: "Check the grammar of this paragraph. Do not change its meaning or add content." We have reviewed, edited, and understood all the content, and we take full accountability for it.

No names, Learner Reference Numbers, scores, or other personal information were entered into the AI tool.