About this sample

This is a teaching sample of a completed action research. The study is invented: no school, teacher, parent or learner took part, and every number was generated to behave like a real class's. The DepEd rules and the research it cites are real. Delete this page when you adapt the report for your own study.

Tri-Return Recall is a routine I designed for this sample. If you use it in your own study, cite it as Llego (2026), with the reference given at nash.com.ph/blog/action-research-mathematics-high-school; if you change it and give it a new name, cite it as adapted from Llego (2026).

Mark Anthony Llego

Tri-Return Recall and the Retention of Grade 11 General Mathematics Skills

An Action Research

[Copy for blind review: the researcher's name, position, and school are removed. The final copy lists them here.]

Conducted from June to October 2026

Abstract

This action research tested whether Tri-Return Recall, a routine in which each lesson ends with three questions answered from memory and the same questions return with new numbers after about two days, one week, and three weeks, would help Grade 11 learners keep General Mathematics skills better than the usual review before a test. In Term 1 of School Year 2026–2027, the 24 skills of Weeks 1 to 8 of the budget of work were paired and split by lot into two sets of 12. One section had Tri-Return Recall on Set A and the usual review on Set B; the other section had the reverse, with the same minutes for every skill. Both halves were tested at the end of the term with 24 short-answer items and again four weeks later on a parallel form. Of 80 learners, 77 took part with consent; 72 took each test. A linear mixed model with crossed random effects for learners and skills estimated the gaps in percentage points. At the end of the term, one to three days after the review, the halves were close: the Tri-Return items led by 1.9 points, 95% CI [−2.3, 6.2]. Four weeks later, learners answered 9.7 points more of the Tri-Return items, 95% CI [5.5, 13.9], $p < .001$, and they had forgotten 18.3 points of the reviewed skills against 10.5 of the Tri-Return skills, a difference of 7.8 points, 95% CI [1.8, 13.8], Holm-adjusted $p = .023$. A logistic model and a per-learner comparison agreed. The study cannot tell whether answering from memory or the spacing did the work. Tri-Return Recall will be used in Term 2, and a follow-up will separate its two ingredients.

Keywords: retrieval practice, spaced practice, retention, General Mathematics, senior high school, within-learner design, mixed-effects model

Acknowledgment

I thank the school head, who approved the study and drew the lots in front of the mathematics department; the five validators, who judged every draft item of the test; the Master Teacher in Mathematics who observed the returns and rescored a fifth of the test papers without knowing the sections' sets; the Grade 12 learners who tried the test forms in the Opening Block; the substitute teacher who held my classes on July 15, 2026; the class advisers, who gave out and collected the consent letters; and the parents and learners of the two Grade 11 sections.

Tri-Return Recall and the Retention of Grade 11 General Mathematics Skills

Context and Rationale

My Grade 11 learners forget what they learn faster than the curriculum moves on. In School Year 2025–2026, I gave my two Grade 11 General Mathematics classes the same five problems on simple and compound interest twice: at the end of the lesson that taught them, and again, unannounced, four weeks later. Of 79 learners, the share of correct answers fell from 71% at the lesson's end to 38% four weeks later. In the quarterly examination, after the review I always held in the last periods before it, the same classes answered 63% of the interest items correctly. The review brought the skills back for the examination, after four weeks in which nearly half of the right answers had been lost.

Forgetting costs more in the new General Mathematics. Under the Strengthened Senior High School Curriculum, General Mathematics is one of five core subjects of Grade 11, with 160 hours "in 1 school year (e.g., 3 terms)" (Department of Education [DepEd], 2026e, para. 13). Its Term 1 budget of work builds week on week: wages, salaries and deductions, then percentages, then patterns, sequences and series, which Week 5 uses to compute loan payments and the growth of investments, then measurement, then functions (DepEd, 2026a). A learner who has lost the geometric series of Week 4 cannot do the loan payments of Week 5.

Research on memory suggests two remedies, and they combine. The first is retrieval practice: recalling information strengthens it more than

reading it again. Students who were tested on a prose passage remembered more of it two days and one week later than students who restudied it, although restudy gave more on a test five minutes later (Roediger & Karpicke, 2006). A meta-analysis found practice tests "more beneficial for learning than restudying and all other comparison conditions" (Adesope et al., 2017, abstract), and a review of 50 classroom experiments found medium or large benefits in 57% of their effects, though only 6% were run outside Western, educated, industrialized, rich, and democratic countries (Agarwal et al., 2021). The second remedy is spacing: practice spread over time is retained longer than the same practice massed in one session, and the gap that works best grows with how long the learning must last (Cepeda et al., 2006). Spacing out retrieval combines the two; a meta-analysis of 29 studies found a strong benefit of spaced over massed retrieval practice (g = 0.74), but no difference between gaps that grow and gaps that stay the same (Latimier et al., 2021).

The evidence in mathematics is thinner. In a precalculus course, retention four weeks after the course "benefited exclusively from increasing spacing" (Lyle et al., 2020, abstract), and Year 7 learners who spread their practice over three sessions a week apart scored higher on a test a month later than those who massed it (Emeny et al., 2021). A meta-analysis of mathematics studies, though, found a small effect of spacing within courses (g = 0.24) and no reliable effect of testing over restudy (Murray et al., 2025), and a study of spaced retrieval in nine introductory STEM courses found significant effects in only two of them (Bego et al., 2024). Whether spaced

retrieval helps in a Philippine senior high school class of 40, with quarter sheets of paper and 90-minute periods, is a question for the classroom.

The usual review is also what learners trust. Asked how they study, 84% of college students reported rereading, and only 11% reported practicing recall (Karpicke et al., 2009); adults who learned under massed and spaced conditions "rated massing as more effective than spacing, even after their own test performance had demonstrated the opposite" (Kornell & Bjork, 2008, p. 585). A routine that asks learners to answer from memory, without notes, may feel harder than a review while it works better.

This study falls under Theme 1 of the Basic Education Research Agenda, Teaching and Learning, sub-theme Instruction (DepEd, 2016). It asked whether Tri-Return Recall would help my Grade 11 learners keep the Term 1 skills four weeks after the term, compared with the usual review before the test, for the same class time.

Innovation, Intervention, and Strategy

The innovation was Tri-Return Recall, a routine that brings each lesson's skill back three times, at growing gaps, for learners to answer from memory.

The Routine

Each lesson on a Tri-Return skill ended with three questions on the board, answered on a quarter sheet of paper without notes or help in about four minutes. I then showed the worked answers, and each learner marked their own sheet and wrote the step they had missed (6 minutes in all). The same three questions came back three times, with new numbers: at the first

class meeting at least 2 days after the lesson, at least 7 days after it, and at least 21 days after it. Each return took the first 6 minutes of a period and followed the same steps: answer from memory, see the worked answers, mark, and note the missed step. The three questions asked for the skill in three ways: a direct computation, the skill inside a short problem, and the reason for a step (Table 1). The sheets were not collected for grades; I recorded only who answered each question correctly, for this study.

Table 1

The Three Questions of the VAT Skill at the Lesson and Its Three Returns

Occasion	Date	Question 1: the VAT in a receipt, VAT included at 12%	Answer	Question 2: the price with 12% VAT	Answer
Lesson end	June 23, 2026	₱1,120	₱120	₱8,500 before VAT	₱9,520
Return 1	June 25, 2026	₱2,240	₱240	₱6,250 before VAT	₱7,000
Return 2	June 30, 2026	₱560	₱60	₱12,000 before VAT	₱13,440
Return 3	July 14, 2026	₱3,360	₱360	₱4,750 before VAT	₱5,320

Note. Question 3 was the same at every return: "A classmate finds the VAT in a VAT-inclusive price by taking 12% of it. What is wrong, and what is the right first step?" Dates are Section 2's, which had the VAT skill as a Tri-Return skill.

The return dates were planned from the school calendar before the term (DepEd, 2026c, Annex B): a return passed over holidays, the teacher-made summative tests of July 6 and July 28, 2026, and days with no General Mathematics period, and went to the next meeting. Every return's planned day and the day it was held are in the return log (Appendix D).

The Comparison: The Usual Review, for the Same Minutes

A reviewed skill was taught the way I had always taught it. Its lesson ended with my usual summary: I restated the steps on the board while learners copied them, for the same 6 minutes. Before the end-of-term test, on August 24 to 26, 2026, I reviewed each reviewed skill once, as I always had before an examination: one worked example on the board, its steps named, and two practice problems answered with the class, notes open, for 18 minutes a skill, four skills a day. So each skill had 24 planned minutes beyond its lesson in either half: a Tri-Return skill's 6 at the lesson and three returns of 6, a reviewed skill's 6 at the lesson and 18 in the review. The two halves differed in when the time came and in whether learners answered from memory, not in how much time they had.

The Skills and the Lot

The 24 skills were three from each of Weeks 1 to 8 of the Term 1 budget of work, each one a skill a short problem can test, in the budget of work's own competencies (Table 2). Weeks 9 and 10 were left out because a three-week return for their lessons would have fallen after the end-of-term test. Within each two-week block, the six skills were paired by the number of steps in their worked answers. On June 11, 2026, in the Opening Block, the school head tossed a coin for each of the 12 pairs: heads sent the first skill of the pair to Set A and the second to Set B, tails the reverse. A second toss decided that Section 1 would have Set A as its Tri-Return skills and Section 2 Set B. Every learner therefore had 12 Tri-Return skills and 12 reviewed skills, three of each

from every two weeks of lessons, and every skill was a Tri-Return skill in one section and a reviewed skill in the other.

Table 2

The 24 Skills, Their Pairs, and the Lot

Week	Skill	Budget of work competency	Steps	Pair	Set
1	S01 Pay from an annual salary	Calculate weekly or monthly wages from an annual salary, wages from an hourly rate, including situations involving overtime and other allowances, and earnings based on commission or piecework.	2	1	A
1	S02 Overtime pay	Calculate weekly or monthly wages from an annual salary, wages from an hourly rate, including situations involving overtime and other allowances, and earnings based on commission or piecework.	3	3	A
1	S03 Gross and net pay	Solve problems involving salaries, wages, benefits, and deductions (tax computations, overtime pay, and gross and net incomes) using appropriate technology.	3	3	B
2	S04 Mark-up and discount	Apply a percentage increase or decrease in various contexts: b. calculating percentage mark-ups and discounts	2	1	B
2	S05 VAT	Apply a percentage increase or decrease in various contexts: c. calculating VAT	2	2	B
2	S06 Profit or loss	Apply a percentage increase or decrease in various contexts: d. calculating profit or loss in both absolute and percentage terms	2	2	A
3	S07 Next terms of a pattern	Determine the next term of a given pattern, including the Fibonacci sequence.	2	4	A
3	S08 The rule of a pattern	Identify the rule of a given pattern.	2	4	B
3	S09 Sequence problems	Solve problems involving arithmetic and geometric sequences.	3	5	A
4	S10 Sigma notation	Represent series using sigma notation and vice versa.	3	5	B
4	S11 Arithmetic series	Find the sum of arithmetic and geometric series.	3	6	B
4	S12 Series problems	Solve problems involving arithmetic and geometric series.	3	6	A
5	S13 Loan payments	Apply sequences and series to solve financial problems, such as periodic payments for loans and mortgages.	4	9	B
5	S14 One deposit, compound interest	Determine the accumulated values of investment over time using sequences and series.	2	7	A

Week	Skill	Budget of work competency	Steps	Pair	Set
5	S15 Regular deposits	Determine the accumulated values of investment over time using sequences and series.	3	9	A
6	S16 Metric prefixes	Determine and apply the appropriate metric units for various contexts, such as units for small and large quantities and scientific measurements.	2	7	B
6	S17 Area and volume units	Convert measurements with different units in the same system and between different systems.	2	8	A
6	S18 Between systems	Convert measurements with different units in the same system and between different systems.	2	8	B
7	S19 Surface area of a cylinder	Determine the surface area of three-dimensional objects, such as prisms, pyramids, cylinders, cones, spheres, and combinations of these.	3	11	A
7	S20 Volume of a cone	Calculate the volume of prisms, pyramids, cylinders, cones, spheres, and combinations of these.	2	10	B
7	S21 Volume of a composite solid	Calculate the volume of prisms, pyramids, cylinders, cones, spheres, and combinations of these.	3	11	B
8	S22 Map and blueprint scales	Use a scale to interpret and compare distances on maps, blueprints, and sketches.	2	10	A
8	S23 Currency conversion	Solve problems involving the conversion of currencies.	3	12	A
8	S24 Time conversions	Apply time conversions in various real-world scenarios such as travel planning, project management, and daily routine schedules for personal and business transactions.	3	12	B

Note. Competencies as printed in the General Mathematics budget of work, Term 1 (DepEd, 2026a); the percentage skills are its parts b, c, and d. Steps: the number of steps in the worked answer to the skill's test item, on which the pairs were matched. Section 1 had Set A as Tri-Return skills and Section 2 Set B.

Strategy and Evaluation Plan

Three criteria for success were set in the proposal, before the term, and two of them rest on the four-week test, since a review given one to three days before the end-of-term test would leave the reviewed skills fresh then: (a) four weeks after the end of the term, learners answer at least 8 percentage points more of the Tri-Return items than of the reviewed items (one skill of the twelve

is 8.3 points), with $H_0 1$ rejected at the .05 level; (b) learners forget less of the Tri-Return skills than of the reviewed skills between the two tests, with $H_0 2$ rejected; and (c) the routine runs as planned, with at least 90% of the returns held within one meeting of their planned day and the two halves' minutes per skill within 10% of each other. The gap at the end of the term would be reported but was not a criterion.

Action Research Questions

Main question: Does Tri-Return Recall help Grade 11 learners keep General Mathematics skills four weeks after the end of the term better than the usual review before the test, for the same class time?

1. How well did the routine run, in returns held, minutes, and attendance, and how did learners do on the returns?
2. At the end of the term, did learners answer more of the Tri-Return items correctly than of the reviewed items?
3. Four weeks later, did they answer more of the Tri-Return items correctly?
4. Between the two tests, did they forget less of the Tri-Return skills than of the reviewed skills?
5. Which half did learners expect to remember better, and how did their expectations compare with their results?

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

$H_0 1$: Four weeks after the end of the term, the percentage of Tri-Return items answered correctly does not differ from that of reviewed items.

$H_0$2: The drop in the percentage answered correctly from the end-of-term test to the four-week test does not differ between Tri-Return and reviewed skills.

$H_0$3: At the end of the term, the percentage of Tri-Return items answered correctly does not differ from that of reviewed items.

Action Research Methods

Research Design

The study used a within-learner design: the skills, not the learners or the sections, were assigned to the two conditions by lot, so every learner had both the routine and the usual review, and each learner's own results on the two halves were compared. The two sections had the sets the other way round, so that every skill was measured both ways and a set that happened to be harder could not pass for a weaker routine. Classroom studies of quizzing have used the same plan, rotating which material was quizzed across sections (Roediger et al., 2011). The term ran from June 8 to September 15, 2026, with its Instructional Block from June 15, 2026 to September 1, 2026 and the Term 1 Examination on August 28 and September 1, 2026 (DepEd, 2026c, Annex B). The end-of-term test was given on August 27, 2026, the last General Mathematics period before the examination, and the four-week test on September 24, 2026, in Term 2, which began on September 16, 2026.

Participants and Other Sources of Data and Information

The participants were the learners of my two Grade 11 sections in a public secondary school, School Year 2026–2027, 40 and 40 learners, the maximum class size DepEd sets for core subjects (DepEd, 2026e, Table 7).

Both met General Mathematics four times a week, Monday to Thursday, for 90 minutes, the first of the time allotments DepEd suggests for a core subject, "1.5 hours, 4x a week" (DepEd, 2026e, Table 8). All 80 learners were invited; the parents of 77 consented and these learners gave their assent (Table 3). Every learner had the lessons, the returns, and the reviews; consent decided only whose answers were recorded. The 77 were 39 girls and 38 boys, 47 aged 16 and 30 aged 17 at the start of the term. One learner transferred to another school on July 24, 2026, before either test.

Table 3

Learners Enrolled, Consenting, and Tested, by Section

Learners	Section 1	Section 2	Total
Enrolled	40	40	80
Parent declined	2	0	2
Consent form not returned	0	1	1
With consent and assent	38	39	77
Transferred before the tests	0	1	1
Took the end-of-term test	37	35	72
Took the four-week test	37	35	72

Note. Every enrolled learner took part in the lessons, returns, and reviews; only consenting learners' answers were recorded. In all, 69 learners took both tests.

Sample Size

With two whole sections, the number of learners was fixed, so the proposal asked how large a difference the study could detect. I simulated the study 400 times under the proposal's assumptions (about 76 learners with consent, 45% of reviewed items right four weeks after the term, and learners and skills as varied as my previous year's tests) and analyzed each simulation as planned. The power to detect a four-week difference of 8 percentage points was .99, and of 5 points .67.

Data Gathering Methods

The Retention Test

The test was my own: 24 short-answer items, one for each skill, in two parallel forms, P and Q, with the same skill in each item and new numbers. Learners wrote their answers and solutions, with a scientific calculator, in 60 minutes. Each answer was scored right or wrong against a written key, whose rules were set before the first test: an amount in pesos to the centavo, the unit named, and any equivalent form accepted. Half of each section, dealt down a shuffled class list, took Form P at the end of the term and Form Q four weeks later; the other half took Q, then P. The end-of-term test was announced; the four-week test was not, so that it measured what learners kept rather than what they studied the night before. Neither test counted toward grades.

Five validators (two Master Teachers in Mathematics, the department head, an Education Program Supervisor in Mathematics, and a mathematics teacher of another senior high school) judged each of 56 draft items: whether it tested its skill as the budget of work states it, whether its numbers kept it to one skill, and whether it matched its parallel item. Lawshe's content validity ratio, CVR = (n_e − N/2)/(N/2), was computed for each item (Lawshe, 1975); with five validators, an item passes at the .05 level only when all five rate it essential, CVR = 1.00 (Ayre & Scally, 2014). Of the 56 drafts, 51 passed; three planned items that did not were replaced by spares that did, and the 48 items kept all had CVR = 1.00. On June 10, 2026, 38 Grade 12 learners tried both forms to check the wording and the time: the median learner finished in 43

minutes and the slowest in 55, so 60 minutes were allowed. Their scores were not used: they were not the study's learners, and they had met most of these skills in earlier grades.

A second scorer, a Master Teacher I in Mathematics of the school who did not teach the sections, rescored a fifth of the papers of each test, drawn by lot by code, without knowing which set each section had. The reliability of each form at each test was computed with the Kuder–Richardson formula 20 (KR-20; Kuder & Richardson, 1937).

The Return Log and the Returns' Answers

For each section, I logged every lesson-end recall, return, lesson-end summary, and review: its planned day, the day it was held, and its minutes. For each recorded learner and Tri-Return question, I noted whether the answer was right; learners absent on a return's day were marked absent for it. The Master Teacher observed 12 returns, six in each section drawn by lot, with a five-step checklist: notes closed, the same three questions with new numbers, about four minutes to answer, the worked answers shown and marked by each learner, and no score recorded.

Learners' Predictions

Before the end-of-term test, each learner answered one question on paper: "Four weeks from now, which skills do you think you will remember better: the skills that came back three times, the skills we reviewed this week, or neither?"

Storage

Papers were kept by code in a locked cabinet in the mathematics faculty room. The list linking codes to names was kept by the school's research coordinator, apart from the papers, and will be destroyed when the report is accepted.

Ethical Considerations

Parents received a letter explaining the study at the Grade 11 parents' orientation on June 9, 2026, with a consent form; learners whose parents consented signed an assent form on June 11, 2026, before the first lesson, as DepEd's quality tools for completed research require of studies with learners below 18 (DepEd, 2022, Annex 2, Item 13). The study followed the research ethics of DepEd's Research Management Guidelines, which refer researchers to the DepEd Child Protection Policy "for interacting with learners" (DepEd, 2017, Sec. VI.C; DepEd, 2012). The returns, the quarter sheets, and both tests were formative and counted toward no grade: "Formative assessment shall be used to monitor learning progress but shall not be used for grade computation" (DepEd, 2026d, para. 15). A learner's answers are personal information, and information about a learner's education is sensitive personal information (Republic Act No. 10173, 2012, Secs. 3(g) and 3(l)(2)), so the data set holds codes, not names. No learner went without the routine or the review: each had both, on different skills. My use of an artificial intelligence (AI) tool is declared in Appendix F (DepEd, 2026b, Annex A).

Data Analysis

Question 1 was answered from the return log and the returns' answers. For Questions 2 to 4, the main analysis was a linear mixed model of every item score (1 right, 0 wrong) at both tests, with crossed random intercepts for learners and skills (Baayen et al., 2008), and fixed effects for the half (Tri-Return or reviewed), the test (end of term or four weeks), their interaction, the section, and the form. With a 0 or 1 outcome, the model's coefficients are differences in the proportion right, reported here as percentage points. The gap at the end of the term ($H_0 3$), the gap at four weeks ($H_0 1$), and the difference in forgetting ($H_0 2$, the interaction) were estimated from the model with Wald 95% confidence intervals. $H_0 1$ was the primary hypothesis; $H_0 2$ and $H_0 3$ were adjusted for two tests by Holm's (1979) procedure. The model was fitted by restricted maximum likelihood with three optimizers, and the converged fit with the highest likelihood was kept.

Three analyses checked the main one. A logistic mixed model with the same random effects, fitted by variational Bayes, gave the gaps as odds ratios with 95% credible intervals. Each learner's percent correct on each half at each test was compared with a paired t test, with the effect size d_z (the mean difference over its standard deviation); because the four-week differences were not normally distributed (Shapiro–Wilk $W = .95$, $p = .010$), a Wilcoxon signed-rank test checked the paired t test there. Each skill's percent correct in the section that had it as a Tri-Return skill was compared with its percent correct in the section that reviewed it, and a sign test counted the skills ahead. The models were fitted in Python with statsmodels (Seabold & Perktold, 2010); the

Excel workbook in Appendix A recomputes every percentage, the per-learner tests, the per-skill comparison, KR-20, and the scorers' agreement with formulas.

Discussion of Results and Reflection

Implementation

Question 1 asked how well the routine ran. Of the 72 returns planned in the two sections, 69 were held on their planned day. The other three were planned for July 15, 2026, when I was on sick leave and the substitute teacher held only the lesson; they were held at the next meeting, a day later. The first returns came 2 to 6 days after their lessons, the second 7 to 8, and the third 21 to 22. A return took 6.3 minutes on average (5 to 8). The two halves had almost the same time (Table 4): 25.2 minutes a skill for the Tri-Return skills and 24.1 for the reviewed skills. In the 12 returns observed, 59 of the 60 checklist steps were done; in one, the questions kept the lesson's own numbers. Criterion (c) was met.

Table 4

Minutes per Skill Beyond the Lesson, by Half and Section

Section	Tri-Return skills: *M* (range)	Reviewed skills: *M* (range)
Section 1	24.7 (22 to 27)	24.2 (22 to 26)
Section 2	25.8 (22 to 29)	24.0 (22 to 26)
Both sections	25.2 (22 to 29)	24.1 (22 to 26)

Note. Tri-Return skills: the lesson-end recall and three returns. Reviewed skills: the lesson-end summary and the skill's share of the review on August 24 to 26, 2026. Means of the 12 skills of each half in each section, from the return log.

Recorded learners were present for 93.7% of the lesson-end recalls and returns of their Tri-Return skills, and for 78.9% of their Tri-Return skills they took all four. On the returns' questions, learners were right on 72.8% at the

end of the lesson, 53.8% at the first return, 58.8% at the second, and 62.1% at the third: the first return was the hardest, and each return after it was easier although the gaps grew.

Questions 2 and 3: The Halves at the End of the Term and Four Weeks Later

Table 5 gives each section's percent correct on each half at each test.

Table 5

Percent Correct by Section, Half, and Test

Section	Half	Set	End of term: *n*	End of term: *M* (*SD*)	Four weeks: *n*	Four weeks: *M* (*SD*)
Section 1	Tri-Return	A	37	69.1 (24.4)	37	58.6 (22.9)
Section 1	Reviewed	B	37	60.8 (25.4)	37	44.1 (21.1)
Section 2	Tri-Return	B	35	55.7 (19.3)	35	44.5 (17.8)
Section 2	Reviewed	A	35	60.2 (19.7)	35	39.5 (20.0)
Both sections	Tri-Return	A and B	72	62.6 (22.9)	72	51.7 (21.7)
Both sections	Reviewed	A and B	72	60.5 (22.7)	72	41.9 (20.6)

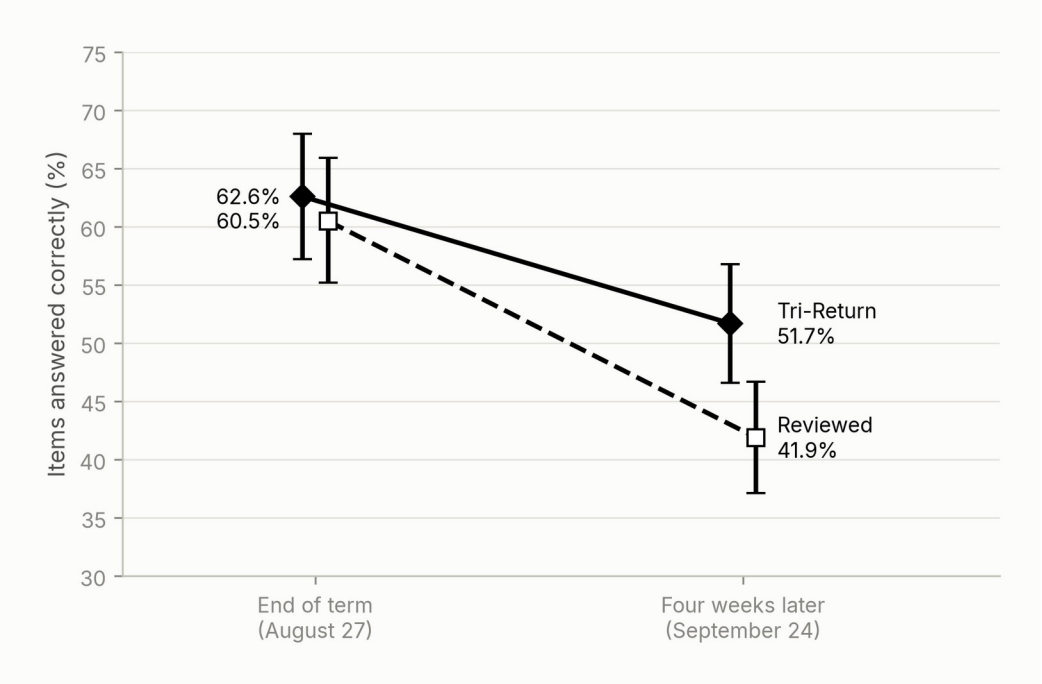
Note. Means and standard deviations of learners' percent correct on the 12 items of each half. Section 1 had Set A as its Tri-Return skills; Section 2, Set B.

At the end of the term, one to three days after the review, the halves were close: learners answered 1.9 percentage points more of the Tri-Return items, 95% CI [−2.3, 6.2], $z = 0.90$, Holm-adjusted $p = .368$ (Table 6). $H_0 3$ is not rejected; the interval allows a gap from a few points in favor of the review to a few points in favor of the routine.

Four weeks later the halves had parted. Learners answered 9.7 percentage points more of the Tri-Return items, 95% CI [5.5, 13.9], $z = 4.48$, $p < .001$. $H_0 1$ is rejected, and the estimate met criterion (a), 8 points. The interval, though, reaches down to 5.5 points, so the true gap may be smaller than the one the criterion asked for.

Figure 1

Percent Correct on Each Half at the End of the Term and Four Weeks Later



Note. Learners' percent correct on their 12 Tri-Return and 12 reviewed skills; error bars are 95% confidence intervals. n = 72 at each test.

Question 4: Forgetting

The model estimated that, between the two tests, learners lost 18.3 percentage points of the reviewed skills, 95% CI [14.0, 22.5], and 10.5 of the Tri-Return skills, 95% CI [6.2, 14.8]: a difference of 7.8 points, 95% CI [1.8, 13.8], z = 2.53, Holm-adjusted p = .023 (Table 6; Figure 1 shows the raw means). $H_0$2 is rejected, and criterion (b) was met.

Table 6

Tests of the Three Null Hypotheses

Hypothesis	Comparison	Estimate (points)	95% CI	z	p	Holm p	Decision
$H_0$1	Tri-Return less reviewed, four weeks	9.7	[5.5, 13.9]	4.48	< .001	(primary)	Rejected

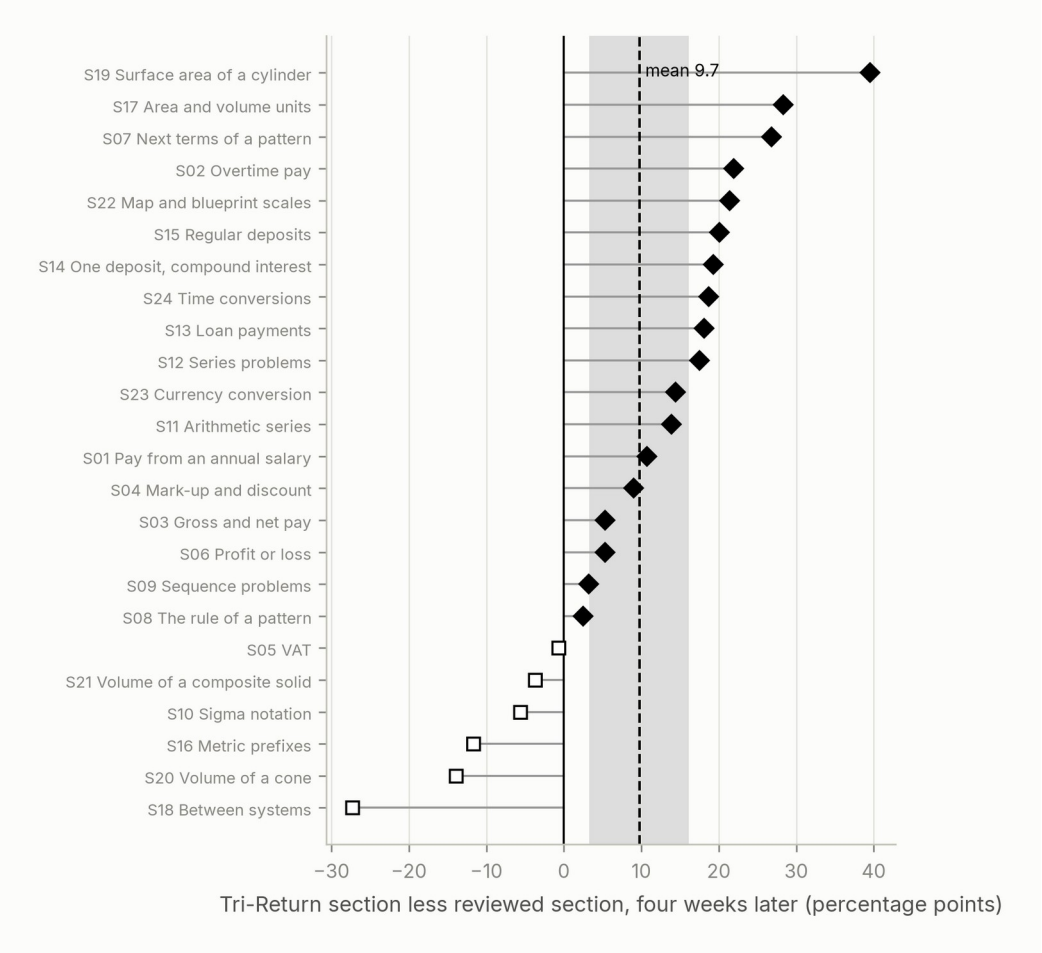
Hypothesis	Comparison	Estimate (points)	95% CI	z	p	Holm p	Decision
$H_0 2$	Drop of the reviewed less drop of the Tri-Return skills	7.8	[1.8, 13.8]	2.53	.011	.023	Rejected
$H_0 3$	Tri-Return less reviewed, end of term	1.9	[−2.3, 6.2]	0.90	.368	.368	Not rejected

Note. Linear mixed model of 3,456 item scores of 75 learners on 24 skills, with crossed random intercepts for learners and skills, and the half, the test, their interaction, the section, and the form as fixed effects; estimates in percentage points, Wald 95% confidence intervals. Holm's adjustment over $H_0 2$ and $H_0 3$; $H_0 1$ was the primary hypothesis.

The three checks agreed. In the logistic mixed model, the odds of a right answer on a Tri-Return item were 1.61 times those on a reviewed item four weeks after the term, 95% credible interval [1.39, 1.87], against 1.13, [0.97, 1.31], at the end of the term. Compared learner by learner, the 72 learners at the four-week test answered 9.8 points more of their Tri-Return items, 95% CI [4.9, 14.7], $t(71) = 4.01$, $p < .001$, $d_z = 0.47$; 49 did better on the Tri-Return half, 17 on the reviewed half, and 6 equally well. The Wilcoxon signed-rank test agreed, $V = 1{,}638$, $p < .001$. At the end of the term the per-learner difference was 2.1 points, 95% CI [−2.2, 6.3], $p = .332$. Skill by skill, 18 of the 24 skills were answered better four weeks later in the section that had them as Tri-Return skills, a mean difference of 9.7 points, 95% CI [3.3, 16.1], sign test $p = .023$ (Figure 2). At the end of the term, 13 of the 24 were ahead.

Figure 2

Each Skill's Four-Week Difference Between Its Tri-Return and Its Reviewed Section



Note. Each skill's percent correct four weeks after the term in the section that had it as a Tri-Return skill, less in the section that reviewed it. Filled diamonds: skills ahead as Tri-Return skills; open squares: skills behind. Dashed line: the mean, 9.7 points; shaded: its 95% confidence interval.

The two sets were not equally hard: Set B's skills were answered 5.6 points less often than Set A's, whichever half they were in. That is why the sections swapped them. Taken alone, Section 1, which had the easier set as its Tri-Return skills, would have put the four-week gap at 14.4 points, 95% CI [3.6, 25.2], and Section 2 at 5.0 points, 95% CI [−3.9, 13.9]; together, each skill counted both ways, the sets' difference cancels.

The test was reliable enough to compare the halves: KR-20 was .83 for Form P and .81 for Form Q at the end of the term, and .73 and .72 at four weeks. The two scorers agreed on 99.4% of 672 answers on 28 papers, κ = .99; the key's rules settled the four disagreements.

Question 5: What Learners Expected

Before the end-of-term test, 28 of 72 learners expected to remember the reviewed skills better four weeks later, 30 the Tri-Return skills, and 14 neither. Of the 28 who expected the review to last, 26 took the four-week test, and 21 of them remembered more of their Tri-Return half.

Reflection

What I Found

Answering three questions from memory, three times over three weeks, did not make the Tri-Return skills look better at the end of the term, when the review was one to three days old. It made them last. Four weeks later, my learners answered about 10 points more of the Tri-Return items, and they had lost little more than half as much of those skills as of the reviewed ones. Most learners and most skills went the same way, so the result was not the work of a few learners or of the easier set.

How It Compares With the Research

The pattern is the one Roediger and Karpicke (2006) found with prose: the comparison condition looked as good or better on an early test, and retrieval pulled ahead on later ones. It also matches the precalculus study in which retention four weeks after the course depended on spacing (Lyle et al., 2020). The returns' own scores resemble what Lyle et al. (2022) found in

calculus, where spaced practice lowered scores on the practice quizzes and raised retention: answering after a gap was harder than at the lesson's end, with the first return the lowest point, and the skills were kept better. Murray et al. (2025) put the effect of spacing within mathematics courses at $g = 0.24$ and suggested it may be smaller than in other domains; this study set three spaced retrievals against one massed review, so the two ingredients were tested together.

What Else Could Explain the Difference

The Tri-Return skills came back four times and the reviewed skills twice, at the lesson's end and in the review, for the same minutes; the study tested the whole routine against the usual practice and cannot say whether answering from memory, the spacing, or the number of occasions did the work. Learners knew which of their skills were Tri-Return skills and may have studied the others more for the Term 1 Examination of August 28, which would have narrowed the four-week gap, not widened it. I taught both sections and knew the sets, but the test was scored against a key, and the second scorer, who did not know the sets, agreed almost entirely. The sections met at different hours, and the swap of sets put each skill in both.

Limits

The study ran in two sections of one school, with one teacher, in one term. The test was my own; at four weeks its reliability (.72 and .73) was adequate for comparing halves of a class, not for judging one learner. The four-week test came in the second week of Term 2, after the End-of-Term Block; a longer wait was not tested. The skills were short computations, so

the study says nothing about longer problems that join several skills. The learners' predictions were one question on paper.

What I Would Change

I would test the two ingredients apart: in one set, three returns from memory; in the other, three short reviews on the same days, for the same minutes. I would add a test at the end of Term 2, to see whether the gap holds over a longer wait, and I would check that every return changes its numbers.

Conclusion

Tri-Return Recall helped my Grade 11 learners keep General Mathematics skills four weeks after the term, for no more class time than the usual review. It did not show at the end of the term, which is when teachers usually check. I will use it for every Term 2 skill in both sections, keep the return log, and recommend it to the department with the planner and the question cards.

Action Plan

Table 7

Action Plan

Activity	Objective	Time frame	Persons involved	Resources	Expected output
Use Tri-Return Recall for every Term 2 skill in both sections	Keep the routine running for all skills, with the return log	September to December 2026	Researcher	Quarter sheets, question cards, the planner	Returns held within one meeting of plan for at least 90% of lessons
Share the routine, the planner, and the question cards at the mathematics department's learning action cell	Let the other Grade 11 teachers try it in Term 3	November 2026	Researcher, department head	Report, planner workbook, sample cards	Teachers who will try it named in the LAC minutes
Write the three questions and their new numbers for the Term 2 and Term 3 skills with the other Grade 11 teachers	Save each teacher's time	November 2026 to January 2027	Grade 11 mathematics teachers	Budget of work, shared folder	A card for every skill of Terms 2 and 3
Test the two ingredients apart in Term 3: returns from memory against short reviews on the same days	Learn whether retrieval or spacing does the work	January to March 2027	Researcher, Master Teacher	The same design, a new lot	A second report
Present the results at the division research forum and submit the report to the division's repository	Share the findings	As the division schedules	Researcher	Report and slides	Presentation and archived report

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

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bego, C. R., Lyle, K. B., Ralston, P. A. S., Immekus, J. C., Chastain, R. J., Haynes, L. D., Hoyt, L. K., Pigg, R. M., Rabin, S. D., Scobee, M. W., & Starr, T. L. (2024). Single-paper meta-analyses of the effects of spaced retrieval practice in nine introductory STEM courses: Is the glass half full or half empty? *International Journal of STEM Education, 11*, Article 9. https://doi.org/10.1186/s40594-024-00468-5

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Department of Education. (2012). *DepEd child protection policy* (DepEd Order No. 40, s. 2012). https://www.deped.gov.ph/wp-content/uploads/2012/05/DO_s2012_40.pdf

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. (2022). *Adopting tools to improve quality management of completed research at the Department of Education* (DepEd Memorandum No. 028, s. 2022). https://www.deped.gov.ph/2022/04/05/april-5-2022-dm-028-s-2022-adopting-tools-to-improve-quality-management-of-completed-research-at-the-department-of-education/

Department of Education. (2026a). *Budget of work: General Mathematics, Grade 11* [Three-term school calendar].

https://www.deped.gov.ph/wp-content/uploads/General-Mathematics-2.pdf

Department of Education. (2026b). *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. (2026c). *Guidelines on the implementation of the three-term school calendar in basic education* (DepEd Order No. 009, s. 2026). https://www.deped.gov.ph/2026/04/16/april-16-2026-do-009-s-2026-guidelines-on-the-implementation-of-the-three-term-school-calendar-in-basic-education/

Department of Education. (2026d). *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/

Department of Education. (2026e). *Strengthened senior high school curriculum* (DepEd Order No. 017, s. 2026). https://www.deped.gov.ph/2026/06/04/june-4-2026-do-017-s-2026-strengthened-senior-high-school-curriculum/

Emeny, W. G., Hartwig, M. K., & Rohrer, D. (2021). Spaced mathematics practice improves test scores and reduces overconfidence. *Applied Cognitive Psychology, 35*(4), 1082–1089. https://doi.org/10.1002/acp.3814

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17*(4), 471–479. https://doi.org/10.1080/09658210802647009

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

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

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review, 33*(3), 959–987. https://doi.org/10.1007/s10648-020-09572-8

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

Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. S. (2020). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review, 32*(1), 277–295. https://doi.org/10.1007/s10648-019-09489-x

Lyle, K. B., Bego, C. R., Ralston, P. A. S., & Immekus, J. C. (2022). Spaced retrieval practice imposes desirable difficulty in calculus learning. *Educational Psychology Review, 34*(3), 1799–1812. https://doi.org/10.1007/s10648-022-09677-2

Murray, E., Horner, A. J., & Göbel, S. M. (2025). A meta-analytic review of the effectiveness of spacing and retrieval practice for mathematics learning. *Educational Psychology Review, 37*(3), Article 75. https://doi.org/10.1007/s10648-025-10035-1

Republic Act No. 10173. (2012). *Data Privacy Act of 2012*. https://lawphil.net/statutes/repacts/ra2012/ra_10173_2012.html

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*(4), 382–395. https://doi.org/10.1037/a0026252

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

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In S. van der Walt & J. Millman (Eds.),

Proceedings of the 9th Python in Science Conference (pp. 92–96).

https://doi.org/10.25080/Majora-92bf1922-011

Financial Report

Table 8

Statement of Expenses

Date	Particulars	Receipt no.	Quantity	Amount (PhP)	Category
June 5, 2026	Printing: parents' consent letters and learners' assent forms	OR 1184	170 pages at 1.00	170.00	Reproduction
June 8, 2026	Printing: pilot forms (two forms, 4 pages each)	OR 1190	320 pages at 1.00	320.00	Reproduction
June 12, 2026	Bond paper, long, for the quarter-sheet return slips (2 reams)	OR 1203	2 reams at 265.00	530.00	Supplies and materials
June 12, 2026	Printing: worked-answer cards for the 24 skills, one set for each section	OR 1203	96 pages at 3.00	288.00	Reproduction
August 24, 2026	Printing: end-of-term test, Forms P and Q (4 pages each)	OR 1377	320 pages at 1.00	320.00	Reproduction
September 21, 2026	Printing: four-week test, Forms P and Q (4 pages each)	OR 1412	320 pages at 1.00	320.00	Reproduction
October 23, 2026	Printing and binding of the final report	OR 1468	3 copies at 280.00	840.00	Reproduction
Total				**2,788.00**	

Note. Fund source: the researcher's own funds. No school funds or grants were used, and no learner paid for anything. Receipt numbers are invented for this sample.

Appendices

Appendix A. Data Set and Notebook

The Excel data set (the lot, the learners by code, every item score at both tests with the second scorer's marks, every answer on the returns, the return log, attendance, the predictions, the validators' ratings, and every statistic Excel can compute as a formula, with Python's values for the mixed models, labeled) and the Jupyter notebook that reproduces every result from the data.

Appendix B. Parents' Consent Letter and Learners' Assent Form

The blank letter and form go with the supporting documents the evaluators check (DepEd, 2022, Annex 2, Item 13); the signed forms stay with the school and are left out of the archived report.

Appendix C. The Retention Test

Forms P and Q, the key with its scoring rules, and the validators' ratings. Kept with the school's copy so that the items can be used again.

Appendix D. The Routine

The three questions and worked answers of each of the 24 skills, with the new numbers of each return; the return log of both sections; and the planner used to set the return dates.

Appendix E. Observation Checklist

The Master Teacher's five-step checklist for the returns, and the draw of the returns observed.

Appendix F. Declaration of AI Use

I, the researcher, declare that I 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." I have reviewed, edited, and understood all the content, and I take full accountability for it. No names, codes, scores, or other personal information were entered into the AI tool (adapted from Annex A of DepEd Order No. 003, s. 2026).