Spaced Reteaching and the Least Mastered Competencies of Grade 7 Learners

An Action Research

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

Conducted August to October 2025

Abstract

This action research tested whether short, spaced reteaching of a Grade 7 Mathematics class's least mastered competencies would raise the learners' mastery of them more than ordinary instruction. The list came from the class's Quarter 1 test in School Year 2025–2026: after an item review corrected a miskeyed item and an item with two right answers, the six competencies with the lowest percent correct were taken. A parallel retest a week later, before any reteaching, gave the baseline. The six were split into two sets, and a lot decided which set came first. Each set had nine 15-minute sessions over three weeks (a retrieval check, worked examples, and practice with feedback), Set A's while Set B waited its turn, and new parallel items measured both sets after each phase. Of the 44 learners with consent, 43 were analyzed in Phase 1 and 42 in Phase 2. The set being retaught gained 16.5 percentage points more than the waiting set in Phase 1, 95% CI [7.2, 25.7], $t(42) = 3.60$, $p < .001$, and 16.3 points more than Set A in Phase 2, 95% CI [7.5, 25.0], $p < .001$, a difference that includes Set A's fall of 5.2 points after its own reteaching. Without any reteaching, the six had risen 4.4 points from the Quarter 1 test to the retest, $p = .042$: regression to the mean, which a before-and-after comparison would have counted as a gain. After reteaching, three of the six competencies were mastered by at least 75% of the learners. Spaced reteaching is recommended, with a fresh baseline before it and a review of each set after it.

Keywords: least mastered competencies, item analysis, regression to the mean, reteaching, Grade 7 mathematics

Acknowledgment

The researchers thank the school head, who approved the study and drew the lot that set the order of the reteaching; the Master Teacher in Mathematics, who wrote the parallel items of the retest and the two probes from the item specifications before the Quarter 1 test, and assembled the probes before the lot that set the order of the reteaching; the Head Teacher of the Mathematics department, who reviewed the two faulty items with us and observed six of the reteaching sessions; the Grade 7 adviser, who sent home and collected the consent forms; and the 46 learners of the class and their parents. The study was funded by the researchers.

Spaced Reteaching and the Least Mastered Competencies of Grade 7 Learners

Context and Rationale

Every quarter, our Mathematics department lists each class's least mastered competencies from the quarterly test and plans their reteaching. In School Year 2025–2026, our Grade 7 learners took the MATATAG curriculum's Quarter 1 competencies: polygons and their angles, percentages, rates, and rational numbers (Department of Education [DepEd], 2023, p. 52). On our Quarter 1 test, given on August 20 and 22, 2025, the class's mean percentage score, after an item review, was 62.73, and six of the eleven competencies the test covered scored below 60%.

DepEd's assessment policy for that school year asked teachers to use assessment results to decide "whether to review, re-teach, remediate, or enrich lessons" (DepEd, 2015, Sec. IV), and its policy on learning gaps in secondary schools lists "Teacher modeling followed by guided practice and independent practice" and "Strategic Intervention Materials (SIMs), which are worksheets prepared by teachers targeting the least mastered competencies" among the interventions (DepEd, 2012, Encl. No. 1, item 5). No issuance says how to find the least mastered competencies or how to tell whether their reteaching worked.

The usual evidence, a competency's percent correct before and after reteaching, has a flaw that our department had never checked. A competency is put on the list because it scored lowest, and a score that is lowest on one test is partly low by chance: a hard item, a confusing one, a bad day. On any

second test, such scores tend to come back toward the average. Regression to the mean "can make natural variation in repeated data look like real change," and it grows "when follow-up measurements are only examined on a sub-sample selected using a baseline value" (Barnett et al., 2005, p. 215). When units are selected for their extreme scores, "they will often have less extreme scores on other variables, an occurrence that can be confused with a treatment effect" (Shadish et al., 2002, p. 55). A before-and-after on the least mastered competencies therefore shows a gain even if the reteaching did nothing.

The reteaching itself drew on three practices with research behind them: worked examples alternated with problems the learner solves (Pashler et al., 2007, p. 9), quizzes that make learners retrieve what they learned, with correct-answer feedback (Pashler et al., 2007, p. 21), a practice for which most of the 49 effect sizes in a review of 50 classroom experiments (57%) showed medium or large benefits (Agarwal et al., 2021, p. 1409), and explicit instruction with "models of proficient problem solving, verbalization of thought processes, guided practice, corrective feedback, and frequent cumulative review" (Gersten et al., 2009, p. 21). In a meta-analysis of 55 studies, learners who studied worked examples improved their mathematics performance by "almost half of a standard deviation (g = 0.48)" compared with those who did not (Barbieri et al., 2023, p. 23).

Whether short, spaced reteaching built on these practices raises the mastery of our own class's least mastered competencies, measured in a way that regression to the mean cannot fake, is the question of this study. It

addresses the Teaching and Learning theme of the Basic Education Research Agenda (DepEd, 2016).

Innovation, Intervention, and Strategy

The innovation was a routine for the least mastered competencies of one class, in four steps: review the test's items before listing the competencies; list the lowest competencies after the review; confirm the list and set a baseline on a parallel retest with new items; and reteach the competencies in two sets, one after the other, in short sessions spaced over three weeks.

Finding the Least Mastered Competencies

The Quarter 1 test had 40 multiple-choice items on the eleven competencies of the quarter that a written test can assess, three or four items each; the two drawing competencies and the financial plan were assessed by performance tasks (DepEd, 2023, p. 52). The item analysis computed each item's difficulty index, the share of learners who answered it right, and its discrimination index, the difference between the shares right in the upper and the lower 27% of the class by total score (Kelley, 1939, p. 20), with learners tied at the cut taken in code order, as the data set's `RANK` and `COUNTIF` formulas take them. Each competency's percent correct was the right answers on its items divided by the answers given.

Two items had negative discrimination indexes. In item 25, on the uses of rates, 39 of 44 learners chose C, the right answer, but the key sheet said B, which 2 chose; the key was wrong. In item 3, on angle pairs, 25 learners chose D and 15 chose the key, B: the item asked which word describes two angles that form a linear pair, and D, "adjacent," is as true as B,

"supplementary." An item in which a single distractor is chosen more often than all the other options, the answer included, calls for a check of the key (Kehoe, 1995, p. 3). Item 3's discrimination index depended on how the learners tied at the 27% cut were split, but its corrected point-biserial correlation, the item against the rest of the test, was −.11, which no split changes. The Head Teacher and the first researcher reviewed both items on August 26, 2025: item 25 was rekeyed to C, and both B and D were credited in item 3. The review raised the class's mean from 23.68 to 25.09 of 40 and the uses of rates from 53.0% to 81.1% correct.

The proposal had fixed the rule before the test: the six competencies with the lowest percent correct after the item review. No DepEd issuance sets a cut-off for a least mastered competency. Six was the number the class could reteach in two three-week phases, and with the review done, the six lowest were also the six below 60%.

The Parallel Retest

On August 28, 2025, before any reteaching, the class took a parallel retest: 44 new items, four on each of the eleven competencies, written by the Master Teacher from the same item specifications as the Quarter 1 test. It did two things. It showed how much the competencies' percent correct moved on a second test with no reteaching between, and it gave each of the six a baseline measured on items that had played no part in selecting it.

The Two Sets and the Lot

The six competencies were split into two sets of three. Competencies that build on each other were kept in the same set, so that reteaching one set

would not teach the other: the exterior angle of a polygon and the measures of a polygon's angles (LC 5 and 6), and percentage increase and decrease and money problems with percentages (LC 7 and 8). Four plans met that rule: two ways of placing problems involving rates (LC 11) and ordering rational numbers (LC 13) beside the two pairs, each in two orders. On August 27, 2025, the school head drew one of the four by lot: Set A, exterior and interior angles (LC 5), angles and sides of polygons (LC 6), and problems involving rates (LC 11), was retaught first; Set B, percentage increase and decrease (LC 7), money problems with percentages (LC 8), and ordering rational numbers (LC 13), second.

The Reteaching Sessions

The class met for Mathematics 45 minutes a day, five days a week (DepEd, 2024, para. 2). On Mondays, Wednesdays, and Fridays, the first 15 minutes of the period went to one competency of the set being retaught, the three in turn, so that each competency had three sessions a week apart. The rest of the period, and every Tuesday and Thursday, followed the Quarter 2 lessons, the same for both sets. Phase 1 ran from September 1 to 19, 2025 for Set A, and Phase 2 from September 24 to October 13, 2025 for Set B: nine sessions each.

The sessions were for the whole class. DepEd's policy on learning gaps in secondary schools asks that interventions be "tailored to individual learning needs" and does not recommend a whole-class approach "unless the deficiencies are common to all students" (DepEd, 2012, Encl. No. 1, item 4). These were not common to all: at the retest, 43% to 55% of the learners had

mastered each of the six. The study chose the whole class for two reasons: about half the class had missed each competency, and the comparison between the sets needed every learner taught the same way. The learners who had still not mastered a competency at Probe C were to get small-group sessions or a strategic intervention material on it after the study, so that the comparison was not disturbed.

Each session had five parts:

1. a retrieval check of two items on the competency, answered without notes before any explanation, the answers then shown;
2. a worked example, the teacher saying each step and why, built on the distractor the class had chosen most often on the Quarter 1 test (for percentage decrease, for example, the one that divides the decrease by the new price instead of the original);
3. a partly worked example that the learners finished in pairs;
4. two problems each learner solved alone; and
5. feedback on every learner's answers before the session ended.

Strategy and Evaluation Plan

The Head Teacher observed three sessions of each phase with a checklist of the five parts (Appendix D). Four criteria for success were set in the proposal: (a) in each phase, at least 80% of the learners attend at least 7 of the 9 sessions, and the observed sessions show at least 80% of the planned parts; (b) each probe has a KR-20 of at least .70; (c) in each phase, the set being retaught gains more than the other set, significant at the .05 level; and

(d) after its phase, at least 75% of the learners master each retaught competency, answering at least 3 of 4 of its probe items right.

Action Research Questions

Main question: Does short, spaced reteaching raise Grade 7 learners' mastery of their class's least mastered competencies more than ordinary instruction does?

1. Which competencies were least mastered on the Quarter 1 test, and how did the item review change the list?
2. How much did the six competencies' percent correct change from the Quarter 1 test to the parallel retest, with no reteaching between them?
3. In Phase 1, did the learners gain more on Set A, being retaught, than on Set B, waiting?
4. In Phase 2, did the learners gain more on Set B, being retaught, than on Set A, retaught before?
5. What share of the learners mastered each of the six competencies before and after its reteaching?

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

$H_0 1$: The learners' percent correct on the six competencies does not differ between the Quarter 1 test and the parallel retest.

$H_0 2$: In Phase 1, the learners' gain on Set A does not differ from their gain on Set B.

$H_0 3$: In Phase 2, the learners' gain on Set B does not differ from their change on Set A.

$H_0$4: For each competency, the share of learners who mastered it does not change during its phase of reteaching.

Action Research Methods

Research Design

The study used a switching-replications design across competencies within one class. When one set of competencies was retaught, the other served as its comparison; when the second set was retaught, the first served as "a continued-treatment control" (Shadish et al., 2002, p. 192). Comparing two sets of the same learners in the same weeks holds constant everything that changes over time for the whole class: the Quarter 2 lessons, the learners' growth, the tests' general difficulty, and the habit of taking probes. Both sets were chosen by the same rule from the same test, so regression to the mean affected both, and the difference between them is not its work; random allocation works this way, as it leaves the groups "equally affected by RTM" (Barnett et al., 2005, p. 217). The baseline of each phase was a test that had not chosen the competencies, which follows the advice to select on one measurement and examine change between the second and the third rather than the first and the third (Shadish et al., 2002, p. 59). A lot decided the sets and their order, and the second phase repeated the comparison with the roles switched, so that what made one set easier to raise could not explain both results.

Participants and Other Sources of Data and Information

The participants were the learners of one Grade 7 class of the first researcher in a public secondary school, School Year 2025–2026. All 46

enrolled learners were invited; 44 had their parents' consent and gave their own assent. One learner transferred to another school on September 12, 2025, before Probe B, and one was absent for Probe C and its make-up day. Table 1 shows how many were analyzed in each part.

Table 1

Learners Enrolled, Excluded, and Analyzed

Learners	Number
Enrolled	46
No parental consent	2
With consent: took the Quarter 1 test and the retest	44
Transferred before Probe B	1
Analyzed in Phase 1 (retest and Probe B)	43
Absent for Probe C and its make-up day	1
Analyzed in Phase 2 and in the mastery analysis (all three probes)	42

Note. Every learner took part in the reteaching as a class activity; consent decided only whose answers were recorded.

Sample Size

Because the whole class took part, the number of learners was fixed. A sensitivity analysis for a paired *t* test showed that the 44 learners with consent give a power of .80 to detect a standardized difference of $d_z = 0.43$ at $\alpha = .05$, two-tailed.

Other Sources

The Master Teacher in Mathematics, who taught other grades, wrote the items of the retest and the probes and assembled the probes. The school's research coordinator, the second researcher, kept the sealed probes, gave them with the first researcher, scored every answer sheet with the key, and kept the list that linked codes to learners. The Head Teacher observed the sessions.

Data Gathering Methods

The Tests

All four tests were multiple choice with four options. The Quarter 1 test (40 items) was the class's quarterly assessment. The retest (44 items, four on each of the eleven competencies) and Probes B and C (24 items each, four on each of the six competencies) were written for the study from the item specifications of the Quarter 1 test: for each competency, the same kinds of item with new numbers, contexts, and figures. Every item was new; no item appeared on two tests. The Master Teacher wrote all the items in July and August 2025, before the Quarter 1 test, for all eleven competencies, so the items for a competency did not depend on whether it would be on the list or in which set. On August 26, 2025, after the list was made and before the lot, the Master Teacher assembled Probes B and C for all six competencies, four items each, and the research coordinator sealed them. Scoring with the key needed no judgment. No item of the retest or the probes had a negative discrimination index.

The retest was given on August 28, 2025, Probe B on September 22, 2025, and Probe C on October 15, 2025, each with a make-up day for absent learners. A probe took about 30 minutes. The teacher saw a probe's items only when it was given; the answer sheets were collected at once and the items were not discussed with the class. The retest and the probes were not graded.

The outcome was each learner's percent correct on a set's items in each test. KR-20 (Kuder & Richardson, 1937) was .83 for the Quarter 1 test after the item review, .84 for the retest, .80 for Probe B, and .79 for Probe C; the

second criterion for success was met. A learner had mastered a competency on a probe when they answered at least 3 of 4 of its items right. Four items make that call rough: the KR-20 of a competency's four items ranged from .18 to .61.

Implementation Records

The first researcher kept a log of the sessions: the date, the competency, and the learners present. The Head Teacher observed three sessions of each phase (September 3, September 12, September 15, 2025; September 29, October 3, October 8, 2025) with the checklist of the five parts.

Storage

The test papers, the answer sheets, and the consent forms were kept in a locked cabinet in the school. The list linking codes to learners stayed with the research coordinator, apart from the data. The data set holds only codes, answers, and the sessions each learner attended.

Ethical Considerations

The school head approved the study in writing on August 8, 2025. On August 11, 2025, the adviser sent home a letter in Filipino and English on the purpose, the tests, the right to refuse or withdraw without any effect on the learner's grades, and how the answers would be kept; 44 of 46 parents signed the consent form by August 15, 2025. On August 18, 2025, the first researcher explained the study to the class, and each learner with consent signed an assent form. DepEd's quality checklist for completed research asks the researcher to "obtain assent from minor participants and consent from their parents and/or legal guardians" (DepEd, 2022, Annex 2, Item 13), and

the national ethical guidelines ask that "the consent of the parent or guardian must be obtained as well as the assent of the minor" (Philippine Health Research Ethics Board Ad Hoc Committee for Updating the National Ethical Guidelines [PHREB Ad Hoc Committee], 2022, p. 93). A learner's answers are personal information (Republic Act No. 10173, 2012, Sec. 3(g)), so the data were coded and the list of names kept apart. Every learner received the reteaching of both sets; the set retaught second waited three weeks, not the whole quarter. The retest and the probes did not enter the grades. The two faulty items were rescored for every learner's Quarter 1 grade, not only for the study. The study followed the Research Management Guidelines (DepEd, 2017). The researchers' use of an artificial intelligence (AI) tool is declared in Appendix E (DepEd, 2026).

Data Analysis

For the first question, the items' difficulty and discrimination indexes and each competency's percent correct were computed with the printed key and after the review. For the second, each learner's percent correct on the six competencies' items on the Quarter 1 test and on the retest was compared with a paired t test, and the same was done for the other five competencies; the competencies' percent correct on the two tests were correlated.

For the third and fourth questions, each learner's gain on each set from one test to the next was computed, and the gain on the set being retaught was compared with the gain on the other set with a paired t test, reported with its 95% confidence interval and d_z, the mean difference divided by the standard deviation of the differences. The Wilcoxon signed-rank test, which

uses only the ranks of the differences, checked each result, and the Shapiro–Wilk test (Shapiro & Wilk, 1965) checked the differences' normality. The paired test compares each learner with themselves, so the learner, not the competency, is the unit of analysis. A model over all three probes and every item checked both phases at once: a generalized estimating equation (GEE) for a right answer, with the competency, the test, and whether the competency had been retaught by that test as predictors, and an exchangeable correlation among each learner's answers (Liang & Zeger, 1986). The same model was also fitted to the retest and Probe B alone, Phase 1's comparison.

For the fifth question, the share of learners who mastered each competency at each probe was reported with its Wilson (1927) 95% confidence interval, for the 42 learners present at all three probes. The change during each competency's phase of reteaching was tested with the exact McNemar test (McNemar, 1947), and the six p values were adjusted with Holm's (1979) procedure.

The analyses were run in Python 3.12 with pandas, SciPy, and statsmodels (Seabold & Perktold, 2010), free software; the Excel data set computes as a formula every statistic Excel has a function for and labels Python's values for the rest (the GEE models, the model of the competency-by-probe cells, Shapiro–Wilk, the Wilcoxon p values, and the power analysis), and the notebook reproduces every result from the data (Appendix A).

Discussion of Results and Reflection

Implementation

In Phase 1, the learners attended a mean of 8.3 of the 9 sessions (range 6–9), and 38 of 43 (88.4%) attended at least 7. In Phase 2, the mean was 8.6 (range 7–9), and 42 of 42 attended at least 7. The six observed sessions showed 28 of the 30 planned parts (93.3%): one session in Phase 1 ran short and left out the problems done alone, and one in Phase 2 left out the pair work. The first criterion for success was met.

Question 1: The Least Mastered Competencies and the Item Review

Table 2 shows each competency's percent correct on the Quarter 1 test with the printed key and after the review. With the printed key, the six lowest included the uses of rates (LC 10), at 53.0%, held down by its miskeyed item; after the review it stood at 81.1%, and ordering rational numbers (LC 13), at 59.8%, took its place. A teacher who had acted on the printed key would have retaught a competency the class had learned. A 75% cut-off would have listed 10 of the eleven competencies with the printed key and 8 after the review.

Table 2

Percent Correct by Competency on the Quarter 1 Test, Before and After the Item Review

Competency	Items	Printed key (%)	After the review (%)	Rank after the review
LC 3 Angle pairs	4	67.6	81.8	11
LC 4 Classifying polygons	3	62.9	62.9	7
LC 5 Exterior and interior angles	4	52.3	52.3	2
LC 6 Angles and sides of polygons	4	53.4	53.4	3
LC 7 Percentage increase and decrease	4	47.7	47.7	1

Competency	Items	Printed key (%)	After the review (%)	Rank after the review
LC 8 Money problems with percentages	4	55.1	55.1	5
LC 10 Uses of rates	3	53.0	81.1	10
LC 11 Problems involving rates	4	54.5	54.5	4
LC 12 Fractions, decimals and percentages	3	79.5	79.5	9
LC 13 Ordering rational numbers	3	59.8	59.8	6
LC 14 Operations on rational numbers	4	69.9	69.9	8

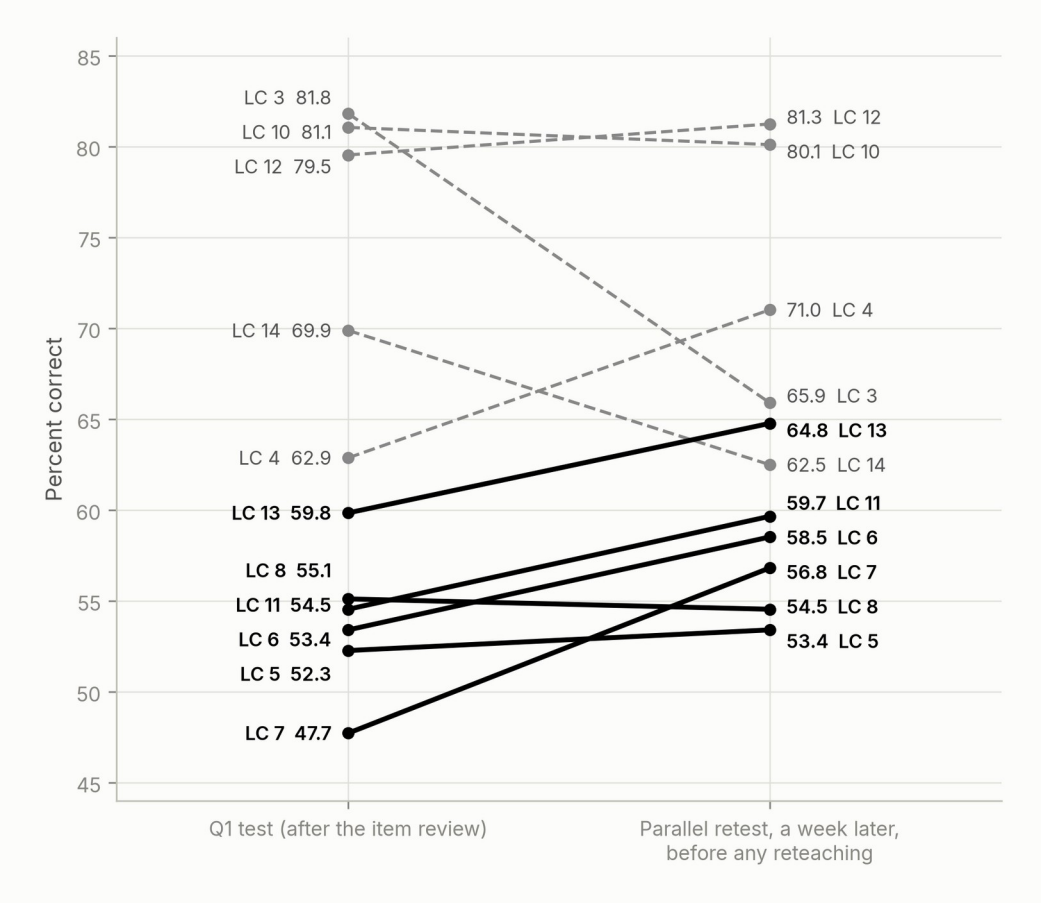
Note. 44 learners. LC: the learning competency's number in the MATATAG curriculum guide, Grade 7, Quarter 1 (DepEd, 2023, p. 52); the names are short forms of DepEd's wording. Rank 1 is the lowest. The six lowest after the review, ranks 1 to 6, were the least mastered competencies of the study.

Question 2: The Retest, Before Any Reteaching

From the Quarter 1 test to the retest a week later, with no reteaching between, each learner's percent correct on the six competencies' items rose from 53.6 to 58.0, a mean change of 4.4 points, 95% CI [0.2, 8.6], $t(43) = 2.10$, $p = .042$, $d_z = 0.32$. $H_0 1$ is rejected. On the other five competencies, it fell from 75.1 to 72.2, a mean change of −3.0 points (computed before rounding), 95% CI [−6.7, 0.8], $p = .116$ (Figure 1). The competencies' percent correct on the two tests correlated, $r = .82$, but not perfectly, and regression to the mean comes from that imperfect correlation (Shadish et al., 2002, p. 58). Of the six, five rose, and angle pairs (LC 3), the highest on the Quarter 1 test at 81.8%, scored 65.9% on the retest. The rise of the six was not learning: no lesson in that week touched them. It is the amount a before-and-after comparison from the Quarter 1 test would have credited to any reteaching.

Figure 1

Each Competency's Percent Correct on the Quarter 1 Test and the Parallel Retest



Note. The six least mastered competencies in bold lines; the other five in dashed lines. The retest had four new items on each competency and was given a week after the Quarter 1 test, before any reteaching.

Question 3: Phase 1

In Phase 1, the learners' percent correct on Set A, being retaught, rose by 20.0 points (SD = 20.1), and on Set B, waiting, by 3.5 points (SD = 19.7), 95% CI [−2.6, 9.5]. Set A gained 16.5 points more, 95% CI [7.2, 25.7], $t(42) = 3.60$, $p < .001$, $d_z = 0.55$; 29 of the 43 learners gained more on Set A than on Set B. $H_0 2$ is rejected. The Wilcoxon signed-rank test agreed, $p = .002$, and the Shapiro–Wilk test did not reject normality of the differences, $W = .96$, $p = .184$ (Table 3 and Figure 2).

Table 3

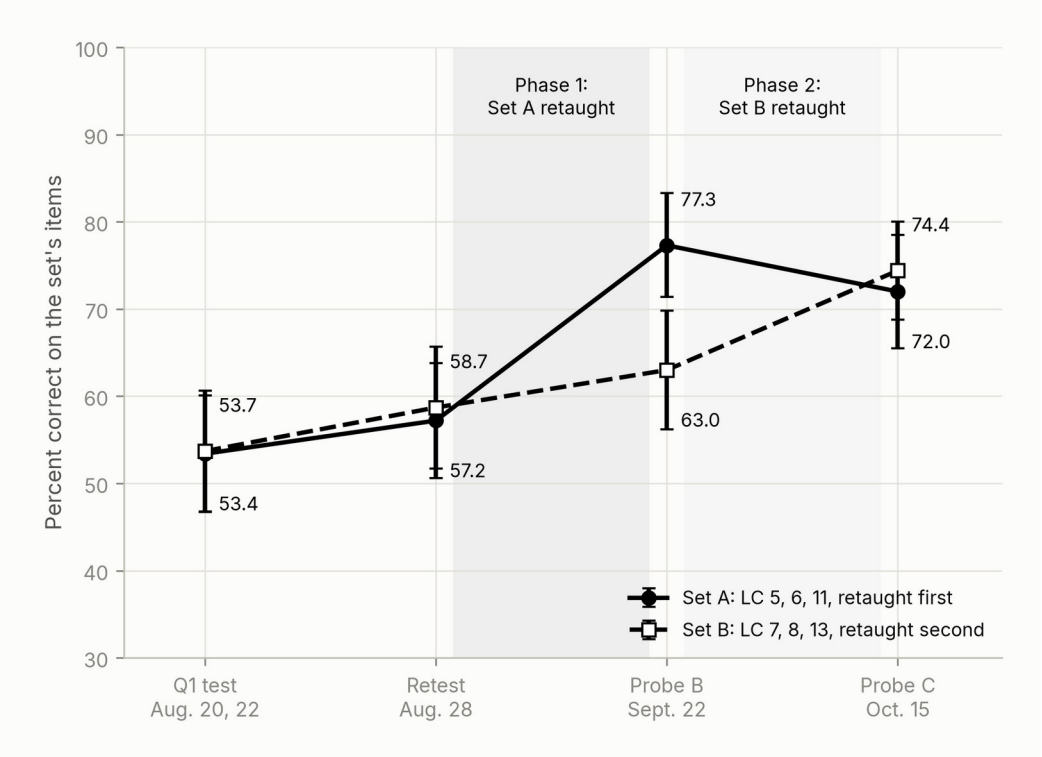
Percent Correct on Each Set at the Four Tests

Set	Quarter 1 test *M* (*SD*)	Retest *M* (*SD*)	Probe B *M* (*SD*)	Probe C *M* (*SD*)
Set A (LC 5, 6, 11), retaught in Phase 1	53.4 (22.0)	57.2 (21.7)	77.3 (19.4)	72.0 (20.9)
Set B (LC 7, 8, 13), retaught in Phase 2	53.7 (22.7)	58.7 (22.9)	63.0 (22.0)	74.4 (17.9)

Note. Learners present: 44 at the Quarter 1 test and the retest, 43 at Probe B, and 42 at Probe C. Phase 1 ran between the retest and Probe B, Phase 2 between Probes B and C. The gains in the text are for the learners present at both tests of a phase (43 in Phase 1, 42 in Phase 2), so they differ slightly from the differences between these columns.

Figure 2

Percent Correct on Each Set at the Four Tests



Note. Means with 95% confidence intervals. Filled circles, Set A; open squares, Set B. The shaded weeks are the two phases.

Question 4: Phase 2

In Phase 2, the learners' percent correct on Set B, now being retaught, rose by 11.1 points (*SD* = 23.0), while on Set A it fell by 5.2 points (*SD* = 15.8). Set B gained 16.3 points more, 95% CI [7.5, 25.0], $t(41) = 3.75$, $p < .001$, d_z =

0.58. $H_0 3$ is rejected, and the third criterion for success was met in both phases. In this phase the other set had been retaught before, not waiting its turn, and 5.2 of the 16.3 points are its fall. The Wilcoxon test agreed, $p = .001$, and the Shapiro–Wilk test did not reject normality, $W = .96$, $p = .210$. Set A's fall in the three weeks after its reteaching, −5.2 points, 95% CI [−10.1, −0.2], $p = .041$, shows that part of what the reteaching built did not last without review.

The model over all three probes and all 3,096 answers of 44 learners agreed: a competency that had been retaught was answered right with 2.18 times the odds of one that had not, 95% CI [1.51, 3.15], $z = 4.17$, $p < .001$, with the competency and the test held constant. By Probe C all six had been retaught, so the model compares the sets at Probe B against their difference at the retest and at Probe C. Set A's fall after its phase therefore counts toward the effect, as it does in the Phase 2 comparison. Fitted to Phase 1 alone, the odds ratio was 2.12, 95% CI [1.42, 3.17], $p < .001$.

Question 5: Mastery, Competency by Competency

Table 4 shows the share of the 42 learners present at all three probes who mastered each competency. During its phase, mastery rose for all six. The rise was significant after Holm's adjustment for exterior and interior angles (LC 5) and angles and sides of polygons (LC 6), and $H_0 4$ is rejected for those two; for the other four, rises of 9.5 to 16.7 points were not significant. After its phase, exterior and interior angles (LC 5), angles and sides of polygons (LC 6), and ordering rational numbers (LC 13) reached 75% mastery; the fourth criterion for success was met for three of the six competencies, and

not for percentage increase and decrease (LC 7), money problems with percentages (LC 8), and problems involving rates (LC 11). Ordering rational numbers (LC 13) also rose while it waited, from 52.4% to 71.4% (p = .077, unadjusted), in the weeks the Quarter 2 lessons placed irrational numbers on the number line.

Table 4

Learners Who Mastered Each Competency at the Three Probes

Competency	Set	Retest % [95% CI]	Probe B % [95% CI]	Probe C % [95% CI]	During its phase: gained / lost	Holm p
LC 5 Exterior and interior angles	A	42.9 [29.1, 57.8]	83.3 [69.4, 91.7]	73.8 [58.9, 84.7]	18 / 1	< .001
LC 6 Angles and sides of polygons	A	42.9 [29.1, 57.8]	78.6 [64.1, 88.3]	57.1 [42.2, 70.9]	18 / 3	.007
LC 11 Problems involving rates	A	50.0 [35.5, 64.5]	61.9 [46.8, 75.0]	69.0 [54.0, 80.9]	11 / 6	.801
LC 7 Percentage increase and decrease	B	45.2 [31.2, 60.1]	54.8 [39.9, 68.8]	64.3 [49.2, 77.0]	10 / 6	.801
LC 8 Money problems with percentages	B	54.8 [39.9, 68.8]	47.6 [33.4, 62.3]	59.5 [44.5, 73.0]	9 / 4	.801
LC 13 Ordering rational numbers	B	52.4 [37.7, 66.6]	71.4 [56.4, 82.8]	88.1 [75.0, 94.8]	11 / 4	.474

Note. 42 learners. Mastered: at least 3 of 4 items right on the competency's probe items. Gained / lost: learners who moved from not mastered to mastered, and back, from the probe before the competency's phase to the probe after it (Set A: retest to Probe B; Set B: Probe B to Probe C). Exact McNemar tests, Holm-adjusted over the six.

Reflection

What We Found

Short, spaced reteaching raised the class's percent correct on the competencies being retaught by 16.5 points more than ordinary instruction raised the competencies waiting their turn in Phase 1. In Phase 2 the newly retaught set gained 16.3 points more than the set retaught before, which fell

5.2 points in those weeks. Over the three probes, the retaught competencies were answered right with 2.18 times the odds of the others, and with 2.12 times in Phase 1 alone. The two competencies on polygons' angles each moved from 42.9% of the learners mastering them to 83.3% and 78.6%.

How a Before-and-After Would Have Read

Had we compared the six competencies on the Quarter 1 test with the same six on Probe C, as our department's usual report does, we would have reported a rise from 54.3% to 73.2%, 18.9 points, $t(41) = 9.88$, $p < .001$, with 41 of 42 learners improving. Of those 18.9 points, 4.9 (26%), for the same 42 learners, had appeared on the retest, before any reteaching. The rest mixes the reteaching with the class's ordinary growth across the eight weeks after the Quarter 1 test. Only the comparison between the sets separates them.

How It Compares With the Research

Our differences, $d_z = 0.55$ and 0.58, sit near the average effect of worked examples on mathematics performance, $g = 0.48$ (Barbieri et al., 2023, p. 23), but the two are not the same measure: theirs compares groups of learners, ours two sets of competencies within the same learners. Our sessions combined several practices, so the study cannot say which part did the work.

What Else Could Explain the Difference

The two sets differed in content, and one set may be easier to raise than the other; the switch of roles in Phase 2 answers this, since each set gained more in the phase it was retaught. The probes' items, though written to the same specifications, cannot be equally hard on every form, and a harder form for one set in one phase would move that phase's difference. The first

researcher both taught and gave the probes, though not the items, which a teacher who did not know the order wrote before the Quarter 1 test and placed in the probes before the lot. The Quarter 2 lessons touched the waiting competencies in places, as they did ordering rational numbers in Phase 1; such overlap makes the difference between the sets smaller, not larger. The class is one class, taught by one teacher.

Limits

In Phase 2 the comparison set had been retaught before, so the Phase 2 difference, and the model over all three probes, include Set A's fall after its phase; the Phase 1 comparison does not. The intervals count the variation among learners, not among items, since each set had one form per probe; a regression of the logit of percent correct in the 18 competency-by-probe cells, where a form's difficulty counts as error, still found the effect, odds ratio 2.26, $p = .013$. Four items on a competency make a single learner's mastery a rough call: the four items' KR-20 was as low as .18, so the mastery shares describe the class better than any one learner. The study lasted eight weeks after the Quarter 1 test and cannot show whether the gains lasted to the end of the school year; Set A's fall in the three weeks after its phase suggests they would not without review.

What We Would Change

We would return to each set's competencies in the weeks after its phase, as cumulative review (Gersten et al., 2009, p. 21) and spacing (Pashler et al., 2007, p. 5) advise, and give the percentage problems (LC 7 and 8) and

problems involving rates (LC 11) more sessions or a strategic intervention material for the learners who had not mastered them.

Conclusion

Spaced reteaching of the least mastered competencies, in 15-minute sessions with a retrieval check, worked examples, and practice with feedback, raised the class's percent correct on the competencies it reached in both phases: against competencies that waited their turn in Phase 1, and against the set retaught before in Phase 2. The list itself needed two checks before reteaching: an item review, which removed a competency the class had learned, and a retest, which showed how much of a before-and-after gain would have been regression to the mean.

Action Plan

Table 5

Action Plan

Objective	Activities	Persons involved	Timeline	Resources (PhP)	Success indicator
Keep the review and the retest in the department's routine	Review items with negative discrimination before listing; confirm the list on a parallel retest before reteaching	Mathematics teachers, Head Teacher	From June 2026	0	Every class's list made after the item review
Build the item bank the routine needs	Four items for every competency of each quarter, written to the item specifications, two forms each	Master Teacher, Mathematics teachers	April to May 2026	600 (printing)	Two forms ready for every Grade 7 competency
Share the routine	A Learning Action Cell session on the item review, the retest, and the two-set reteaching, with this study's data	Researchers, Mathematics department, school head	March 2026	0	Each Mathematics teacher runs the routine with one class
Hold the gains	Review each retaught set in the three weeks after its phase	Mathematics teachers	From June 2026	0	Retaught competencies keep their probe scores
Test the routine across the Grade 7 classes	Proposal for a study in the school's Grade 7 classes, each with its own list and its own lot	Researchers, school research committee	Next call for proposals	Up to 30,000 (BERF, school level)	Proposal submitted and approved

References

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

Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K. (2023). A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review, 35*(1), Article 11. https://doi.org/10.1007/s10648-023-09745-1

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology, 34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Department of Education. (2012). *Policy guidelines on addressing learning gaps and implementing a reading and writing program in secondary schools effective school year (SY) 2012–2013* (DepEd Order No. 39, s. 2012). https://www.deped.gov.ph/2012/05/11/do-39-s-2012-policy-guidelines-on-addressing-learning-gaps-and-implementing-a-reading-and-writing-program-in-secondary-schools-effective-school-year-sy-2012-2013/

Department of Education. (2015). *Policy guidelines on classroom assessment for the K to 12 Basic Education Program* (DepEd Order No. 8, s. 2015). https://www.deped.gov.ph/2015/04/01/do-8-s-2015-policy-guidelines-on-classroom-assessment-for-the-k-to-12-basic-education-program/

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. (2023). *MATATAG curriculum: Mathematics (Grades 1–10)*. https://www.deped.gov.ph/wp-content/uploads/9-FINAL-MATATAG-Mathematics-CG-2023-Grades-1-10.pdf

Department of Education. (2024). *Amendment to DepEd Order No. 010, s. 2024 (Policy guidelines on the implementation of the MATATAG curriculum)* (DepEd Order No. 012, s. 2024). https://www.deped.gov.ph/2024/09/18/september-18-2024-do-012-s-2024-amendment-to-deped-order-no-010-s-2024-policy-guidelines-on-the-implementation-of-the-matatag-curriculum/

Department of Education. (2026). *Foundational guidelines on artificial intelligence 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/

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED504995.pdf

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70. https://www.jstor.org/stable/4615733

Kehoe, J. (1995). *Basic item analysis for multiple-choice tests* (ERIC/AE Digest). ERIC Clearinghouse on Assessment and Evaluation. https://files.eric.ed.gov/fulltext/ED398237.pdf

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*(1), 17–24. https://doi.org/10.1037/h0057123

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

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153–157. https://doi.org/10.1007/BF02295996

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007-2004). National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED498555.pdf

Philippine Health Research Ethics Board Ad Hoc Committee for Updating the National Ethical Guidelines. (2022). *National ethical guidelines for research involving human participants 2022*. Philippine Council for Health Research and Development. https://www.pchrd.dost.gov.ph/wp-content/uploads/2023/05/2022-NEGRIHP_Official-Gazatte_Ver-5-1.pdf

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

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

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

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association, 22*(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

Financial Report

Table 6

Statement of Expenses

Date	Particulars	Receipt no.	Amount (PhP)	Category
August 9, 2025	Consent letters in Filipino and English and assent forms, 100 copies	OR 3112	150	Reproduction
August 27, 2025	Retest, 44 items, 50 copies of 4 pages	OR 3175	300	Reproduction
August 27, 2025	Answer sheets for the retest and the two probes, 150 copies	OR 3176	150	Reproduction
September 19, 2025	Probe B, 24 items, 50 copies of 3 pages	OR 3241	225	Reproduction
October 13, 2025	Probe C, 24 items, 50 copies of 3 pages	OR 3302	225	Reproduction
August 29, 2025	Worked example and partly worked example sheets for the 18 sessions	OR 3190	540	Reproduction
March 6, 2026	Printing and binding of the report, 3 copies	OR 3518	780	Reproduction
Total			**2,370**	

Note. Fund source: the researchers' own funds. No school funds or grants were used. Receipt numbers are invented for this sample.

Appendices

Appendix A. Data Set and Notebook

The Excel data set (every learner's answer to every item of the four tests, the keys before and after the review, the sessions each learner attended, the observation checklists, every statistic Excel can compute as a formula, and Python's values for the rest, labeled) and the Jupyter notebook that reproduces every result from the data.

Appendix B. Consent Letter and Assent Form

Kept with the school's copy, with the signed forms, and left out of the archived report.

Appendix C. Item Specifications, the Retest, and the Probes

The specification of each competency's items, and the retest and the two probes with their keys. Kept with the school's copy so that the items can be used again.

Appendix D. Session Plan and Observation Checklist

The five parts of a reteaching session, a sample session for each of the six competencies, and the Head Teacher's checklist.

Appendix E. 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, codes, scores, or other personal information were entered into the AI tool (adapted from Annex A of DepEd Order No. 003, s. 2026).