# Response Cards and Learning in Large Grade 8 Araling Panlipunan Classes

An Action Research

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

Conducted November 2025 to January 2026

**Abstract**

This action research tested whether write-on response cards would raise what Grade 8 learners kept from an Araling Panlipunan lesson to the next day, in classes of about 60 in a double-shift junior high school. In usual practice, a question asked of 60 learners is answered by a volunteer or in chorus, and the teacher cannot tell who understood. With the cards, every learner wrote an answer and held it up on a cue, the teacher counted the wrong cards, and a rule fixed in advance decided when to reteach. The study used a randomized alternating treatments design replicated across the teacher's four sections (57 to 63 learners each, 228 with consent): each of 16 lessons of Quarter 3, School Year 2025–2026, was taught in all four sections, with the same plan and the same nine questions, and a lot drawn before the unit gave the cards to two sections in each lesson. A 5-item quiz opened the next meeting. Over 3,237 quizzes, learners scored 6.2 percentage points higher after lessons with cards, 95% CI [4.4, 7.9], $p < .001$, in a mixed model with lesson and section effects; a randomization test that redrew the lot 1,679,616 ways agreed, $p < .001$. On the Quarter 3 test eight weeks later, items from card lessons were answered 2.0 points better, 95% CI [−0.2, 4.2], Holm-adjusted $p = .077$. Among learners drawn by lot for observation, 100.0% answered at least once in card lessons against 12.5% in usual lessons, and the question rounds took 3.1 more minutes of the 40-minute period. Response cards are recommended for large classes, with the time they take planned into the lesson.

*Keywords:* response cards, large classes, double shift, active student response, Araling Panlipunan, alternating treatments design

**Acknowledgment**

The researchers thank the school head, who approved the study and drew the lot; the Master Teacher in Araling Panlipunan, who wrote the quiz items before the lot without learning it; the Araling Panlipunan teacher of the afternoon shift, who observed 32 periods, and the department's Head Teacher, who observed eight of them a second time; the four class advisers, who sent home and collected the consent forms and witnessed the learners' assent; and the 239 learners of Grade 8 Narra, Molave, Yakal, and Kamagong and their parents. The study was funded by the Basic Education Research Fund (school level).

# Response Cards and Learning in Large Grade 8 Araling Panlipunan Classes

## Context and Rationale

Our school holds classes in two shifts because it has too few rooms. In School Year 2025–2026, 2,986 learners in 50 sections shared 25 instructional rooms, according to the Learner Information System and our facility records; the Grade 8 classes met in the morning shift, from 6:00 a.m. to 12:00 noon, and the 711 Grade 8 learners filled 12 sections. The first researcher's four sections enrolled 57 to 63 learners each. DepEd's guidelines on the opening of classes set a range of 15 to 60 learners a class for Grade 5 to high school (Department of Education [DepEd], 2012, item 3.g), and two of the four sections were above it. To fit two shifts into a day, the school's class program, approved by the Schools Division Superintendent, gave Araling Panlipunan 40 minutes a day, five days a week: the 200 minutes a week that DepEd allows as the least for the subject when a school proposes its own time allotment (DepEd, 2024b, para. 2.b).

The records showed a teaching problem as well as a crowded one. On the Quarter 1 test of August 2025, the 228 learners of the four sections who later joined this study had a mean percentage score of 49.28, and 22 of the 40 items were answered right by fewer than half of them. In class, the same lessons had seemed understood. When the department's Head Teacher tallied who answered aloud in four of our periods in September 2025, one in each section, 34 of the 41 questions were answered by one learner and 7 in chorus;

the 34 answers came from 17 of the 232 learners present, and 9 of them gave 25. The other 215 gave no answer the teacher could hear alone.

This is the usual pattern of questioning a whole class. "It is also generally the case that only a few pupils in a class answer the teacher's questions," and the teacher who accepts answers from a few "can keep the lesson going but is actually out of touch with the understanding of most of the class" (Black & Wiliam, 1998, pp. 143–144). Calling on one learner "often results in more frequent responses by high-achieving students and few or no responses by low-achieving students" (Heward et al., 1996, p. 5). Class size makes it worse: in observations of 686 pupils, "Classroom engagement decreased in larger classes," and at secondary level "this was particularly marked for lower attaining pupils" (Blatchford et al., 2011, p. 715). Rosenshine (2012) lists among the principles of effective instruction to "Ask a large number of questions and check the responses of all students," for example by having all students "Write the answer on a card and then hold it up" (pp. 12, 14).

Response cards do this. In an early study, fourth graders in a social studies class held up an answer on a white board after the cue "Hold up your cards," and "Most students scored higher on daily quizzes following sessions in which response cards were used than they did on quizzes that followed hand-raising sessions" (Narayan et al., 1990, pp. 483, 486). With fifth graders in science, active responding was "14 times higher with response cards than with hand raising," and all 22 learners scored higher on next-day quizzes after response-card lessons (Gardner et al., 1994, p. 63). In a meta-analysis of 18

studies, learners "performed 1.08 SD higher on quizzes and 0.38 SD higher on tests in the response card condition than in the hand-raising condition" (Randolph, 2007, p. 125), and write-on cards of "white card stock covered by a plastic sheet protector" with a dry-erase marker were among those tested (Swanson, 1999, as cited in Randolph, 2007, p. 116). The classes in these studies were small: 20 learners (Narayan et al., 1990), 22 (Gardner et al., 1994), 23 in a ninth-grade class (Cavanaugh et al., 1996), and 27 in a university course (Marmolejo et al., 2004); the largest classes in Randolph's review were university classes of 47 and 52 (Randolph, 2007, p. 120). In Narayan et al.'s study the teacher scanned 20 cards with no difficulty, and the authors wrote that "there is no doubt a limit to the size of the class in which response cards could be used effectively," though "it is likely that response cards could be used with somewhat larger classes as well" (Narayan et al., 1990, p. 489). Whether a teacher can scan 60 cards, decide, and still finish a 40-minute lesson is the question our classes raise.

The cards also answer a question of resources. Our school had no devices on which every learner in a section could answer, and no budget for them; a card made of a sheet protector, a piece of white cartolina, a marker, and a cloth costs a few pesos and can be used again. This study addresses the Basic Education Research Agenda's theme of teaching and learning, whose topics on instruction include "Class size," "Contact time," and "Materials and resources," and its topic of "Classroom assessment (i.e. formative and summative)" (DepEd, 2016).

## Innovation, Intervention, and Strategy

The innovation was a routine of write-on response cards for classes of about 60, built so that a teacher could check every learner's answer within a short period and decide by a rule, not by impression, whether to reteach.

### The Cards

Each learner had a card: a quarter sheet of white cartolina, trimmed to A4 to fit inside a clear A4 sheet protector, a whiteboard marker, and a square of cotton cloth to erase with. The cards were kept in one plastic box for each section and given out by row at the start of a period, which took about a minute. The 260 cards, one for each of the 239 learners enrolled and spares, cost PhP 7,240.00 in all, including a box of markers bought in the third week to replace markers that had dried out: PhP 30.29 for each learner, or PhP 3.79 for each learner for each of the eight lessons that used them. Markers were 67% of the cost.

### The Routine

The teacher asked nine questions in each lesson, in three rounds of three, at points fixed in the lesson plan: after the opening, in the middle, and before the closing. Each question had an answer of one to three words or a letter. For each question:

1. the teacher asked the question and said "Isulat." (Write.);
2. after about ten seconds, the teacher said "Itaas!" (Hold it up!), and every learner held up the card, answer facing front;
3. the teacher scanned the room row by row and counted the cards that were wrong or blank;

4. if 15 or more cards were wrong or blank, about a fourth of a class of 60, the teacher reteached the point with a new example and asked a parallel question; if fewer, the teacher gave the right answer with a one-sentence reason and went on; and
5. the teacher said "Ibaba at burahin." (Put it down and erase.)

The rule was the study's own. Its threshold follows Heward et al. (1996), who advise that when "a fourth or more of the class" answers wrong, the teacher "state or display the correct answer and immediately repeat the same question or item" (p. 9). Instead of repeating the item, the rule reteaches the point, as the instructor in Marmolejo et al. (2004) reviewed the material when fewer than 75% of the class answered correctly (p. 407). The teacher wrote the count and the decision for each question on a tally sheet (Appendix D).

**The Comparison**

In lessons without cards, the teacher followed the usual practice of our department: the same lesson plan and the same nine questions at the same points, answered by a volunteer the teacher called on, by a learner called by name, or by the class in chorus, with the teacher reteaching when the answer heard was wrong or the chorus was weak. The only difference between the two kinds of lessons was the cards and the rule.

**Strategy and Evaluation Plan**

Four criteria for success were set in the proposal: (a) the routine is run as planned, with the rule followed on at least 90% of the card questions and all nine planned questions asked in at least 90% of the observed periods; (b) the two observers agree, with the lower bound of the 95% confidence interval

of their intraclass correlation at least .75, "good" reliability (Koo & Li, 2016, p. 158); (c) learners score higher on the next day's quiz after lessons with cards, significant at the .05 level both in the mixed model and in the randomization test; and (d) the question rounds with cards take no more than four minutes more of a period than the usual rounds.

## Action Research Questions

Main question: Do write-on response cards, used with a rule for reteaching, raise what Grade 8 learners in classes of about 60 keep from an Araling Panlipunan lesson, compared with the usual oral questioning?

1. How did the learners score on the next day's quiz after lessons with response cards and after lessons without them?
2. Did the difference remain on the Quarter 3 test, eight weeks after the unit?
3. How many times did each observed learner answer in a period, and how many answered at least once, with and without the cards?
4. How many minutes of the 40-minute period did the question rounds take with and without the cards?
5. Did learners in the lower half of their section by the Quarter 2 grade gain more from the cards than learners in the upper half?

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

$H_0$1: The learners' next-day quiz scores do not differ between lessons with and without response cards.

$H_0$2: The learners' scores on the Quarter 3 test items do not differ between the lessons taught with and without response cards.

$H_0$3: The number of answers each observed learner gave in a period does not differ between periods with and without response cards.

$H_0$4: The share of observed learners who answered at least once does not differ between periods with and without response cards.

$H_0$5: The minutes spent on question rounds do not differ between periods with and without response cards.

$H_0$6: The difference that response cards make to next-day quiz scores does not differ between the lower and the upper half of the learners by Quarter 2 grade.

## Action Research Methods

### Research Design

The study used a randomized alternating treatments design replicated across the first researcher's four Grade 8 sections. An alternating treatments design compares two conditions by the "fast alternation of two different treatments or conditions" in the same learners (Barlow & Hayes, 1979, p. 200); because both conditions run in the same weeks, it controls most threats to internal validity (p. 203). Each of the 16 lessons of the unit was taught in all four sections on the same day, with the same plan and the same questions. The lessons were grouped in 8 blocks of two, Monday and Tuesday, and Wednesday and Thursday, of each week. For each block, a lot gave the cards to two of the four sections in the first lesson and to the other two in the second, so that every lesson had two sections with cards and two without, and every section used the cards in one lesson of each block, eight lessons in all. This is a blocked randomization of the kind recommended for alternating

treatments, in which "conditions of the same type (A or B) cannot appear in more than two consecutive time periods" (Kratochwill & Levin, 2010, pp. 133, 136). Comparing sections in the same lesson holds the lesson's difficulty constant; comparing lessons in the same section holds the section constant; and the lot, not the teacher, decided which section used the cards in which lesson.

The unit was the first four learning competencies of Quarter 3 in the MATATAG curriculum for Grade 8, which our Grade 8 classes followed from School Year 2025–2026 (DepEd, 2024a, para. 8): World War I, the world after the war, totalitarianism as a threat to democratic government, and World War II (DepEd, 2023, p. 55), four lessons each, from November 3 to 27, 2025. The lessons were taught in Filipino, the medium of instruction for Araling Panlipunan in Grades 7 to 10 (DepEd, 2024a, Annex 2, Table 4).

On October 24, 2025, after the quiz items were written and sealed, the school head drew one of six slips from a box, one slip for each pair of sections (Narra and Molave, Narra and Yakal, and so on), for each of the 8 blocks, putting the slip back before the next draw. The research coordinator recorded the schedule. There were 1,679,616 possible lots, 6 for each of the 8 blocks.

**Participants and Other Sources of Data and Information**

The participants were the learners of the first researcher's four Grade 8 sections in a public junior high school on double shift, School Year 2025–2026. All 239 enrolled learners were invited; the parents of 228 consented and these learners gave their assent. Table 1 shows the flow. Every learner

took part in every lesson, with or without cards as the lot decided; consent decided only whose quizzes and answers were recorded.

**Table 1**

*Learners Enrolled, Consenting, and Analyzed, by Section*

| Section | Enrolled | Consented | Analyzed |
|---|---|---|---|
| Narra | 61 | 58 | 58 |
| Molave | 58 | 56 | 56 |
| Yakal | 63 | 59 | 59 |
| Kamagong | 57 | 55 | 55 |
| Total | 239 | 228 | 228 |

*Note.* Of the 11 learners not in the data, the parents of 7 declined and the forms of 4 were not returned. One learner with consent transferred to another school after the ninth lesson; that learner's quizzes up to then are analyzed. Analyzed: took at least one quiz after attending its lesson.

### *Sample Size*

The number of learners was fixed by the four sections. Before the study, a simulation of 200 studies with this design, using figures from our Quarter 2 records (a quiz-to-quiz standard deviation of about 25 points, of which about 15 was the learner's own level, and about 3 points of chance in how a lesson went in one section) and 6% of quizzes missed, gave a power of .995 for the mixed model and .995 for the randomization test to detect a difference of 5 percentage points, a quarter of one item on a five-item quiz, at α = .05. Four sections, each answering sixteen quizzes, give many quizzes even when the learners are few. The same simulation gave a power of .825 to detect a difference of 4 points between the lower and the upper half of the learners: 5 points for one half against 1 for the other would likely be found, a smaller difference likely not.

### *Other Sources*

The Master Teacher in Araling Panlipunan, who taught Grade 10 in the afternoon shift, wrote the quiz items and did not learn the lot. An Araling Panlipunan teacher of the afternoon shift observed one card period and one usual period in every lesson, and the department's Head Teacher observed eight of these periods a second time, on four days, independently. The research coordinator, the second researcher, drew the lot for the observed periods, kept the sealed quizzes, scored them, and kept the list that linked codes to learners.

## Data Gathering Methods

### *Next-Day Quizzes*

Each lesson had a quiz of 5 multiple-choice items with four options, in Filipino, written to the lesson's objectives by the Master Teacher before the lot and sealed by the research coordinator on October 22, 2025. The quiz opened the next meeting of the class and took five minutes; Thursday's quiz opened Friday's period, which then went to written work. The teacher saw each quiz only when it was given, and the papers were collected at once. The research coordinator scored them with the key; scoring a letter against a key needs no judgment, so knowing the lot could not change a score. The outcome was the percent of the five items a learner answered right. A single five-item quiz measures one learner roughly, with a KR-20 (Kuder & Richardson, 1937) from .34 to .52; for the 94 learners who took all 16 quizzes, the KR-20 of the 80 items was .91. The analysis therefore pools all the quizzes. The quizzes were formative assessment: under DepEd's classroom assessment policy for that

school year, teachers keep a record of formative results, but "this should not be used as bases for grading" (DepEd, 2015, Sec. IV).

### *The Quarter 3 Test*

The Quarter 3 test of January 22 and 23, 2026, the department's usual test of 50 items, had two items on each lesson of the unit; the other 18 covered the quarter's later weeks. Every lesson was taught with cards in two sections and without in two, so an item's difficulty weighed on both conditions alike. The Quarter 3 test's KR-20 was .88.

### *Observation*

Before each observed period, the observer drew eight class numbers by lot from the learners with consent who were present, and counted, for each, the answers the learner gave alone that the teacher could see or hear: a card held up with an answer, or an answer spoken when the learner was called on. A card held up blank was not counted, and chorus answers were not counted for anyone, since no one can tell who said what in a chorus of 60. With a stopwatch, the observer timed each question round from the teacher's first question to the end of its feedback, and in usual periods counted the hands raised for each question. Both observers practiced the counts in two Grade 7 periods before the study.

### *The Teacher's Tally*

In card periods the teacher wrote, for each question, the number of cards held up with an answer, the number wrong or blank, and the decision; in usual periods, whether one learner or the class in chorus answered, and whether the point was retaught. The tally counted the room, with no names.

### *Storage*

The quiz papers, the observation forms, 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 codes, sections, the answers to the quizzes and to the Quarter 1 and Quarter 3 tests, attendance, a lower-half flag from the Quarter 2 grades, and the observers' counts, all by code; it holds no names, birthdays, or grades.

## Ethical Considerations

The Schools Division Research Committee approved the proposal on October 3, 2025 and endorsed it to the Regional Office, which confirmed it on October 8, 2025 (DepEd, 2017, Sec. V.A); the memorandum of agreement for the grant was signed on October 9, 2025. On October 13, 2025, the advisers sent home a letter in Filipino and English on the purpose of the study, the quizzes, the observation, the use of the Quarter 1 and Quarter 3 test results and the Quarter 2 grade by code, and the right to refuse or withdraw with no effect on the learner's grades, and the consent form; 228 of 239 parents signed by October 20, 2025. On October 21, 2025, the first researcher explained the study in each section, and each learner with consent signed an assent form in front of the class adviser, who was not a member of the research team, as witness. DepEd's quality tools for completed research ask 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). The lessons were the regular lessons of the quarter, and every learner had the cards in half of them; learners whose parents declined were taught the same, took the quizzes as class work, and were left out of the data and of the observers' draws. 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. The quizzes did not enter the grades. The study followed the Research Management Guidelines (DepEd, 2017). The researchers' use of an artificial intelligence (AI) tool is declared in Appendix F (DepEd, 2026).

**Data Analysis**

For $H_0 1$, the quiz percent of each learner after each lesson the learner attended was modeled with a linear mixed model with the condition (cards or usual), the lesson, and the section as fixed effects and two random intercepts: one for each learner, and one for each lesson in each section, the unit the lot assigned. Learners in the same section on the same day share whatever went well or badly in that period, so a model that treats their quizzes as independent would overstate its precision. The difference is reported in percentage points with its 95% confidence interval. Quizzes missed because of absence were left out; the mixed model uses every quiz taken and assumes that whether a quiz was missed did not depend on the score the learner would have had, and the share missed was compared between the conditions.

As a check that rests only on the lot, a randomization test compared the mean of the 32 lesson-in-section averages with cards and the mean of the

32 without, and found where the observed difference stood among the differences for all 1,679,616 lots the box could have given. The test re-draws the lot exactly as it was drawn, because "the permutation of the data to obtain the distribution of test statistics must correspond to the random assignment that is actually used" (Onghena & Edgington, 1994, p. 784). The two-sided $p$ value is the share of lots with a difference at least as large in either direction.

For $H_0 2$, the same mixed model was fitted to each learner's percent right on the two Quarter 3 test items of each lesson the learner attended. For $H_0 3$, $H_0 4$, and $H_0 5$, the observed card period and the observed usual period of each lesson were paired, and the mean difference over the 16 lessons was tested with a paired $t$ test. Holm's (1979) procedure adjusted the $p$ values of the four secondary tests ($H_0 2$ to $H_0 5$). For $H_0 6$, the mixed model added the interaction of the condition with the lower half, the learners below the median Quarter 2 grade of their section, as set in the proposal.

Agreement between the observers was the intraclass correlation for absolute agreement of a single observer, ICC(A,1), judged by its 95% confidence interval (Koo & Li, 2016, p. 161), for the counts of the 64 learner-periods both observed and for the minutes of the 8 periods.

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 mixed models, the randomization test, the intraclass correlations, and the power simulation), and the notebook reproduces every result from the data (Appendix A).

## Discussion of Results and Reflection

### Implementation

The routine ran as planned. In the 288 questions of the 32 card periods, 96.2% of the cards in the room were held up with an answer, and the count of wrong or blank cards ranged from 2 to 28, a mean of 10.8. The rule called for reteaching 57 times (19.8%), and the teacher followed it on 284 of the 288 questions (98.6%); of the 4 departures, 3 were questions the teacher retaught with fewer than 15 wrong cards, and 1 was a question the teacher moved past with more when the period was running out. In usual periods, 74 of the 288 questions (25.7%) were answered in chorus; in the 16 observed usual periods, a median of 5 hands went up for each of the 111 questions answered by one learner, in rooms with a mean of 57.4 learners present. The teacher retaught 59 times in card periods and 34 times in usual periods. All nine planned questions were asked in every observed period. The first criterion for success was met.

The observers agreed. For the counts, ICC(A,1) = .998, 95% CI [.997, .999], over 64 learner-periods, and the two counts were identical for 60 of them (93.8%). Because the counts differ so much between the conditions, agreement within each condition is the stricter check: .89, 95% CI [.80, .95], in card periods and .92, 95% CI [.84, .96], in usual periods. For the minutes, ICC(A,1) = .99, 95% CI [.94, 1.00], over 8 periods. Every lower bound was at least .75, so the second criterion for success was met.

Of the 3,640 quizzes the learners with consent could have taken, 237 (6.5%) were missed through absence: 7.0% after card lessons and 6.0% after

usual lessons, a difference of 0.9 points within learners, $p = .205$. Another 166 quizzes were taken by learners who had missed the lesson itself and were left out of the main analysis, as planned. The analysis used 3,237 quizzes of 228 learners.

**Question 1: The Next-Day Quizzes**

After lessons with response cards, the learners' mean quiz score was 65.1% ($SD = 26.3$); after usual lessons, 58.9% ($SD = 27.6$). In the mixed model, the cards raised the next day's score by 6.2 percentage points, 95% CI [4.4, 7.9], $z = 6.89$, $p < .001$, with the lesson and the section held constant; $H_0 1$ is rejected. The difference is 0.31 of an item on a five-item quiz, or $d = 0.22$ against the spread of single quiz scores after usual lessons. The learners' own levels varied with a standard deviation of 17.2 points and the lessons in each section with 2.3 points. The randomization test agreed: the observed difference between the lesson-in-section averages, 6.3 points, was matched or exceeded in size by 34 of the 1,679,616 possible lots, $p < .001$. The third criterion for success was met.

Every section scored higher with the cards, by 2.6 to 9.2 points (Table 2), and the card sections were ahead in fifteen of the 16 lessons (Figure 1). Counting also the quizzes of learners who had missed the lesson itself, the difference was 5.6 points, 95% CI [4.0, 7.2].

**Table 2**

*Mean Next-Day Quiz Score by Section and Condition*

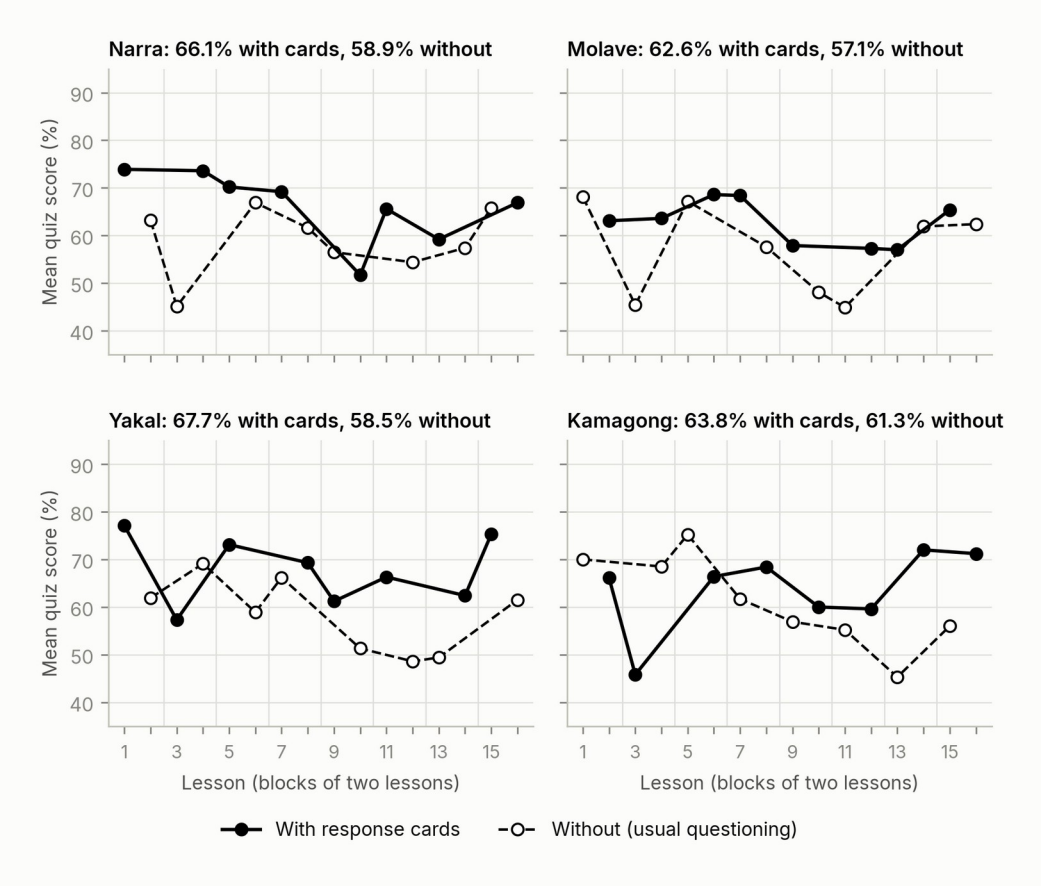
| Section | Learners | With cards (%) | Without cards (%) | Difference (points) |
|---|---|---|---|---|
| Narra | 58 | 66.1 | 58.9 | +7.2 |
| Molave | 56 | 62.6 | 57.1 | +5.5 |

| Section | Learners | With cards (%) | Without cards (%) | Difference (points) |
|---|---|---|---|---|
| Yakal | 59 | 67.7 | 58.5 | +9.2 |
| Kamagong | 55 | 63.8 | 61.3 | +2.6 |

*Note.* Each section used the cards in eight lessons and not in the other eight; a section's eight lessons with cards were not the same lessons as its eight without, so a section's difference includes the difference between its lessons. The mixed model holds the lessons constant.

**Figure 1**

*Each Section's Mean Next-Day Quiz Score After Each Lesson, With and Without Response Cards*




*Note.* Filled points: lessons with response cards. Open points: lessons without. Each lesson was taught in all four sections; the lot gave the cards to two sections in each lesson.

## Question 2: The Quarter 3 Test

The Quarter 3 test was taken by 225 learners: the learner who had transferred and two who were absent when their section took it did not. Eight weeks after the unit, the learners answered the items on lessons taught with

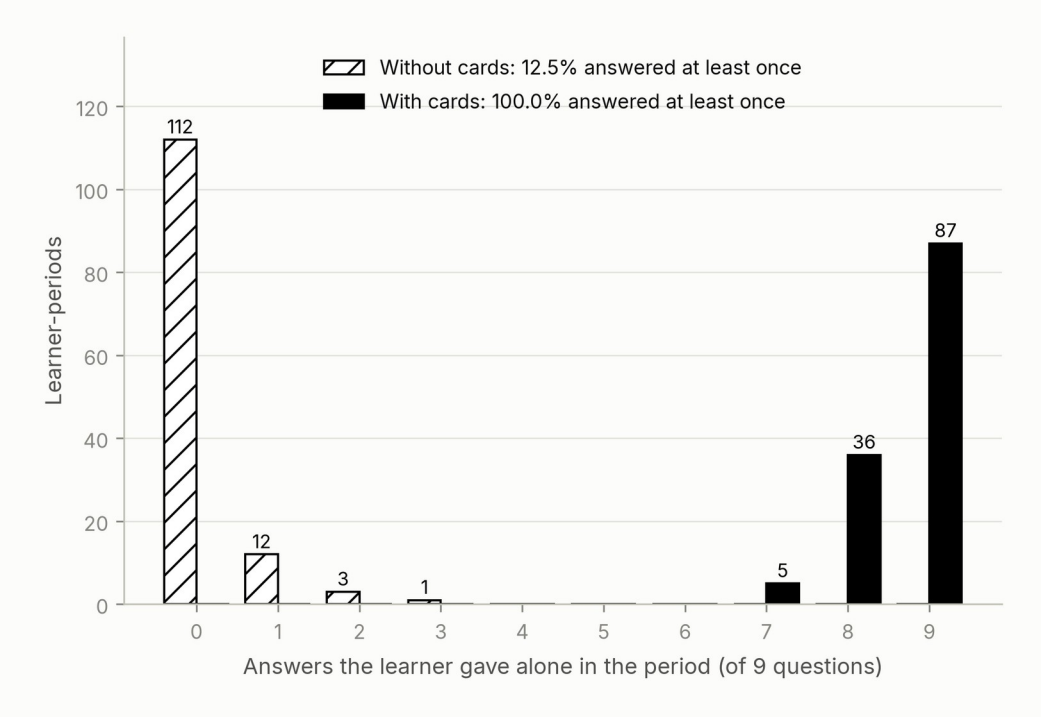
cards 59.4% right and the items on lessons taught without 57.5% right. In the same mixed model, the difference was 2.0 points, 95% CI [−0.2, 4.2], $p$ = .077, Holm-adjusted $p$ = .077. $H_0$2 is not rejected: the difference could have been anywhere from −0.2 to 4.2 points, so the study cannot say whether the cards' advantage lasted to the quarterly test.

**Question 3: How Many Answered**

In usual periods, the observed learners answered alone a mean of 0.16 times a period; 112 of the 128 learner-periods (87.5%) had no answer at all, and no learner answered more than 3 times. In card periods, they answered a mean of 8.64 of the nine questions, and 87 of 128 (68.0%) answered all nine. The paired difference over the 16 lessons was 8.48 answers a period, 95% CI [8.30, 8.65], $t(15) = 104.83$, Holm-adjusted $p < .001$ (Figure 2). In card periods, 100.0% of the observed learners answered at least once, against 12.5% in usual periods, a difference of 87.5 points, 95% CI [78.7, 96.3], Holm-adjusted $p < .001$. $H_0$3 and $H_0$4 are rejected.

**Figure 2**

*How Many Times Each Observed Learner Answered in a Period, With and Without Response Cards*




*Note.* 128 learner-periods in each condition: eight learners drawn by lot in each of the 16 observed periods of each condition. An answer is one the learner gave alone and the teacher could see or hear; chorus answers are not counted.

## Question 4: The Minutes

The three question rounds took a mean of 6.2 minutes in usual periods and 9.3 minutes with cards, 23% of the 40-minute period. The cards added 3.1 minutes a period, 95% CI [2.2, 3.9], $t(15) = 7.37$, Holm-adjusted $p < .001$; $H_0 5$ is rejected. The fourth criterion for success, no more than four extra minutes, was met. The added minutes came from the rest of the lesson: on card days the teacher shortened the closing activity, so the quiz difference is the effect of the cards net of the time they took.

### Question 5: The Lower Half

Among learners below their section's median Quarter 2 grade, the cards raised the next-day score by 7.5 points, 95% CI [5.2, 9.8]; among those above, by 5.0 points, 95% CI [2.9, 7.2]. The difference between the halves, 2.5 points, 95% CI [−0.2, 5.2], $p$ = .073, fits no difference at all as well as one of five points; $H_0$6 is not rejected. The study was planned to find a 4-point difference most of the time, and the interval still leaves room for one, so the lower half may gain more from the cards, but this study does not show it.

## Reflection

### *What We Found*

With response cards, our learners kept more of each lesson to the next day: 6.2 percentage points more on a five-item quiz, in all four sections, with a lot rather than the teacher deciding where the cards were used. The cards turned a lesson in which 12.5% of the observed learners answered anything into one in which every observed learner answered nearly every question, and they let the teacher count the wrong answers instead of guessing from the few who spoke. They took about 3.1 more minutes of a 40-minute period.

### *How It Compares With the Research*

Our difference is smaller than the average effect on quizzes in the studies Randolph (2007) pooled, $d$ = 1.08, and smaller than the 10-point difference in a university course (Marmolejo et al., 2004, p. 409). The measures differ: our $d$ divides by the spread of single five-item quiz scores, which mixes learners and lessons, and our comparison was not hand raising alone but the teacher's usual mix of volunteers, names called, and chorus,

with the same questions. Like Cavanaugh et al. (1996), whose teacher found that "what he had perceived to be an effective lesson was visibly contradicted when many students could not respond correctly" (p. 405), our teacher retaught more often with the cards (59 times against 34), because the cards showed what the chorus had hidden.

### *What Else Could Explain the Difference*

The teacher knew which lessons had cards and may have taught them with more energy; the lot kept such effects from favoring any section or lesson, but not from favoring the condition itself. The observers could not be blind to the condition, since the cards are seen, so the counts and minutes could lean toward what the observers expected; the quizzes, written before the lot and scored with a key, could not. The sections met in fixed periods, from early to late in the morning, and the teacher taught each lesson four times a morning; the lot spread the cards over early and late periods alike. Learners and teacher moved between the conditions from one lesson to the next, and a habit of the card lessons, such as the teacher checking more learners, may have carried into the usual lessons, which would make the difference smaller (induction); the reverse, usual lessons that seemed duller beside the card lessons (contrast), would make it larger (Barlow & Hayes, 1979, pp. 204–205).

### *Limits*

This is one teacher in one school, in classes of 57 to 63, in one quarter's unit on world history. The quizzes measured the next day, and the Quarter 3 test eight weeks later showed a smaller difference. The observers counted eight learners a period, not every learner. Carryover between the two

conditions could have made the difference smaller or larger. The extra gain of the lower half, 2.5 points, was smaller than the 4 points the study was planned to detect, and its interval includes zero.

### *What We Would Change*

We would plan the minutes the cards take into the lesson plan rather than take them from the closing activity, and return to each card lesson's points in a later review, since the difference faded by the Quarter 3 test.

### *Conclusion*

In classes of about 60 learners with 40-minute periods, write-on response cards with a rule for reteaching raised next-day learning in every section, made every observed learner answer, and showed the teacher when to reteach, at a cost of about 3.1 minutes a period and PhP 30.29 a learner.

## Action Plan

**Table 3**

*Action Plan*

| Objective | Activities | Persons involved | Timeline | Resources (PhP) | Success indicator |
|---|---|---|---|---|---|
| Use the cards in every Grade 8 Araling Panlipunan class | Make cards for the other eight Grade 8 sections, with supplies bought from the school's MOOE; plan the minutes the routine takes into each lesson plan | First researcher, Grade 8 Araling Panlipunan teachers, school head | From June 2026 | 5,600 (MOOE supplies) | Every Grade 8 section has a box of cards |
| Share the routine | A Learning Action Cell session on the cue, the scan, and the rule, with this study's data, for all Araling Panlipunan teachers, including those teaching outside their major | Researchers, Head Teacher, Araling Panlipunan department | June 2026 | 0 | Each teacher runs the routine in one class and brings a tally sheet to the next session |
| Hold the gains | A short card review of each lesson's points in the week after it | Araling Panlipunan teachers | From June 2026 | 0 | Term examination items on reviewed lessons |
| Test the routine in other subjects | Proposal for a study in Grade 8 Science, which shares the morning shift's short periods | Researchers, school research committee | Next call for proposals | Up to 30,000 (BERF, school level) | Proposal approved |

## References

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*(2), 199–210. https://doi.org/10.1901/jaba.1979.12-199

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–144, 146–148.

Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction, 21*(6), 715–730. https://doi.org/10.1016/j.learninstruc.2011.04.001

Cavanaugh, R. A., Heward, W. L., & Donelson, F. (1996). Effects of response cards during lesson closure on the academic performance of secondary students in an earth science course. *Journal of Applied Behavior Analysis, 29*(3), 403–406. https://doi.org/10.1901/jaba.1996.29-403

Department of Education. (2012). *Revised guidelines on the opening of classes* (DepEd Order No. 41, s. 2012). https://www.deped.gov.ph/2012/05/24/do-41-s-2012-revised-guidelines-on-the-opening-of-classes-amended-by/

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/2016/06/10/do-39-s-2016-adoption-of-the-basic-education-research-agenda/

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: Araling Panlipunan (Grades 4–10)*. https://www.deped.gov.ph/wp-content/uploads/1-FINAL-MATATAG-Araling-Panlipunan-CG-2023-Grades-4-10.pdf

Department of Education. (2024a). *Policy guidelines on the implementation of the MATATAG curriculum* (DepEd Order No. 010, s. 2024). https://www.deped.gov.ph/2024/07/23/july-23-2024-do-010-

s-2024-policy-guidelines-on-the-implementation-of-the-matatag-curriculum/

Department of Education. (2024b). *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/

Gardner, R., III, Heward, W. L., & Grossi, T. A. (1994). Effects of response cards on student participation and academic achievement: A systematic replication with inner-city students during whole-class science instruction. *Journal of Applied Behavior Analysis, 27*(1), 63–71. https://doi.org/10.1901/jaba.1994.27-63

Heward, W. L., Gardner, R., III, Cavanaugh, R. A., Courson, F. H., Grossi, T. A., & Barbetta, P. M. (1996). Everyone participates in this class: Using response cards to increase active student response. *Teaching Exceptional Children, 28*(2), 4–10. https://doi.org/10.1177/004005999602800201

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

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*(2), 124–144. https://doi.org/10.1037/a0017736

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

Marmolejo, E. K., Wilder, D. A., & Bradley, L. (2004). A preliminary analysis of the effects of response cards on student performance and participation in an upper division university course. *Journal of Applied Behavior Analysis, 37*(3), 405–410. https://doi.org/10.1901/jaba.2004.37-405

Narayan, J. S., Heward, W. L., Gardner, R., III, Courson, F. H., & Omness, C. K. (1990). Using response cards to increase student participation in an elementary classroom. *Journal of Applied Behavior Analysis, 23*(4), 483–490. https://doi.org/10.1901/jaba.1990.23-483

Onghena, P., & Edgington, E. S. (1994). Randomization tests for restricted alternating treatments designs. *Behaviour Research and Therapy, 32*(7), 783–786. https://doi.org/10.1016/0005-7967(94)90036-1

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

Randolph, J. J. (2007). Meta-analysis of the research on response cards: Effects on test achievement, quiz achievement, participation, and off-task behavior. *Journal of Positive Behavior Interventions, 9*(2), 113–128. https://doi.org/10.1177/10983007070090020201

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

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*(1), 12–19, 39.

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 4

*Statement of Expenses*

| Date | Particulars | Receipt no. | Quantity | Amount (PhP) | Category |
|---|---|---|---|---|---|
| October 10, 2025 | Consent letters in Filipino and English and assent forms | OR 1184 | 500 copies at 1.00 | 500.00 | Reproduction |
| October 20, 2025 | Clear sheet protectors, A4, packs of 100 | OR 2207 | 3 packs at 230.00 | 690.00 | Supplies and materials |
| October 20, 2025 | Whiteboard markers, boxes of 12 | OR 2207 | 22 boxes at 210.00 | 4,620.00 | Supplies and materials |
| October 20, 2025 | White cartolina, each sheet cut into four and trimmed to A4 | OR 2208 | 70 sheets at 13.00 | 910.00 | Supplies and materials |
| October 20, 2025 | Cotton cloth for erasers | OR 2208 | 6 meters at 55.00 | 330.00 | Supplies and materials |
| October 20, 2025 | Plastic boxes with lids, one for each section's cards | OR 2209 | 4 boxes at 120.00 | 480.00 | Supplies and materials |
| October 22, 2025 | Quiz sheets: 16 quizzes, 250 copies each, four quizzes to a page | OR 1203 | 1,000 pages at 1.00 | 1,000.00 | Reproduction |
| October 22, 2025 | Observation forms and the teacher's tally sheets | OR 1203 | 150 copies at 1.00 | 150.00 | Reproduction |
| November 17, 2025 | Whiteboard markers, box of 12, to replace dried-out markers | OR 2311 | 1 box at 210.00 | 210.00 | Supplies and materials |
| March 13, 2026 | Printing and binding of the final report | OR 1377 | 3 copies at 300.00 | 900.00 | Reproduction |
| **Total** | | | | **9,790.00** | |

*Note.* Fund source: Basic Education Research Fund, school level, PhP 10,000.00, released 8,000.00 with the memorandum of agreement and the work plan, and 2,000.00 on the acceptance of the final report (DepEd, 2017, Sec. V.C.i.5). Unspent balance of PhP 210.00 returned. Every line is an expense the fund allows: supplies and materials, and reproduction, printing, and binding (Sec. V.C.i.4). Receipt numbers are invented for this sample.

# Appendices

## Appendix A. Data Set and Notebook

The Excel data set (the lot and the schedule, every learner's answer to every quiz item and to the Quarter 3 test, attendance, the observers' counts and timings, the teacher's tally, the purchases, 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. The Quizzes and the Test Map

The sixteen quizzes with their keys, and the Quarter 3 test's map of items to lessons. Kept with the school's copy so that the items can be used again.

## Appendix D. The Response Card Routine

How the cards were made, the cue, the rule, the tally sheet, and a sample quiz (the kit published with this report).

## Appendix E. Observation Protocol

The draw of learners, what counts as an answer, the timing of the rounds, and the observation form.

## Appendix F. 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).