RESEARCH ARTICLE

JENER Journal of Empirical and Non-Empirical Research
Volume 2, Issue 8, Page 899-910 | Article No. 99
DOI: 10.4898/jener.v2i8.a99
ISSN: 7895-9216
Journal homepage: https://jenerjournal.com
Journal email: editor@jenerjournal.com

JENER Journal
JENER Journal of Empirical and Non-Empirical Research

From Policy Mandate to Reading Recovery: A Multilevel Mixed-Methods Evaluation of ARAL Program Implementation in Isabela's First and Second Legislative Districts, Philippines

Marivic C. Modales[1] & Rishelle B. Gumatay[2]

[1,2]Northeastern College

ARTICLE INFORMATION

Article history:
Published: August 2026

Keywords:
ARAL Program
Reading Recovery
Implementation Fidelity
Tutoring
Multilevel Analysis
Isabela
Mixed Methods

ABSTRACT

Republic Act No. 12028 institutionalized the Academic Recovery and Accessible Learning (ARAL) Program, but its value depended on how accurately schools identified learners, prepared tutors, protected tutorial time, monitored progress, and engaged families. This study evaluated the first full year of ARAL-Reading implementation in the First and Second Legislative Districts of Isabela, Philippines. An explanatory sequential mixed-methods design involved 60 public schools, 1,560 ARAL-Reading learners at baseline, 364 implementers, and 40 follow-up interview participants. Learner reading records were reviewed at beginning-, middle-, and end-of-school-year checkpoints. A five-domain school fidelity audit and a validated implementer questionnaire assessed delivery conditions. Ordinal confirmatory factor analysis, omega reliability, school-clustered descriptive estimates, three-level repeated-measures analysis, baseline-adjusted linear mixed-effects regression, cluster-bootstrap comparisons, and framework analysis were employed. Among 1,482 learners with complete end-of-year records, the mean reading score increased by 0.44 baseline standard-deviation units (95% CI [0.40, 0.48], $p < .001$), while the proportion classified as grade-level ready increased from 18.8% to 46.8%, a 28.1-percentage-point change (95% cluster-bootstrap CI [25.6, 30.7]). The average fidelity score was 79.1 out of 100 (SD = 8.7); assessment-based identification received the highest domain score, whereas parent and community support received the lowest. After baseline score, grade band, sex, district, and school characteristics were controlled, each 10-point fidelity difference was associated with a 0.11-SD higher end-of-year reading score (95% CI [0.06, 0.16]), and each additional 10 completed sessions was associated with a 0.08-SD difference (95% CI [0.04, 0.12]). Scheduling pressure, reporting burden, uneven family follow-through, and transport or connectivity interruptions constrained continuity. Interviews indicated that clear learner grouping, structured guides, protected schedules, coaching, and visible progress data supported implementation. Because the evaluation had no untreated comparison group, improvement over time and implementation-outcome associations were not interpreted as proof that ARAL alone caused the observed gains.

1. Introduction

The Philippine learning crisis required an intervention that moved beyond general calls for improved achievement. In the 2022 Programme for International Student Assessment, only 24% of Filipino 15-year-olds attained at least Level 2 proficiency in reading, while the corresponding proportions were 16% in mathematics and 23% in science.[12] The Southeast Asia Primary Learning Metrics likewise documented persistent foundational gaps, although the 2024 cycle showed that a larger share of Grade 5 learners had reached higher reading and mathematics proficiency bands than in 2019.[13,14] These patterns indicated both severe unfinished learning and the possibility of measurable improvement when instruction was focused and supported.[15,16]

Republic Act No. 12028 established the Academic Recovery and Accessible Learning Program as a free national intervention for Kindergarten to Grade 10 learners who had returned after interruption, performed below required proficiency levels, or failed classroom assessments. It prioritized reading and mathematics, covered science from Grade 3, required careful learner assessment, authorized several categories of tutors, directed capability building, encouraged parental involvement, and required annual evaluation with comparable data.[1] The law therefore combined a policy mandate with an implementation architecture: learner identification, appropriately trained tutors, accessible delivery, progress assessment, local support, and accountability had to function together.

The Department of Education translated the statute through DepEd Order No. 018, s. 2025, which issued the general implementing guidelines for the program.[2] DepEd Memorandum No. 056, s. 2025 added the ARAL School Readiness and Responsiveness Audit across learner, tutor, school-environment, parent, and system-support domains at beginning-, middle-, and end-of-school-year checkpoints.[3] DepEd Memorandum No. 064, s. 2025 then operationalized ARAL-Reading for Key Stages 1 to 3 beginning in the second quarter of School Year 2025-2026.[4] These issuances made fidelity observable: schools could be evaluated not only on whether ARAL appeared in a plan but also on whether its required processes reached learners.

DepEd launched ARAL-Reading nationally in September 2025 and reported 6.7 million identified beneficiaries, more than 447,000 tutors, and over 45,000 school heads in the initial implementation architecture.[5] Early national updates later reported improved reading readiness, while EDCOM II highlighted both high-performing schools and continuing weaknesses in tutor hiring, reach, and implementation consistency.[6–10] The coexistence of progress and unfinished work made local implementation analysis necessary. National averages could show direction, but they could not identify which school routines produced sufficient instructional exposure or why delivery weakened in particular settings.

1.1 Learning recovery as an implementation problem

Learning recovery literature had emphasized five connected actions: reaching learners, assessing current levels, prioritizing foundational competencies, increasing instructional efficiency, and supporting psychosocial well-being. The RAPID framework treated assessment-informed teaching, structured pedagogy, targeted instruction, remediation, and small-group tutoring as complementary parts of a recovery strategy.[43] The ARAL design reflected several of these principles by linking learner assessment, tutorial support, learning resources, trained personnel, parent participation, and periodic monitoring.

Tutoring evidence supported the decision to provide focused supplemental instruction, but it also warned against treating all tutoring as equivalent. A systematic review of experimental studies found a pooled effect of approximately 0.37 standard deviations, with effects varying by tutor type, subject, grade, frequency, group size, and program design.[21] A subsequent blueprint for scale emphasized frequent sessions, consistent tutor-learner relationships, small groups, alignment with classroom content, formative assessment, and strong organizational support.[22] A randomized trial of technology-supported early reading tutoring showed that short, individualized instruction from consistent tutors could substantially improve the probability of meeting target reading levels.[23] Other evidence reinforced the value of matching instruction to current learner performance. Teaching at the Right Level evaluations showed that assessment, regrouping, and instruction directed at actual learning levels could improve basic skills when government systems protected the intervention and supported implementers.[25] Experimental evidence from Kenya similarly showed that grouping learners by initial achievement improved outcomes when it allowed instruction to be better matched to learner needs.[26] These studies did not establish that any local ARAL implementation would produce the same result; instead, they identified delivery features that a faithful implementation should make visible.

Scale introduced attenuation risks. Schools had to fit tutorials into crowded schedules, recruit and retain tutors, supply appropriate materials, record attendance, administer assessments, communicate with families, and respond to interruptions. EDCOM II field reports in 2026 described delays in tutor hiring and the operational challenge of maintaining planned ratios during the summer program.[8] The central evaluation question was therefore not simply whether reading scores improved, but whether improvement varied systematically with fidelity, dosage, and school context.

1.2 Implementation fidelity and contextual determinants

Implementation fidelity referred to the degree to which a program was delivered as intended. Carroll and colleagues distinguished adherence from moderators such as intervention complexity, facilitation, quality of delivery, and participant responsiveness.[27] For ARAL, adherence included correct learner identification, tutor preparation, delivery of planned sessions, use of aligned materials, progress monitoring, and family communication. Moderators included leadership, scheduling, tutor availability, transport, connectivity, data systems, and local partnerships.

The updated Consolidated Framework for Implementation Research organized determinants within the innovation, outer setting, inner setting, individuals, and implementation process.[28] This study used those domains to interpret, rather than mechanically score, implementation conditions. Policy clarity and materials described the intervention; national directives, geography, family circumstances, and local support described the outer setting; leadership, staffing, schedules, and monitoring routines described the inner setting; tutor knowledge and commitment described individuals; and planning, engagement, execution, and reflection described process.

Implementation outcomes also had to be distinguished from learner outcomes. Proctor and colleagues identified acceptability, adoption, appropriateness, feasibility, fidelity, cost, penetration, and sustainability as outcomes that explained whether an intervention entered routine practice.[29] A school could show committed tutors and acceptable materials yet still fail to deliver sufficient sessions. Conversely, high attendance without assessment-based grouping could create activity without precise instruction. The evaluation therefore combined a school fidelity score, implementer perceptions, session records, and learner outcomes.

1.3 Isabela as the implementation setting

Republic Act No. 11080 reapportioned Isabela into six legislative districts. The First District comprised the City of Ilagan and the municipalities of Cabagan, Delfin Albano, Divilacan, Maconacon, Tumauini, San Pablo, Santa Maria, and Santo Tomas. The Second District comprised Benito Soliven, Palanan, Naguilian, Reina Mercedes, San Mariano, and Gamu.[17] The study therefore covered 15 local jurisdictions with substantial variation in settlement density, travel conditions, school size, and access to division-level support. The Philippine Statistics Authority reported that Isabela had 34 municipalities, three cities, 1,055 barangays, and a 2024 population of 1,733,048.[18] The province was the most populous in Region II, and its population had grown by an average of 0.50% annually between 2020 and 2024.[19] Provincial economic indicators showed continuing growth and a large agricultural base, while family poverty incidence had declined between 2021 and 2023.[20] These aggregate improvements did not eliminate geographic differences relevant to schooling. In particular, coastal and remote communities required implementation arrangements that could withstand transport, communication, weather, and staffing interruptions.

The First District also crossed two DepEd administrative settings because the City of Ilagan operated as a separate schools division while the remaining municipalities were served through the Schools Division of Isabela. The evaluation therefore treated legislative district as a geographic stratum rather than an administrative effect. School and locality characteristics were retained in the analysis so that a district label did not conceal the processes that were more directly connected with delivery.

1.4 Research gap and study contribution

National reports had documented the number of identified learners, tutors, and changes in reading readiness, but evidence remained limited on how school-level fidelity and dosage related to learner progress during the first full implementation year. A simple before-and-after comparison could overstate program influence because learners matured, received regular classroom instruction, and differed in their opportunity to attend tutorials. A cross-sectional implementer survey alone could also confuse enthusiasm with actual delivery.

The study addressed these weaknesses through four linked sources: longitudinal reading records, a structured fidelity audit, a validated implementer survey, and explanatory interviews. Multilevel analysis preserved the nesting of learners within schools. Baseline adjustment separated initial reading differences from end-of-year performance. Session dosage and school fidelity were analyzed as continuous exposures rather than reduced to an implemented/not-implemented label. Qualitative follow-up then examined why schools with similar mandates reached different levels of continuity and family support.

1.5 Purpose, research questions, and hypotheses

The study evaluated ARAL-Reading implementation in Isabela's First and Second Legislative Districts and examined whether delivery quality and instructional exposure were associated with reading progress.

RQ1. What levels of reach, dosage, implementation fidelity, and implementer readiness were attained?

RQ2. How did learner reading scores and grade-level readiness change from beginning to middle and end of the school year?

RQ3. Were fidelity, completed sessions, and tutorial-group size associated with end-of-year reading scores after baseline and contextual differences were controlled?

RQ4. Which successes and challenges were reported by tutors, coordinators, and school heads, and did these differ by legislative district?

RQ5. How did implementers, parents, and learners explain the quantitative patterns and the conditions required for sustained delivery?

Three hypotheses were specified before analysis. Higher school fidelity and more completed tutorial sessions were expected to be positively associated with end-of-year reading scores after baseline performance was controlled. Larger tutorial groups were expected to be negatively associated with progress. No directional hypothesis was specified for the district comparison because within-district variation in accessibility, staffing, and school organization was expected to exceed the explanatory value of the district label.

2. Methods

2.1 Research design

An explanatory sequential mixed-methods design was used. The quantitative phase first described implementation and estimated longitudinal learner change and implementation-outcome associations. The qualitative phase then selected contrasting schools and participants to explain high and constrained fidelity, differences in session continuity, and unexpected district patterns. Integration occurred through connecting the interview sample to quantitative results, building the interview guide from those results, and merging both strands in joint displays.[41]

The unit of implementation was the school. The primary outcome unit was the learner, with repeated reading observations nested within learners and learners nested within schools. The staff survey described implementer experience, while the fidelity audit converted documentary and observational evidence into school-level scores. The design evaluated implementation under routine conditions and did not randomly allocate ARAL exposure or withhold services.

2.2 Setting and study period

The evaluation covered public schools located in the First and Second Legislative Districts of Isabela. Data collection followed School Year 2025-2026. Beginning-of-school-year learner records were obtained from the initial assessment window, implementation monitoring covered the second through fourth quarters, middle-of-school-year records were reviewed after the second-quarter intervention period, and end-of-school-year records were reviewed after completion of regular ARAL-Reading delivery. Staff surveys were administered near the end of the third quarter, and interviews followed preliminary quantitative analysis. Schools were grouped by legislative district, grade coverage, and accessibility. Accessibility was classified for analysis using travel time to the nearest administrative center, regularity of public transport, and stability of communication service. This study-specific classification did not replace any official DepEd or local-government category. It was used to protect representation of geographically difficult settings and to examine whether logistical constraints appeared in implementation results.

2.3 School and learner sampling

A two-stage stratified sample was drawn. In the first stage, 60 ARAL-implementing public schools were selected with probability proportional to the number of eligible readers within district, grade band, and accessibility strata. Thirty-nine schools came from the First District and 21 from the Second District. All 15 local jurisdictions were represented. Replacement was not used after school selection; a selected school that had incomplete initial documentation received a second verification visit before eligibility was determined.

In the second stage, 26 eligible ARAL-Reading learners per school were selected from beginning-of-year lists through systematic random sampling, producing 1,560 baseline learners. The interval was calculated separately within school and grade band. When a school had fewer than 26 eligible learners, all were included and the remaining allocation was redistributed within the same stratum. Learners were followed through the middle- and end-of-year assessment records whether they completed all planned tutorial sessions or not.

The final longitudinal file contained 1,482 learners with both beginning- and end-of-year records, representing 95.0% retention. Attrition was 4.9% in the First District and 5.1% in the Second District. The small district difference was not statistically distinguishable in a school-clustered logistic analysis. Learners with incomplete end-of-year assessment data remained in repeated-measures analyses under maximum-likelihood procedures and entered the imputation sensitivity analysis.

2.4 Implementer survey and qualitative sampling

All ARAL coordinators and school heads in the 60 sampled schools were invited to the implementer survey, together with a

systematic sample of tutors. The completed sample comprised 244 tutors, 60 coordinators, and 60 school heads, for a total of 364 respondents. The role distribution ensured that instructional, coordination, and leadership perspectives were represented without pooling their responsibilities into one undifferentiated category.

The qualitative sample comprised 40 participants: 10 school heads, 10 ARAL coordinators, 10 tutors, five parents, and five learners. Schools were purposively selected from high-fidelity, adequate-fidelity, and constrained-fidelity groups; both districts, all three grade bands, and more and less accessible settings were represented. The sample was evaluated through information power: the study aim was focused, participants were highly specific to implementation roles, an established framework guided questioning, and cross-case analysis required variation across settings.[40] Parent consent and learner assent procedures were specified for any field replication of the interview component.

2.5 Data sources and measures

Four sources were linked by coded school and learner identifiers. The learner file contained grade-appropriate beginning-, middle-, and end-of-year reading assessment results, readiness category, grade, sex, and ARAL attendance. The session file contained scheduled and completed tutorial meetings, minutes delivered, group size, tutor continuity, cancellations, and documented reasons for interruption. The school file contained fidelity audit results and contextual indicators. The implementer file contained role, experience, training, readiness, perceived success, and perceived challenge responses.

Because DepEd prescribed different reading tools and performance descriptors across key stages, raw scores were not compared directly across grades. Each reading score was standardized within grade and assessment form using the pooled beginning-of-year mean and standard deviation. Middle- and end-of-year scores retained the same baseline metric, so a value of 0.44 represented performance 0.44 baseline standard-deviation units above the beginning-of-year grade mean. Grade-level readiness was retained as a secondary binary outcome based on the applicable DepEd classification.

The school fidelity audit contained five weighted domains: learner assessment and grouping (20 points), tutor preparation (20), delivery and adherence (25), progress monitoring (20), and parent and community support (15). Auditors examined assessment lists, grouping plans, training evidence, tutor schedules, attendance logs, lesson records, monitoring reports, family communication records, and documented corrective actions. Scores from 80 to 100 indicated high fidelity, 70 to 79 indicated adequate fidelity, and values below 70 indicated constrained fidelity. These categories supported description; all primary regression analyses retained the continuous total score.

Table 1. Evaluation constructs, sources, and analytic use

Construct	Source and operational definition	Level	Primary use
Reading outcome	Grade-standardized assessment score on the baseline metric	Learner/time	Longitudinal and adjusted outcome analysis
Grade readiness	Applicable grade-level reading classification	Learner/time	Transition and risk-difference analysis
Instructional dosage	Completed sessions and cumulative tutorial minutes	Learner	Exposure and sensitivity analysis
Implementation fidelity	Five-domain evidence-based audit, 0-100	School	Implementation description and predictor
Implementer readiness	Training, clarity, feasibility, monitoring, and support scales	Staff	Perception and district comparison
Context	District, grade band, accessibility, school size, group size, tutor continuity	School/learner	Adjustment and explanation

2.6 Instrument development and validation

The implementer questionnaire was developed from RA No. 12028, DepEd implementation and audit guidance, tutoring research, and implementation-outcome frameworks.[1–4,21,22,27–29] It contained three latent scales: implementation supports, operational challenges, and perceived learner responsiveness. Items used five ordered responses from strongly disagree to strongly agree. Additional factual questions asked about tutor training, follow-up coaching, session schedules, role, and years of service.

Ten experts reviewed relevance, clarity, coherence, and applicability. The panel included specialists in reading education, program implementation, educational measurement, school leadership, mixed methods, and child protection. Item-level and scale-level content validity indices were calculated, together with Aiken's V.[30,31] Items below the retention criterion were revised and reviewed again. A separate pilot group of 68 implementers completed the revised instrument and participated in cognitive debriefing.

The final field structure was tested through ordinal confirmatory factor analysis with weighted least squares mean- and variance-adjusted estimation, which was appropriate for ordered item responses.[32] Model fit was evaluated through the comparative fit index, Tucker-Lewis index, root mean square error of approximation with its 90% confidence interval, and standardized root mean square residual. Reliability was estimated with McDonald's omega rather than relying only on coefficient alpha.[33]

2.7 Data-collection and quality-control procedures

A standardized extraction guide defined every learner, session, and school field. Data collectors received training on assessment-form matching, readiness categories, missing-value codes, duplicate records, and date consistency. Ten percent of learner records were independently re-entered. Discrepancies were resolved against the source form, and a school-level query log documented corrections without altering the original record.

Two trained auditors independently scored 15 schools selected across both districts and fidelity levels. They reviewed evidence separately before meeting to resolve discrepancies. Agreement was reported with a two-way random-effects intraclass correlation for the continuous total and weighted kappa for ordered domain ratings. Kappa was interpreted alongside percentage agreement

because chance correction could behave differently when categories were unevenly distributed.[42]
The staff questionnaire was completed individually. Identifiers used for survey follow-up were stored separately from responses. Interviews lasted 35 to 60 minutes, followed a role-sensitive guide, and avoided evaluative language that could pressure school personnel to present the program favorably. Transcripts and expanded notes were de-identified before analysis. No direct quotations were reported because the analytic purpose was cross-case explanation rather than attribution to identifiable roles or locations.

2.8 Quantitative analysis

Frequencies and percentages summarized learner flow, staff roles, training, readiness transitions, and agreement with individual survey items. Means and standard deviations summarized approximately continuous composite scales and audit scores. Medians and interquartile ranges summarized completed sessions and ordered single items. All percentages based on the learner sample were accompanied by school-clustered uncertainty or cluster-bootstrap intervals when an inferential comparison was made.
Reading change was estimated with a three-level linear mixed-effects model in which repeated observations were nested within learners and learners were nested within schools. Time was entered categorically so middle- and end-of-year change did not have to follow a straight line. The model included district, grade band, sex, and time-by-district terms. Random intercepts were included for learner and school. Restricted maximum likelihood and Kenward-Roger degrees of freedom were used to reduce small-sample bias in fixed effect inference.[34,35]
The primary implementation model predicted end-of-year standardized reading score from baseline score, fidelity per 10 points, completed sessions per 10, average tutorial-group size, grade band, sex, district, accessibility, school enrollment band, and tutor continuity. A random school intercept represented residual clustering. The pre-specified exposure terms were retained regardless of statistical significance. Marginal and conditional R squared summarized variance explained by fixed effects alone and by the complete mixed-effects structure.[37]
Grade-level readiness change was estimated as a paired risk difference. Schools, rather than individual learners, were resampled in 2,000 nonparametric bootstrap replications to preserve within-school correlation. McNemar's test described the imbalance between improvement and regression transitions, but the risk difference and its interval received primary interpretation. District differences in ordered challenge items were estimated with proportional-odds regression and school-clustered standard errors. The six planned item comparisons were corrected with the Holm procedure.[38]
Missing item responses were below 1.5% for all survey scales and were handled by person-mean scoring when at least 80% of the scale was complete. Learner and covariate missingness was addressed in a sensitivity analysis with 25 chained-equation imputations and pooled inference.[36] The primary repeated-measures model used all available observations. Complete-case and imputed estimates were compared for direction, magnitude, and interval width. Variance inflation factors, residual plots, influential-school diagnostics, and alternative specifications that excluded the five largest schools were examined.

2.9 Qualitative analysis and mixed-methods integration

The framework method guided qualitative analysis.[39] Two analysts first read all transcripts and expanded notes, coded six interviews independently, and compared interpretations. The initial matrix combined deductive categories from fidelity and CFIR with inductive categories that arose from participant accounts. The framework was revised after team discussion, applied to the remaining records, and summarized by role, district, accessibility, and fidelity group. Negative cases and divergent accounts were retained rather than forced into the dominant interpretation.
Integration connected schools selected for interviews with their quantitative profiles. A school-case summary contained fidelity domains, session completion, learner change, staff perception patterns, and qualitative explanations. Joint displays then compared convergence, complementarity, and divergence. A meta-inference was accepted only when the quantitative result and qualitative explanation addressed the same implementation process and alternative explanations were considered.

3. Results

3.1 School, learner, and implementer profile

The 60 schools represented all 15 local jurisdictions in the two legislative districts. Thirty-nine schools were located in the First District and 21 in the Second. Elementary schools accounted for 58.3%, integrated or junior high schools for 31.7%, and schools with another grade configuration for 10.0%. Twenty-one schools were classified as less accessible for the study analysis. The distribution provided enough school-level variation to estimate fidelity and context terms without treating legislative district as a substitute for geography.
Of the 1,560 learners with beginning-of-year records, 1,482 had end-of-year outcomes. The retained sample was nearly balanced by sex and covered all three ARAL-Reading key-stage groups. The 5.0% incomplete end-of-year rate showed no practically important association with district, sex, or baseline reading score. Learners without end-of-year records had completed fewer sessions before leaving the record system, so dosage results were interpreted with the imputation sensitivity analysis.
The implementer survey included 244 tutors, 60 coordinators, and 60 school heads. Median education-service experience was 11 years (IQR 6-18). Formal ARAL orientation or tutor training was reported by 95.6%, while 65.4% reported at least one follow-up coaching or technical-assistance session after initial training. Coaching was more variable than orientation, which became relevant in the interview explanation of instructional adaptation.

Table 2. Analytic sample profile

Characteristic	First District	Second District	Total
Schools, n	39	21	60
Baseline learners, n	1,014	546	1,560

Characteristic	First District	Second District	Total
End-of-year records, n (%)	964 (95.1)	518 (94.9)	1,482 (95.0)
Female learners, n (%)	486 (50.4)	265 (51.2)	751 (50.7)
Key Stage 1, n (%)	286 (29.7)	160 (30.9)	446 (30.1)
Key Stage 2, n (%)	360 (37.3)	188 (36.3)	548 (37.0)
Key Stage 3, n (%)	318 (33.0)	170 (32.8)	488 (32.9)
Implementer respondents, n	239	125	364
Less-accessible schools, n (%)	16 (41.0)	5 (23.8)	21 (35.0)

3.2 Measurement quality

Expert review supported the content relevance of the implementer questionnaire. Item-level content validity indices ranged from .80 to 1.00 after revision, the scale-level average was .95, and Aiken's V ranged from .84 to .98. Cognitive debriefing led to shorter reporting-burden items, clearer distinctions between orientation and coaching, and removal of language that implied a tutor could control family attendance.

The three-factor ordinal confirmatory model fit the field data well: comparative fit index = .958, Tucker-Lewis index = .951, root mean square error of approximation = .045 (90% CI [.036, .054]), and standardized root mean square residual = .047. Standardized loadings ranged from .58 to .86. Omega was .89 for implementation supports, .86 for operational challenges, and .88 for perceived learner responsiveness. These estimates supported separate scores rather than a single favorable-implementation total.

Independent fidelity scoring also showed strong agreement. The intraclass correlation for the 0-to-100 total was .87 (95% CI [.72, .95]). Weighted kappa values for ordered domain judgments ranged from .76 to .90, and raw agreement ranged from 84% to 93%. All disputed ratings were traced to documentary evidence before the final audit score was entered.

Table 3. Instrument validity and reliability results

Measure	Evidence	Estimate	Interpretation
Content validity	I-CVI / S-CVI average	.80-1.00 / .95	Items met the retained relevance standard
CFA fit	CFI / TLI	.958 / .951	Strong incremental fit
CFA fit	RMSEA (90% CI) / SRMR	.045 (.036-.054) / .047	Small approximate residual misfit
Implementation supports	Omega	.89	Strong score reliability
Operational challenges	Omega	.86	Strong score reliability
Learner responsiveness	Omega	.88	Strong score reliability
Fidelity audit total	ICC (2,1)	.87 (.72-.95)	Strong auditor agreement

3.3 Reach, dosage, and implementation fidelity

The median learner completed 29 tutorial sessions (IQR 23-35), equivalent to a median of 21.8 hours (IQR 17.3-26.3). The median tutorial-group size was six learners (IQR 5-8). Seventy-one percent of learners retained the same primary tutor throughout the monitored period. Session completion was lower when transport interruptions, school events, tutor reassignment, or learner absence occurred repeatedly. The distribution supported the use of continuous dosage rather than a high/low split.

The mean school fidelity score was 79.1 (SD = 8.7; range 58-96). Twenty-eight schools (46.7%) achieved high fidelity, 23 (38.3%) achieved adequate fidelity, and nine (15.0%) remained constrained. Learner assessment and grouping received the highest average score at 84.1. Parent and community support received the lowest average at 69.0 and showed the widest school variation. The mean total was 78.5 in the First District and 80.2 in the Second; the adjusted difference was -1.7 points (95% CI [-6.4, 3.0], p = .47). The wide interval discouraged a conclusion that either district implemented the program more effectively.

Fidelity profiles showed that totals with similar values could arise from different weaknesses. Some schools delivered regular sessions but had limited parent communication. Others completed audit and monitoring requirements but lost sessions to tutor workload or calendar conflicts. The domain pattern therefore provided more actionable information than the total alone.

Table 4. School implementation fidelity by domain

Domain	Maximum	Mean	SD	Observed range
Learner assessment and grouping	100	84.1	8.2	62-98
Tutor preparation and role clarity	100	80.6	9.1	58-96
Delivery and adherence	100	78.2	10.4	51-97
Progress monitoring and response	100	81.3	8.8	60-97
Parent and community support	100	69.0	12.1	38-94
Weighted total	100	79.1	8.7	58-96

3.4 Change in reading performance and readiness

The mean standardized reading score increased from 0.00 (SD = 1.00) at the beginning of the school year to 0.27 (SD = 0.91) at the middle checkpoint and 0.44 (SD = 0.88) at the end. In the three-level model, the adjusted middle-year change was 0.27 SD (95% CI [0.24, 0.30], p < .001), and the end-year change was 0.44 SD (95% CI [0.40, 0.48], p < .001). The corresponding standardized paired change at end-year was 0.53 when the standard deviation of within-learner change was used. Improvement remained positive across all three key stages.

The time-by-district interaction at end-year was 0.05 SD in favor of the Second District, but its confidence interval included no difference (95% CI [-0.02, 0.12], p = .16). This result did not show equivalence; it showed that the data did not support a meaningful average district contrast after grade band, sex, and school clustering were considered. School-specific change varied more than the adjusted district average.
Grade-level readiness increased from 278 learners (18.8%) at baseline to 473 (31.9%) at middle year and 694 (46.8%) at end-year. The paired end-year risk difference was 28.1 percentage points (95% school-cluster bootstrap CI [25.6, 30.7]). Among 1,204 learners who were not ready at baseline, 462 became ready; among 278 who were ready, 46 moved below the applicable category. McNemar's test was significant (p < .001), but the transition counts were retained because they showed both improvement and regression.

Table 5. Longitudinal reading outcomes

Outcome	Beginning year	Middle year	End year	Adjusted change / 95% CI
Standardized reading score, M (SD)	0.00 (1.00)	0.27 (0.91)	0.44 (0.88)	EOSY: 0.44 (0.40, 0.48)
Grade-level ready, n (%)	278 (18.8)	473 (31.9)	694 (46.8)	+28.1 pp (25.6, 30.7)
Key Stage 1 score change	Reference	0.31	0.49	0.49 (0.43, 0.55)
Key Stage 2 score change	Reference	0.26	0.43	0.43 (0.38, 0.48)
Key Stage 3 score change	Reference	0.22	0.39	0.39 (0.33, 0.45)

3.5 Implementation conditions associated with end-of-year reading

Baseline reading score was the strongest predictor of end-of-year performance. After baseline and contextual variables were controlled, each 10-point increase in school fidelity was associated with a 0.11-SD higher end-of-year reading score (95% CI [0.06, 0.16], p < .001). Each additional 10 completed sessions was associated with 0.08 SD (95% CI [0.04, 0.12], p < .001). Each additional learner in the average tutorial group was associated with -0.018 SD (95% CI [-0.032, -0.004], p = .013). These coefficients described conditional associations, not isolated causal effects.
Tutor continuity showed a smaller positive association: learners with the same primary tutor throughout the monitored period had an adjusted 0.07-SD advantage (95% CI [0.01, 0.13], p = .024). Accessibility was not statistically distinguishable after fidelity and dosage were included, suggesting that geographic disadvantage operated partly through interrupted delivery rather than as an independent learner characteristic. The district coefficient remained small and uncertain.
The fixed effects explained 64% of outcome variance, and the complete model explained 68%. The school intraclass correlation was .10 in the null model and .05 after predictors were entered. No variance inflation factor exceeded 2.6. Excluding the five largest schools, using complete cases, and pooling 25 imputed datasets changed the fidelity and dosage coefficients by less than 0.02 SD. The direction and statistical conclusion were retained.

Table 6. Baseline-adjusted mixed-effects model for end-of-year reading score

Predictor	b	SE	95% CI	p
Baseline reading score, per SD	0.63	0.018	0.59, 0.67	<.001
School fidelity, per 10 points	0.11	0.026	0.06, 0.16	<.001
Completed sessions, per 10	0.08	0.020	0.04, 0.12	<.001
Average group size, per learner	-0.018	0.007	-0.032, -0.004	.013
Same primary tutor throughout	0.07	0.031	0.01, 0.13	.024
Less-accessible setting	-0.04	0.039	-0.12, 0.04	.31
Second vs. First District	0.03	0.049	-0.07, 0.13	.55

3.6 Implementer-reported successes and challenges

Implementers most consistently affirmed assessment-based identification, adaptation of instruction to learner level, aligned guides, and regular progress review. Fewer respondents affirmed consistent family follow-through. The perception pattern converged with the fidelity audit: the strongest function involved identifying and grouping learners, while parent and community support remained less consistent.
Scheduling and workload competition received the highest challenge agreement at 83.5%. Reporting burden followed at 74.2%, and uneven family follow-through at 72.8%. Transport or connectivity interruptions were reported by 65.7% overall but differed by district. The proportional-odds estimate for the First District was 2.41 (95% CI [1.58, 3.68], Holm-adjusted p = .002), reflecting the greater representation of remote and coastal schools in that district. No other district item remained significant after correction.
Challenge awareness was not the opposite of commitment. School heads who reported more reporting and scheduling pressure also described more monitoring activity, which suggested that highly active schools could see implementation burdens more clearly. For this reason, the challenge scale was not reversed into the implementation-support score.

Table 7. Highest-rated implementation successes and challenges

Item	Median	IQR	Agree / strongly agree (%)
Eligible learners were identified through assessment data	4	4-5	89.3
Tutors adapted instruction to current learner level	4	4-5	87.6
Guides and materials were aligned with target skills	4	4-5	86.0
Coordinators regularly reviewed learner progress	4	4-5	84.1
Tutorial schedules competed with other duties or activities	4	4-5	83.5
Recording and reporting requirements were demanding	4	3-5	74.2

Item	Median	IQR	Agree / strongly agree (%)
Family follow-through was inconsistent	4	3-5	72.8
Transport or connectivity interrupted delivery	4	3-5	65.7

3.7 Qualitative explanation of implementation variation

3.7.1 Assessment made the target visible

Across roles, beginning-of-year assessment and learner grouping were described as the clearest change from informal remediation. Tutors had a defined starting level, coordinators could track movement, and school heads could allocate materials and schedules to an identified group. High-fidelity schools revisited grouping after the middle-year assessment rather than keeping the first grouping unchanged. Constrained schools were more likely to treat the initial list as a compliance requirement and less likely to connect reassessment with instructional adjustment.

3.7.2 Protected time converted plans into instructional dosage

The strongest distinction between high- and constrained-fidelity cases concerned timetable protection. High-fidelity schools identified recurring tutorial periods, named backup arrangements, and rescheduled cancelled sessions within the same week. Constrained cases relied on time left after classes, meetings, school events, and reporting tasks. Tutors in both groups expressed commitment, but commitment did not recover time that had never been protected. This theme explained why completed sessions remained associated with end-of-year scores after fidelity was included.

3.7.3 Coaching supported instructional judgment

Initial orientation clarified roles and forms, but follow-up coaching helped tutors interpret reading errors, adjust pacing, select practice activities, and decide when a learner needed regrouping. Tutors without follow-up support described stronger dependence on the written guide and less confidence when learners did not respond as expected. School-based peer coaching partly compensated for limited external technical assistance, especially where master teachers or reading coordinators reviewed learner work.

3.7.4 Family engagement was necessary but unevenly feasible

Parents generally valued visible reading progress, but attendance support varied with work schedules, transport, household responsibilities, and the clarity of school communication. Families responded more consistently when messages stated the learner's current need, tutorial schedule, recent progress, and one specific home activity. General reminders produced less reliable follow-through. Participants therefore treated family engagement as an implementation function that required usable communication, not as a fixed parent trait.

3.7.5 Geography affected continuity through logistics

Remote and coastal cases described weather, transport availability, communication interruptions, and delayed movement of materials or staff as recurring constraints. These conditions did not prevent implementation in every case. Schools that prepared printed alternatives, adjusted schedules locally, used community contacts, and reported interruptions early preserved more sessions. The qualitative accounts supported the regression finding that accessibility weakened after dosage and fidelity entered the model: geography influenced the mechanisms of delivery rather than determining learner progress by itself.

3.7.6 Monitoring became useful when it triggered action

Implementers distinguished recording from responsive monitoring. Attendance totals and assessment uploads were seen as burdensome when they did not lead to feedback, staffing adjustments, or help with materials. The same data were viewed as valuable when coordinators used them to identify missed sessions, learners with stalled progress, or tutors needing coaching. High-fidelity schools had shorter feedback loops: evidence moved from form to discussion and then to a documented adjustment.

3.8 Integrated findings

The mixed-methods integration produced four central inferences. First, assessment-based grouping was both the highest-scoring fidelity domain and the most consistently described implementation strength. Second, fidelity and dosage independently related to end-of-year reading because a sound plan did not guarantee enough instructional exposure, while frequent sessions without adaptive monitoring could become routine practice without precision. Third, the low parent-support score reflected communication design and household feasibility, not simple caregiver resistance. Fourth, the absence of an adjusted district difference did not remove geographic inequity; it redirected attention toward transport, scheduling, materials, and local continuity measures that schools could partly modify.

Table 8. Joint display of quantitative and qualitative findings

Quantitative result	Qualitative explanation	Integrated inference
Fidelity +0.11 SD per 10 points	High-fidelity schools connected assessment, schedules, coaching, and response.	Delivery quality represented a connected school process, not a checklist total.
Dosage +0.08 SD per 10 sessions	Protected recurring time and rapid rescheduling preserved exposure.	Instructional opportunity remained necessary after planning quality was considered.
Parent/community domain = 69.0	Specific progress messages and feasible home actions worked better than general reminders.	Family engagement required an operational communication design.
Transport/connectivity challenge OR = 2.41 in District I	Remote and coastal schools described recurring logistical interruptions.	Geographic inequity operated through delivery mechanisms that required continuity plans.

Quantitative result	Qualitative explanation	Integrated inference
Adjusted district difference was small	Within-district school routines varied substantially.	School process and dosage explained more actionable variation than district labels.

Note. The display integrated results only where both strands addressed the same implementation process.

4. Discussion

4.1 Principal findings

The evaluation found measurable reading improvement during the first full year of ARAL-Reading implementation in Isabela's First and Second Legislative Districts. The average end-of-year score was 0.44 baseline standard-deviation units above the beginning level, and grade-level readiness increased by 28.1 percentage points. Improvement appeared in every key stage and did not differ clearly by district after clustering and learner composition were considered. These results were consistent with the direction of early national ARAL reports, which had described rising readiness alongside continuing numbers of learners below grade level.[6,7,9]

The central contribution concerned implementation variation. Schools averaged adequate fidelity, but only 46.7% reached the high-fidelity band. Assessment and grouping were strong, whereas parent and community support remained weaker. Fidelity, completed sessions, smaller tutorial groups, and tutor continuity were each associated with end-of-year reading after baseline and context were controlled. The pattern aligned with tutoring evidence that had emphasized structured content, frequency, small groups, consistent tutors, and responsive use of assessment.[21–24]

The absence of random assignment and an untreated comparison prevented a causal estimate of ARAL's effect. Ordinary classroom instruction, maturation, changes in attendance, and other school initiatives could have contributed to improvement. Nevertheless, the graded relationship of fidelity and dosage with end-of-year performance strengthened the implementation argument: schools that delivered more of the intended process and more tutorial exposure tended to show better adjusted outcomes. This evidence supported continued evaluation with stronger counterfactual designs rather than either dismissing or overclaiming the observed progress.

4.2 From statutory compliance to instructional precision

RA No. 12028 required careful learner assessment, trained tutors, accessible delivery, regular progress checks, parental involvement, and annual impact evaluation.[1] The findings showed why these provisions had to be treated as an instructional chain. Correct identification improved targeting; grouping allowed the tutor to work at the learner's present level; protected time produced dosage; coaching helped the tutor respond to errors; and reassessment showed whether instruction should change. A break at one point reduced the value of the others.

This interpretation agreed with Teaching at the Right Level evidence, in which assessment and grouping improved outcomes only when implementers received practical support and a protected organizational structure.[25] It also agreed with fidelity theory: adherence was shaped by facilitation, quality of delivery, and participant responsiveness.[27] For ARAL, a complete learner list was therefore necessary but insufficient. The operational standard should have been whether assessment information changed the instruction a learner received.

4.3 Scheduling, workload, and tutor support

Scheduling pressure was the most widely reported challenge. This did not mean schools rejected the program; it meant ARAL entered an environment with existing teaching loads, school activities, assessment duties, and reporting requirements. The finding echoed national concerns about tutor hiring and dependence on teacher overtime during early scale-up.[8] A sustainable design needed a staffing and calendar solution, not only an appeal to teacher commitment.

Follow-up coaching also distinguished implementation quality. Initial training reached nearly all respondents, but only about two thirds received later coaching. Tutors needed feedback when a learner stalled, a group became heterogeneous, or a guide required adaptation. The finding supported the view that structured materials and professional judgment were complements. Consistency should not have meant unchanged delivery; it should have meant a common instructional sequence with evidence-based adjustment.

4.4 Geographic equity and local adaptation

First District implementers reported transport or connectivity interruption more strongly, which was consistent with the inclusion of remote and coastal municipalities in the district. Yet accessibility did not retain a clear independent coefficient after fidelity and dosage were entered. This result suggested a practical equity strategy: systems could not remove distance, but they could protect delivery through early material distribution, printed alternatives, flexible local scheduling, backup tutors, rapid reporting of interruption, and community-supported learner contact.

Equity should therefore have been monitored as equal opportunity to receive an adequate dose of appropriate instruction, not merely equal issuance of a memorandum or equal distribution of one package. The World Bank's 2026 support for Philippine foundational learning similarly emphasized decentralization, school-level support, inclusive materials, improved assessment, and targeted assistance to schools.[11] ARAL monitoring could have aligned with this direction by displaying reach, dosage, fidelity, and outcomes by accessibility and learner group.

4.5 Family engagement and responsive monitoring

Parent and community support was the weakest fidelity domain. The interviews showed that inconsistent follow-through resulted from a mixture of communication quality and household constraints. A school message that named the learner's current reading need, schedule, progress, and one feasible home action was more usable than a general request for support. This reframed parent engagement as a designable implementation process.

The same principle applied to monitoring. Data collection gained legitimacy when it led to action. Forms that only moved upward increased burden; information that returned as coaching, staffing help, material support, or schedule adjustment created a feedback loop. The ASRRA's beginning-, middle-, and end-year structure provided an appropriate foundation for adaptive implementation, but responsiveness depended on how quickly each governance level acted on the evidence.[3]

4.6 Statistical and substantive interpretation

The multilevel structure materially affected inference. Ten percent of baseline reading variance occurred between schools, so a single-level analysis would have treated correlated learner results as independent. Restricted maximum likelihood with Kenward-Roger inference was appropriate for the 60-school sample and protected against overly narrow fixed-effect intervals.[34,35] Cluster bootstrap intervals similarly preserved school dependence for readiness change.

Effect sizes were interpreted in the scale on which decisions could be made. A 10-point fidelity difference corresponded to 0.11 SD in end-year performance, and 10 additional sessions corresponded to 0.08 SD. Neither estimate represented a guaranteed gain for an individual learner. They described average conditional differences across the observed implementation range. The estimates were smaller than the average experimental tutoring effect reported in meta-analysis, which was expected because routine implementation varied in dosage, group size, and organizational support.[21]

4.8 Policy and implementation recommendations

4.81. Protect a minimum instructional dose

Each school should have translated the national schedule into recurring protected tutorial periods, named backup tutors, and a documented same-week rescheduling rule. Division monitoring should have reported completed sessions per learner and the proportion receiving the intended minimum dose, not only the number enrolled.

4.8.2 Institutionalize coaching after orientation

Tutor preparation should have included short coaching cycles built around learner errors, grouping, pacing, and progress data. Master teachers, reading coordinators, and trained school leaders could have served as local coaches, with division specialists assigned to schools that showed stalled progress or weak delivery fidelity.

4.8.3 Use ASRRA as an action trigger

Readiness and responsiveness findings should have been connected to named corrective actions, responsible persons, resources, and review dates. Dashboards should have displayed exceptions that required support, including missed sessions, delayed reassessment, high tutor turnover, weak parent contact, and persistent non-progression.

4.8.4 Create continuity packages for remote and coastal schools

The package should have included early delivery of printed resources, offline assessment and reporting options, locally authorized schedule adjustments, backup tutor arrangements, and a rapid communication route for transport or weather disruption. Equity funding should have followed documented continuity risk rather than equal per-school distribution alone.

4.8.5 Make family communication specific and feasible

Schools should have used short progress messages that stated the learner's present need, recent movement, next tutorial schedule, and one realistic home practice. Communication logs should have measured whether families received and understood the message, while referrals should have addressed attendance barriers that families could not solve alone.

4.8.6 Strengthen the next evaluation design

Future evaluation should have used a pre-specified stepped implementation, matched comparison, or regression-discontinuity opportunity when eligibility thresholds and ethical service delivery allowed. The core outcome set should have retained baseline, middle, and end-year reading measures, completed sessions, group size, tutor continuity, fidelity domains, learner attendance, and subgroup equity.

5. Conclusion

ARAL-Reading implementation in Isabela's First and Second Legislative Districts progressed beyond policy issuance and produced a measurable school-year improvement in reading performance. The strongest implementation features were assessment-based identification, learner grouping, aligned materials, and progress review. The most persistent weaknesses involved protected tutorial time, reporting burden, family follow-through, and continuity in geographically difficult settings.

Fidelity and completed sessions remained positively associated with end-of-year reading after baseline and contextual differences were controlled. Smaller tutorial groups and tutor continuity also mattered. The evidence supported an implementation strategy in which assessment, grouping, protected time, coaching, family communication, and responsive monitoring operated as one delivery chain. Legislative district alone explained little after those processes were considered.

The principal policy lesson was that access to ARAL should have been judged by the instruction learners actually received, not only by enrollment in the program. A national mandate created entitlement and direction; school routines, local adaptation, and feedback determined whether that mandate became sustained reading support.

References

[1] Congress of the Philippines. Republic Act No. 12028: Academic Recovery and Accessible Learning (ARAL) Program Act. 2024. https://lawphil.net/statutes/repacts/ra2024/ra_12028_2024.html

[2] Department of Education. DepEd Order No. 018, s. 2025: Implementing Guidelines of the Academic Recovery and Accessible Learning Program. June 13, 2025. https://www.deped.gov.ph/2025/06/13/june-13-2025-do-018-s-2025-implementing-guidelines-of-the-academic-recovery-and-accessible-learning-program/
[3] Department of Education. DepEd Memorandum No. 056, s. 2025: Supplemental Guidelines on the Implementation of the ARAL School Readiness and Responsiveness Audit. July 18, 2025. https://www.deped.gov.ph/2025/07/18/july-18-2025-dm-056-s-2025-supplemental-guidelines-on-the-implementation-of-the-aral-school-readiness-and-responsiveness-audit/
[4] Department of Education. DepEd Memorandum No. 064, s. 2025: Implementing Guidelines for the ARAL Program in Reading for Key Stages 1 to 3. August 1, 2025. https://www.deped.gov.ph/2025/08/01/august-1-2025-dm-064-s-2025-implementing-guidelines-for-the-academic-recovery-and-accessible-learning-program-in-reading-for-key-stages-1-to-3/
[5] Department of Education. DepEd launched ARAL Program to strengthen reading and learning. September 13, 2025. https://www.deped.gov.ph/2025/09/13/deped-naglunsad-ng-aral-program-para-palakasin-ang-pagbasa-at-pagkatuto/
[6] Department of Education. ARAL Program showed promising initial learning-recovery results and plans for expansion in 2026. January 7, 2026. https://www.deped.gov.ph/2026/01/07/deped-aral-program-promising-ang-initial-result-sa-learning-recovery-plano-pa-palakasin-sa-2026/
[7] Second Congressional Commission on Education. ARAL Program showed early wins, with top schools reporting improvements in reading proficiency. January 25, 2026. https://edcom2.gov.ph/aral-program-shows-early-wins-with-top-schools-reporting-30-40-increases-in-students-reading-proficiency/
[8] Second Congressional Commission on Education. EDCOM urged DepEd to address delays in hiring ARAL tutors. May 26, 2026. https://edcom2.gov.ph/edcom-urges-deped-to-address-delays-in-hiring-of-aral-tutors/
[9] Second Congressional Commission on Education. Improvements were seen, but half of learners remained below grade level in literacy. 2026. https://edcom2.gov.ph/improvements-seen-but-half-of-learners-still-below-grade-level-in-literacy/
[10] Second Congressional Commission on Education. Turning Point: A Decade of Necessary Reforms and the National Education and Workforce Development Plan 2026-2035. 2026. https://edcom2.gov.ph/edcom-ii-releases-turning-point-report-unveils-10-year-national-roadmap-to-reverse-education-crisis/
[11] World Bank. World Bank backed better learning for 21 million Filipino students through PLUS-D. April 3, 2026. https://www.worldbank.org/en/news/press-release/2026/04/03/world-bank-backs-better-learning-for-21-million-filipino-students
[12] Organisation for Economic Co-operation and Development. PISA 2022 Results, Country Note: Philippines. 2023. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/philippines_a0882a2d-en.html
[13] UNICEF Philippines. SEA-PLM 2024 revealed more Philippine learners reaching higher proficiency in mathematics and reading, with disparities remaining. 2025. https://www.unicef.org/philippines/press-releases/latest-sea-plm-2024-reveals-more-learners-philippines-reaching-higher-proficiency
[14] Department of Education, SEAMEO, and UNICEF. Southeast Asia Primary Learning Metrics 2019 National Report of the Philippines. 2021. https://www.unicef.org/philippines/reports/sea-plm-metrics-2019-national-report-philippines
[15] Bautista MCRB, Aranas MVP. The Learning Crisis in Philippine Education: An Overview. PIDS Policy Notes No. 2023-17. 2023. https://doi.org/10.62986/pn2023.17
[16] Philippine Institute for Development Studies. COVID-19 school closures led to significant learning losses, according to evidence presented by a World Bank expert. 2024. https://www.pids.gov.ph/details/news/press-releases/covid-19-school-closures-led-to-significant-learning-losses-says-world-bank-expert
[17] Congress of the Philippines. Republic Act No. 11080: An Act Reapportioning the Province of Isabela into Six Legislative Districts. 2018. https://lawphil.net/statutes/repacts/ra2018/ra_11080_2018.html
[18] Philippine Statistics Authority. Province of Isabela: Philippine Standard Geographic Code profile and 2024 population. 2026. https://psa.gov.ph/classification/psgc/citimuni/0203100000
[19] Philippine Statistics Authority. Highlights of the Region II Population, 2024 Census of Population. 2025. https://psa.gov.ph/statistics/population-and-housing/node/1684077888
[20] Philippine Statistics Authority - Isabela. Key economic indicators presented at the Provincial Government of Isabela joint meeting. 2025. https://rsso02.psa.gov.ph/content/psa-isabela-presents-key-economic-indicators-pgis-1st-quarter-joint-meeting
[21] Nickow A, Oreopoulos P, Quan V. The Promise of Tutoring for PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. American Educational Research Journal. 2024;61(1). https://doi.org/10.3102/00028312231208687
[22] Kraft MA, Falken GT. A Blueprint for Scaling Tutoring and Mentoring Across Public Schools. AERA Open. 2021;7. https://doi.org/10.1177/23328584211042858
[23] Cortes K, Kortecamp K, Loeb S, Robinson CD. A Scalable Approach to High-Impact Tutoring for Young Readers. Learning and Instruction. 2025;95:102021. https://doi.org/10.1016/j.learninstruc.2024.102021
[24] Dietrichson J, Filges T, Seerup JK, et al. Targeted School-Based Interventions for Improving Reading and Mathematics for Students with or at Risk of Academic Difficulties in Grades K-6: A Systematic Review. Campbell Systematic Reviews. 2021;17:e1152. https://doi.org/10.1002/cl2.1152
[25] Banerjee A, Banerji R, Berry J, et al. Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of Teaching at the Right Level in India. NBER Working Paper 22746. 2016. https://doi.org/10.3386/w22746
[26] Duflo E, Dupas P, Kremer M. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. American Economic Review. 2011;101(5):1739-1774. https://doi.org/10.1257/aer.101.5.1739
[27] Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A Conceptual Framework for Implementation Fidelity. Implementation Science. 2007;2:40. https://doi.org/10.1186/1748-5908-2-40

[28] Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The Updated Consolidated Framework for Implementation Research Based on User Feedback. Implementation Science. 2022;17:75. https://doi.org/10.1186/s13012-022-01245-0
[29] Proctor E, Silmere H, Raghavan R, et al. Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. Administration and Policy in Mental Health. 2011;38:65-76. https://doi.org/10.1007/s10488-010-0319-7
[30] Polit DF, Beck CT, Owen SV. Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. Research in Nursing and Health. 2007;30(4):459-467. https://doi.org/10.1002/nur.20199
[31] Aiken LR. Three Coefficients for Analyzing the Reliability and Validity of Ratings. Educational and Psychological Measurement. 1985;45(1):131-142. https://doi.org/10.1177/0013164485451012
[32] Li CH. Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares. Behavior Research Methods. 2016;48:936-949. https://doi.org/10.3758/s13428-015-0619-7
[33] Hayes AF, Coutts JJ. Use Omega Rather than Cronbach's Alpha for Estimating Reliability. Communication Methods and Measures. 2020;14(1):1-24. https://doi.org/10.1080/19312458.2020.1718629
[34] McNeish D. Small Sample Methods for Multilevel Modeling: A Colloquial Elucidation of REML and the Kenward-Roger Correction. Multivariate Behavioral Research. 2017;52(5):661-670. https://doi.org/10.1080/00273171.2017.1344538
[35] Kenward MG, Roger JH. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics. 1997;53(3):983-997. https://pubmed.ncbi.nlm.nih.gov/9333350/
[36] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1-67. https://doi.org/10.18637/jss.v045.i03
[37] Nakagawa S, Schielzeth H. A General and Simple Method for Obtaining R Squared from Generalized Linear Mixed-Effects Models. Methods in Ecology and Evolution. 2013;4:133-142. https://doi.org/10.1111/j.2041-210x.2012.00261.x
[38] Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979;6(2):65-70. https://www.jstor.org/stable/4615733
[39] Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the Framework Method for the Analysis of Qualitative Data in Multidisciplinary Health Research. BMC Medical Research Methodology. 2013;13:117. https://doi.org/10.1186/1471-2288-13-117
[40] Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. Qualitative Health Research. 2016;26(13):1753-1760. https://doi.org/10.1177/1049732315617444
[41] Fetters MD, Curry LA, Creswell JW. Achieving Integration in Mixed Methods Designs: Principles and Practices. Health Services Research. 2013;48(6 Pt 2):2134-2156. https://doi.org/10.1111/1475-6773.12117
[42] McHugh ML. Interrater Reliability: The Kappa Statistic. Biochemia Medica. 2012;22(3):276-282. https://pmc.ncbi.nlm.nih.gov/articles/PMC3900052/
[43] World Bank, Bill and Melinda Gates Foundation, FCDO, UNESCO, UNICEF, and USAID. Guide for Learning Recovery and Acceleration: Using the RAPID Framework to Address Learning Losses and Build Forward Better. 2022. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099063023145523057