JENER Journal of Empirical and Non-Empirical Research
Volume 2, Issue 8, Page 726-735 | Article No. 79
DOI: 10.4898/jener.v2i8.a79
ISSN: 7895-9216
Journal homepage: https://jenerjournal.com
Journal email: editor@jenerjournal.com

JENER Journal
JENER Journal of Empirical and Non-Empirical Research

GEN-LINK for Grade 8 Genetics: Development and Validation of an Integrated Science-English Strategy for Conceptual Understanding and Scientific Communication at Quezon National High School, Nueva Vizcaya, Philippines

Divina Gracia P. Arzadon[1] & Marites B. Fernandez[2]

[1,2]Northeastern College

ARTICLE INFORMATION

Article history:
Published: August 2026

Keywords:
Genetics Education
Integrated Curriculum
Scientific Communication
Conceptual Understanding
Grade 8
Philippines

ABSTRACT

This study developed and validated GEN-LINK, an integrated Science-English strategy for improving Grade 8 learners' conceptual understanding and scientific communication in genetics. It combined phenomenon-based engagement, multimodal reading, explicit language support, linked representations, evidence interpretation, structured talk, and claim-evidence-reasoning writing. A developmental phase was followed by a randomized pretest-posttest validation with 84 learners at Quezon National High School, Nueva Vizcaya, Philippines. Learners were allocated equally to GEN-LINK and time-matched comparison instruction. Seven experts examined relevance, clarity, coherence, and feasibility. Outcomes were measured through a 30-point Genetics Conceptual Understanding Test and a 24-point Scientific Communication Rubric. The strategy guide obtained strong content-validity evidence (scale-level content validity index/average = .94; Aiken's V = .89-.98). Reliability was strong (KR-20 = .86; rubric omega = .88), and scoring showed excellent inter-rater agreement (ICC[2,k] = .91, 95% CI [.84, .95]). After adjustment for pretest and section, GEN-LINK learners achieved higher conceptual-understanding scores (adjusted mean difference = 2.98, 95% CI [1.84, 4.12], F[1,78] = 27.01, $p < .001$, partial eta squared = .26) and scientific-communication scores (adjusted mean difference = 2.73, 95% CI [1.91, 3.55], F[1,78] = 43.92, $p < .001$, partial eta squared = .36). Fidelity averaged 88.4%, and acceptability was high (M = 4.36/5). Integrating disciplinary language with genetics instruction strengthened both what learners understood and how clearly they communicated scientific explanations. Multisite replication with delayed outcomes was recommended.

1. Introduction

Genetics is a central strand of biological literacy because it allows learners to connect inheritance, variation, cell division, molecular information, probability, and evidence-based decisions. In the Philippine Grade 8 Science curriculum, learners were expected to explain meiosis, maintain chromosome-number reasoning, and predict phenotypic expression under simple patterns of inheritance.[1] These expectations required more than recalling vocabulary or completing Punnett squares. Learners needed to coordinate entities and processes that operate at organismal, cellular, chromosomal, and molecular levels, then communicate the resulting explanation in language precise enough to distinguish probability from certainty.

This instructional challenge was consequential in a national context where science and reading performance remained low. In PISA 2022, Filipino 15-year-olds averaged 356 points in science and 347 points in reading, both substantially below the corresponding OECD averages.[3] The figures did not diagnose the performance of any single school, yet they underscored a system-wide need for classroom designs that treat scientific knowledge and disciplinary language as mutually reinforcing rather than separate concerns. Quezon National High School was established in the Municipality of Quezon, Province of Nueva Vizcaya, under Batas Pambansa Blg. 188.[2] The site therefore represented an appropriate public-school setting for a locally usable intervention aligned with national curriculum expectations.

1.1 Genetics as a multilevel conceptual problem

Research has repeatedly shown that genetics is difficult because explanations must move across ontologically different levels. A learner may recognize the words gene, DNA, chromosome, protein, and trait but still fail to explain how these entities are related or how a change at one level produces an observable outcome at another.[11] A learning progression for Grades 5-10 accordingly emphasized coordinated inheritance, meiotic, and molecular models instead of isolated terminology.[12] Carefully sequenced project-based and inquiry-rich instruction has helped learners develop more mechanistic accounts of genetic phenomena.[13,14]

Middle-school learners have also demonstrated that they can reason productively about inheritance, meiosis, and molecular mechanisms when instruction makes the relations among models visible.[15] Persistent errors have included treating genes, DNA, and chromosomes as interchangeable; interpreting dominant as stronger or more common; believing Punnett-square outcomes are certain; and directly linking genes to traits without mediating mechanisms or environmental interaction.[16,17] These difficulties suggested that an effective Grade 8 strategy should repeatedly link words, diagrams, physical models, data displays, and causal sentences.

1.2 Scientific communication as a mechanism for learning

Scientific literacy includes the capacity to read and write scientific texts, interpret representations, and participate in disciplined explanation.[6] Science practices create language demands but also provide authentic opportunities for language development: learners name observations, compare patterns, qualify claims, justify interpretations, and revise explanations.[7] The language of science is densely informational and often uses technical nouns, nominalizations, passive constructions, and logical connectors that differ from everyday talk.[22] Without explicit support, learners may know part of the science yet be unable to express causal or probabilistic relationships with sufficient precision.

Argumentation and collaborative critical discourse can deepen conceptual learning because learners must make their reasoning public, inspect the adequacy of evidence, and respond to alternatives.[8] The claim-evidence-reasoning framework operationalizes this process by separating the conclusion, the data supporting it, and the scientific principle that warrants the connection. Middle-school studies have shown that explicit and gradually faded scaffolds improve learners' capacity to construct scientific explanations when support is later removed.[9] Writing-to-learn approaches likewise make relationships among ideas visible and create opportunities for metacognitive revision.[23]

1.3 Rationale for integrated Science-English instruction

Integrated science-literacy instruction has produced gains because reading, writing, discussion, and investigation are organized around the same conceptual problem. A large upper-elementary study found that an integrated unit promoted science learning while engaging learners in texts, notes, firsthand investigation, and discussion.[18] Randomized work with middle-grade learners also reported positive effects when inquiry lessons embedded explicit vocabulary, reading, and writing supports.[19] Related longitudinal evidence indicated that science and English-language development can be mutually supportive when literacy activities serve content learning instead of displacing it.[20]

The integration in the present study was therefore disciplinary rather than merely thematic. English activities were selected because they performed a scientific function: unpacking a dense explanation, comparing meanings of related terms, interpreting probability language, rehearsing a causal chain orally, and composing a concise explanation from evidence. This distinction mattered because adding unrelated reading or writing could increase lesson time without strengthening the target concept. The strategy preserved the Grade 8 genetics competencies while embedding language objectives within the same assessment evidence.

1.4 Research gap and contribution

Existing studies established the promise of science-literacy integration, but much of the experimental evidence concerned elementary grades or English learners outside the Philippines.[18-21] Genetics studies, meanwhile, offered strong guidance about learning progressions and representations but did not directly test a locally developed Science-English routine at the Grade 8 level in Nueva Vizcaya.[12-15] The gap was both substantive and methodological: a useful local strategy needed an explicit theory of action, documented development decisions, content-valid instruments, reliable scoring, a time-matched comparison, baseline-adjusted outcome analysis, and implementation-fidelity evidence.

GEN-LINK was developed to address that gap. Its name represented six recurring instructional functions: Engage, Navigate, Link, Interpret, Negotiate, and Keep and refine. The contribution of the study lay not only in estimating outcome differences but also in specifying a replicable instructional sequence and measurement system that a school could review, adapt, and test.

1.5 Purpose, research questions, and hypotheses

The study aimed to develop and validate an integrated Science-English teaching strategy for improving Grade 8 learners' conceptual understanding and scientific communication in genetics at Quezon National High School.

- What curriculum, conceptual, language, and assessment features were incorporated into the GEN-LINK strategy?
- What evidence of relevance, clarity, coherence, feasibility, and instrument quality was obtained during expert review and pilot testing?
- After adjustment for baseline performance and section, did GEN-LINK learners outperform learners who received time-matched conventional genetics instruction in conceptual understanding and scientific communication?
- How consistently was GEN-LINK implemented, and how acceptable did learners find its components?

The primary null hypotheses stated that, after adjustment for corresponding pretest scores and section, there would be no difference between groups in posttest conceptual understanding or scientific communication. The two outcomes were treated as co-primary and tested with Holm control of the familywise error rate.

2. Methods

2.1 Design and setting

The study used a multiphase development-and-validation design. Phase 1 involved needs specification, design, expert review, revision, and small-group pilot testing. Phase 2 used a stratified randomized pretest-posttest control-group design to examine classroom effectiveness. The study was situated at Quezon National High School in the Municipality of Quezon, Nueva Vizcaya, Philippines, a public secondary school established under Batas Pambansa Blg. 188.[2] The intervention addressed the Grade 8 sequence on cellular reproduction, meiosis, and Mendelian genetics reflected in DepEd learning resources.[1]

The outcome trial lasted eight weeks and comprised sixteen 60-minute sessions. To protect instructional-time equivalence, both conditions completed the same genetics content, number of sessions, performance tasks, and summative schedule. The conditions differed in how language and conceptual reasoning were scaffolded. The study design emphasized credible baseline adjustment and transparent reporting of effect sizes, consistent with guidance for pretest-posttest comparisons and educational intervention research.[28,31]

2.2 Participants, allocation, and sample-size basis

Eighty-four Grade 8 learners from four intact sections participated. Each section contributed 21 learners. Within section, learners were allocated in equal overall numbers to GEN-LINK (n = 42) and comparison instruction (n = 42) using stratified random assignment. Allocation balanced instructional section while preserving each learner's regular enrollment. The same science teacher delivered both conditions through a counterbalanced timetable, used common content targets, and avoided sharing intervention materials during comparison sessions. An English teacher contributed to material development and teacher preparation but did not deliver a separate English intervention.

The target sample was selected to detect a standardized between-group effect of approximately 0.62 with 80% power at a two-sided alpha of .05 under a balanced design; baseline covariates were expected to improve precision.[32] Learners were eligible if they were enrolled in Grade 8, attended the genetics unit, and completed the pretest. The analytic dataset retained all 84 allocated learners, producing no differential attrition.

2.3 Development process

Development followed five iterative functions: analysis, design, production, formative review, and revision. Curriculum analysis mapped the DepEd competencies to prerequisite concepts, recurring misconceptions, language functions, and assessment evidence. The design phase converted this map into paired content-language objectives. Production yielded a teacher guide, eight lesson sets, learner reading cards, representation mats, discussion prompts, claim-evidence-reasoning frames, assessment tasks, and a fidelity checklist. Formative review used expert ratings and written comments; revision eliminated ambiguous genetic language, reduced reading load, improved diagram-text alignment, and sequenced supports from explicit to partially faded.

Table 1. GEN-LINK design specifications and intended learning mechanisms

Design feature	Classroom enactment	Intended mechanism
Paired objectives	Every lesson stated a genetics objective and a language function.	Kept language activity accountable to conceptual learning.
Phenomenon anchor	A familiar inheritance pattern, family resemblance case, or variation problem initiated inquiry.	Activated prior knowledge and created a reason to explain.
Multimodal navigation	Learners annotated short text, diagrams, Punnett squares, pedigrees, and data tables.	Reduced fragmentation across verbal and visual representations.
Concept linking	Representation mats connected cell, chromosome, DNA, gene, allele, meiosis, and trait.	Supported movement across biological organization levels.
Language scaffolds	Word families, contrast frames, probability markers, causal connectors, and sentence starters were used.	Made disciplinary meanings and relationships explicit.
Structured discourse	Think-pair-explain and evidence circles required claims, evidence, reasoning, and peer questions.	Externalized reasoning for inspection and revision.
Faded writing support	Full explanation frames became partial prompts and then independent responses.	Promoted transfer beyond the scaffold.
Formative feedback	Exit explanations were coded for concept and language needs before the next lesson.	Enabled responsive reteaching and targeted revision.

2.4 The GEN-LINK instructional cycle

The strategy repeated a six-part cycle. Engage introduced a problem that demanded a genetic explanation. Navigate guided learners through a concise text-representation set. Link required learners to connect terms and models across levels. Interpret used data, crosses, or family evidence while emphasizing probabilistic language. Negotiate organized peer talk around claim, evidence, and reasoning. Keep and refine required oral or written revision, with scaffolds progressively reduced. This cycle drew on research showing that scientific literacy requires work with text, representations, discourse, and explanation, not vocabulary recall alone.[6-9,22]

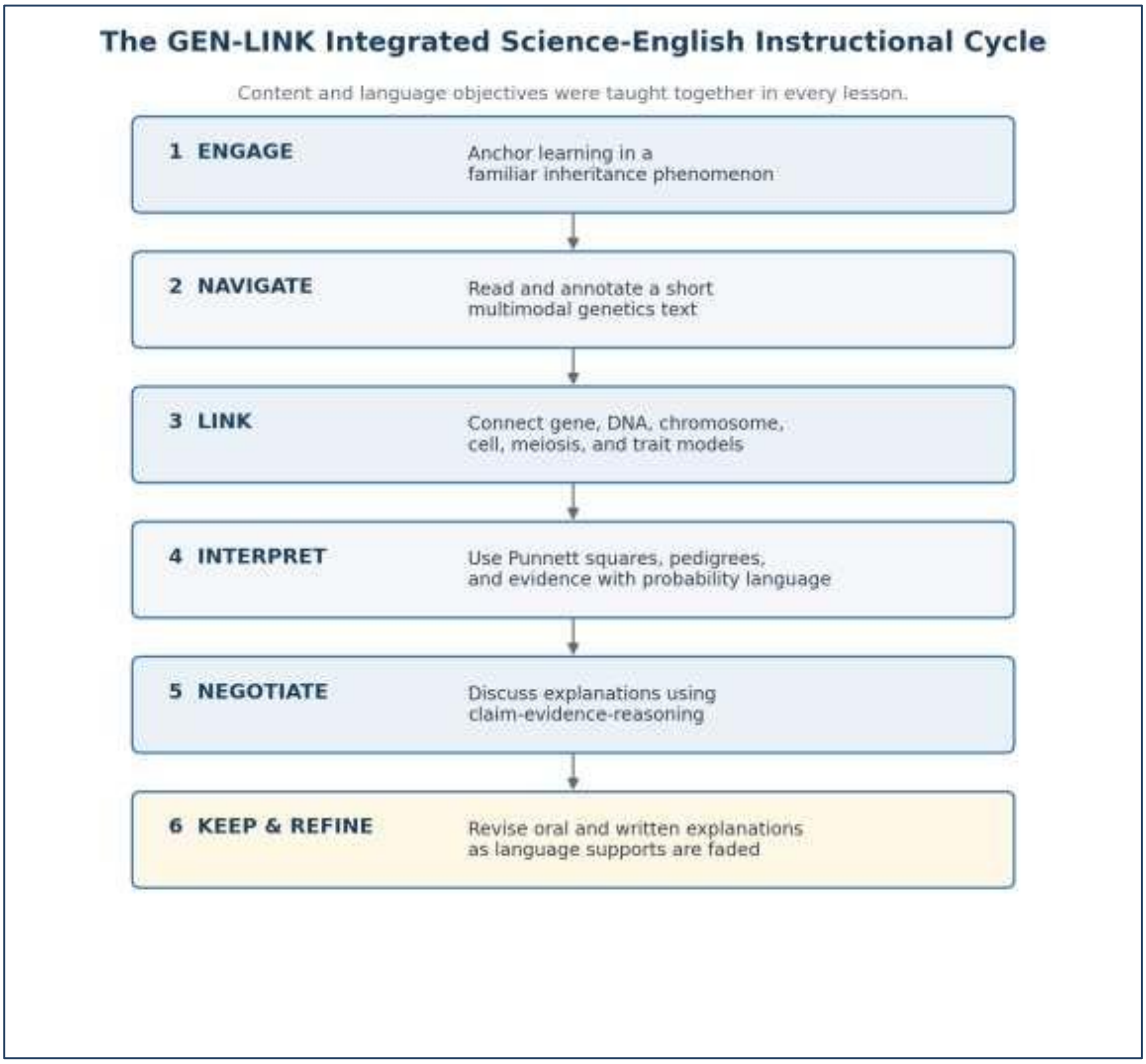

Figure 1. The GEN-LINK integrated instructional cycle.

Table 2. Eight-week instructional sequence

Week	Genetics focus	Scientific communication focus	Major evidence
1	Cell, nucleus, chromosome, DNA, gene, allele	Definition, classification, and part-whole language	Annotated representation map
2	Mitosis, meiosis, chromosome number	Sequence and causal connectors	Oral process explanation
3	Gametes, segregation, fertilization	Conditionals and probability markers	Evidence table with interpretation
4	Mendelian inheritance and Punnett squares	Claim-evidence-reasoning paragraph	Worked cross plus explanation
5	Genotype, phenotype, dominance	Contrast and qualification	Misconception-refutation note
6	Pedigrees and family-pattern evidence	Questioning and evidence sufficiency	Small-group evidence conference
7	Gene-environment interaction and variation	Multicausal explanation	Independent explanatory response
8	Synthesis and transfer	Concise oral and written scientific communication	Performance task and posttest

2.5 Comparison instruction

The comparison condition followed the same competencies and instructional time through teacher explanation, textbook or module reading, guided examples, individual exercises, and review. Learners completed Punnett-square practice and answered content questions. The condition did not use the GEN-LINK cycle, paired language objectives, representation mats, structured evidence circles, or progressively faded explanation frames. Common pretests and posttests were administered under identical directions.

2.6 Measures

2.6.1 Genetics Conceptual Understanding Test

The Genetics Conceptual Understanding Test (GCUT) contained 30 two-tier items. Each item paired a selected response with a reason; one point was awarded only when both were correct, limiting credit for guessing. The blueprint included conceptual relations among gene, DNA, and chromosome (8 points); meiosis and inheritance (8 points); Punnett-square and probability reasoning (8 points); and mechanism, variation, and gene-environment reasoning (6 points). Total scores ranged from 0 to 30. Distractors represented documented misconceptions rather than arbitrary incorrect statements.[11,15-17]

2.6.2 Scientific Communication Rubric

The Scientific Communication Rubric for Genetics (SCRG) scored an evidence-based explanation task across four six-point domains: scientific accuracy, selection and use of evidence, causal or probabilistic reasoning, and disciplinary language and organization. Total scores ranged from 0 to 24. Two raters independently scored pilot responses and a randomly selected 30% of trial responses while blinded to condition and time point. Disagreements greater than one scale point were resolved after independent scoring was preserved.

2.6.3 Fidelity and acceptability measures

A 20-indicator observation checklist assessed adherence to lesson sequence, paired objectives, representation linking, structured discourse, scaffold fading, formative checks, and time allocation. Each indicator was coded 0 (not observed), 1 (partly observed), or 2 (fully observed), then converted to a percentage. Eight of 16 sessions were independently observed. A 16-item learner acceptability scale measured clarity, usefulness, engagement, and manageability on a five-point response format. Attendance and completion of major tasks provided dose and receipt indicators.

2.7 Expert validation and pilot testing

Seven validators were selected to represent biology education, English or language education, curriculum and instruction, and educational measurement. They rated strategy materials and instruments on four-point relevance and clarity scales and supplied written recommendations. Item-level content validity indices (I-CVI) and the scale-level average (S-CVI/Ave) were computed. I-CVI values of .78 or higher were treated as acceptable for the panel size, while S-CVI/Ave values at or above .90 supported strong scale-level content validity.[25] Aiken's V quantified the degree of ordered agreement in ratings.[24]

The GCUT was piloted with 36 Grade 8 learners outside the outcome sample. Item difficulty, corrected point-biserial discrimination, and Kuder-Richardson 20 internal consistency were examined because the final score was dichotomous.[26] Items outside the intended difficulty range or with weak discrimination were revised and repiloted. The SCRG was applied to 30 pilot explanations. Internal consistency was summarized with McDonald's omega, and inter-rater agreement was estimated using a two-way random-effects, absolute-agreement ICC for the mean of two raters, reported as ICC(2,k).[27,33]

2.8 Procedure and quality safeguards

Teacher preparation consisted of two workshops and one rehearsal lesson. The workshops clarified the genetics content model, demonstrated each GEN-LINK routine, calibrated scoring with anchor responses, and planned time-equivalent sessions. Learners completed the GCUT and SCRG performance task one week before instruction. The eight-week unit then proceeded according to the assigned condition. Posttests used parallel prompts and were administered within three school days of the final lesson.

Several safeguards reduced alternative explanations. Allocation was stratified by section; learning time and content targets were matched; the same teacher taught both conditions; test directions and scoring guides were standardized; raters were blinded to condition; fidelity observations verified delivery; and comparison materials were stored separately until posttesting. Treatment-fidelity procedures addressed design, teacher preparation, delivery, receipt, and enactment.[30]

2.9 Statistical analysis

Descriptive statistics summarized scores, validation evidence, fidelity, and acceptability. Baseline balance was described with standardized mean differences rather than null-hypothesis tests; absolute values below 0.10 indicated close balance. For each co-primary outcome, an analysis of covariance predicted posttest score from condition and the corresponding centered pretest score, with fixed effects for instructional section. ANCOVA was selected because baseline-adjusted analysis is generally more efficient and interpretable than comparing raw change scores alone.[28]

Ypost = beta0 + beta1(GEN-LINK) + beta2(Ypre) + section fixed effects + error

Homogeneity of regression slopes was checked by adding the condition-by-pretest interaction. Residual normality and homoscedasticity were examined using the Shapiro-Wilk and median-centered Levene tests, supplemented by residual inspection. The condition coefficient was reported as an adjusted mean difference with a 95% confidence interval, F statistic, p value, and partial eta squared. Hedges' g for change-score differences was reported as a sensitivity effect size with small-sample correction.[29] Holm adjustment controlled the two co-primary tests. Statistical significance was set at an adjusted alpha of .05; estimation and practical magnitude remained central to interpretation.

3. Results

3.1 Product development and validation

The development phase produced a 74-page teacher guide, eight lesson sets, 24 learner reading and representation cards, eight explanation tasks, a 30-item conceptual test, a scientific communication rubric, and fidelity and acceptability tools. Expert comments led to three substantive revisions. First, gene, allele, DNA, and chromosome were consistently represented as related but noninterchangeable concepts. Second, dominance statements were rewritten to avoid implying strength, desirability, or population frequency. Third, probability expressions such as may, is likely to, and one expected outcome were added beside Punnett-square examples.

Table 3. Validation and reliability evidence

Material or measure	Evidence	Estimate	Interpretation
GEN-LINK guide	Relevance S-CVI/Ave	.96	Strong content relevance
GEN-LINK guide	Clarity S-CVI/Ave	.93	Strong clarity

Material or measure	Evidence	Estimate	Interpretation
GEN-LINK guide	Coherence S-CVI/Ave	.95	Strong alignment
GEN-LINK guide	Feasibility S-CVI/Ave	.91	Feasible with minor revisions
GEN-LINK guide	Aiken's V range	.89-.98	High ordered agreement
GCUT	I-CVI / S-CVI/Ave	.86-1.00 / .95	Adequate item and scale content validity
GCUT	Difficulty / discrimination	.34-.78 / .25-.58	Useful spread and positive discrimination
GCUT	KR-20	.86	Strong internal consistency
SCRG	I-CVI / S-CVI/Ave	.86-1.00 / .96	Adequate item and scale content validity
SCRG	McDonald's omega	.88	Strong internal consistency
SCRG	ICC(2,k), 95% CI	.91 [.84, .95]	Excellent inter-rater reliability
Acceptability scale	McDonald's omega	.90	Strong internal consistency

Note. I-CVI = item-level content validity index; S-CVI/Ave = scale-level content validity index averaged across items; GCUT = Genetics Conceptual Understanding Test; SCRG = Scientific Communication Rubric for Genetics; ICC(2,k) = two-way random, absolute-agreement intraclass correlation for mean ratings.

3.2 Participant profile and baseline balance

The 84 learners had a mean age of 13.67 years (SD = 0.46); 44 (52.4%) were female and 40 (47.6%) were male. Each of the four sections contributed 21 learners, and each condition included 42 learners. Baseline standardized mean differences were 0.03 for conceptual understanding and 0.02 for scientific communication, indicating close balance.

Table 4. Baseline characteristics by condition

Characteristic	Comparison (n = 42)	GEN-LINK (n = 42)	Standardized difference
Age, years, M (SD)	13.66 (0.45)	13.68 (0.47)	0.04
Female, n (%)	22 (52.4)	22 (52.4)	0.00
GCUT pretest, M (SD)	11.26 (3.30)	11.38 (3.60)	0.03
SCRG pretest, M (SD)	8.39 (1.81)	8.44 (2.18)	0.02
Sessions attended, M (%)	15.1 (94.4)	15.2 (95.0)	0.04

3.3 Descriptive changes in the two outcomes

Both groups improved, but improvement was larger under GEN-LINK. Mean GCUT scores increased by 9.29 points in the GEN-LINK group and 6.33 points in the comparison group. Mean SCRG scores increased by 8.21 and 5.49 points, respectively. The standardized difference in change favored GEN-LINK for conceptual understanding (Hedges' g = 1.06) and scientific communication (Hedges' g = 1.41). These sensitivity estimates were interpreted alongside the baseline-adjusted models rather than as replacements for them.

Table 5. Pretest, posttest, and change scores

Outcome / condition	Pretest M (SD)	Posttest M (SD)	Change M (SD)	Hedges' g for change
GCUT - comparison	11.26 (3.30)	17.60 (3.58)	6.33 (2.54)	
GCUT - GEN-LINK	11.38 (3.60)	20.67 (3.65)	9.29 (2.98)	1.06
SCRG - comparison	8.39 (1.81)	13.88 (2.38)	5.49 (2.16)	
SCRG - GEN-LINK	8.44 (2.18)	16.65 (2.35)	8.21 (1.63)	1.41

Note. Positive Hedges' g values favored GEN-LINK. Paired changes were significant within all four condition-outcome combinations (all p < .001), but between-condition estimates were used for effectiveness conclusions.

3.4 Baseline-adjusted effectiveness

The conceptual-understanding model satisfied the homogeneity-of-slopes test (interaction p = .640), and its residuals did not depart detectably from normality (W = .978, p = .152) or equal variance (Levene p = .899). The adjusted posttest mean was 20.62 for GEN-LINK and 17.64 for comparison instruction. The adjusted difference of 2.98 points was statistically significant and educationally substantial, F(1,78) = 27.01, p < .001, partial eta squared = .26.

The scientific-communication model likewise satisfied the slope requirement (interaction p = .511); residual normality (W = .979, p = .189) and equal variance (Levene p = .092) were acceptable. Adjusted means were 16.63 for GEN-LINK and 13.90 for comparison instruction. The 2.73-point difference favored GEN-LINK, F(1,78) = 43.92, p < .001, partial eta squared = .36. Both co-primary findings remained significant after Holm adjustment.

Table 6. ANCOVA estimates for posttest outcomes

Outcome	Adjusted mean: comparison	Adjusted mean: GEN-LINK	Adjusted difference (95% CI)	F(1,78)	p	Partial eta squared
Conceptual understanding	17.64	20.62	2.98 (1.84, 4.12)	27.01	<.001	.26
Scientific communication	13.90	16.63	2.73 (1.91, 3.55)	43.92	<.001	.36

3.5 Conceptual performance by genetics domain

Posttest advantages appeared across all four content domains. The largest absolute difference occurred in the linked concepts and representations domain, where learners had to distinguish gene, allele, DNA, and chromosome while placing them within cells. The mechanism and variation domain remained the most difficult in both groups, indicating that multicausal and gene-environment explanations needed continued instructional attention.

Table 7. GCUT scores by domain

Domain (maximum)	Comparison pre	GEN-LINK pre	Comparison post	GEN-LINK post	Post difference
Concepts and representations (8)	3.00	3.05	4.80	5.75	0.95
Meiosis and inheritance (8)	3.10	3.11	4.65	5.55	0.90
Punnett and probability reasoning (8)	3.20	3.19	4.90	5.65	0.75
Mechanism, variation, environment (6)	1.96	2.03	3.25	3.72	0.47
Total (30)	11.26	11.38	17.60	20.67	3.07

3.6 Scientific communication by rubric domain

GEN-LINK learners showed the strongest posttest advantage in scientific accuracy and evidence use. Reasoning remained lower than language and organization, but the group difference indicated that the structured talk and explanation frames helped learners state why evidence supported a genetic claim. The pattern supported the strategy's theory that communication quality depended on conceptual resources and explicit discourse support operating together.

Table 8. SCRG posttest scores by domain

Rubric domain (maximum 6)	Comparison M	GEN-LINK M	Mean difference
Scientific accuracy	3.65	4.45	0.80
Use of evidence	3.30	4.10	0.80
Causal/probabilistic reasoning	3.15	3.95	0.80
Disciplinary language and organization	3.78	4.15	0.37
Total	13.88	16.65	2.77

3.7 Misconception-sensitive indicators

Item-cluster inspection provided a more diagnostic view of learning. GEN-LINK learners more often rejected the idea that dominant traits were automatically stronger or more common, distinguished genes from DNA and chromosomes, explained segregation through meiosis, and treated Punnett-square results as probabilities rather than guaranteed family outcomes. The differences were consistent with the intervention's explicit contrast routines and repeated requirement to qualify claims.

Table 9. Percentage demonstrating mastery on misconception-sensitive item clusters

Target conception	Comparison (%)	GEN-LINK (%)	Difference (percentage points)
Gene, DNA, and chromosome were related but not interchangeable	47.6	73.8	26.2
Dominant did not mean stronger, better, or necessarily more common	50.0	76.2	26.2
Meiosis explained allele segregation and chromosome-number reduction	54.8	76.2	21.4
Punnett squares represented probabilities, not guaranteed outcomes	45.2	69.0	23.8
Traits reflected mechanisms and possible environmental influence	42.9	64.3	21.4

3.8 Implementation fidelity and acceptability

All sixteen planned sessions were delivered. Mean observed fidelity was 88.4% (range = 81.3%-94.4%). Strongest adherence occurred for paired objectives, representation linking, and structured discourse. The lowest component was scaffold fading, mainly because two lessons retained sentence frames longer than planned. Mean attendance was 94.7%, and 91.7% of learners completed all major explanation tasks. Overall acceptability was 4.36 on a five-point scale.

Table 10. Fidelity and learner acceptability

Implementation indicator	Estimate	Interpretation
Sessions delivered	16/16 (100%)	Complete planned dose
Overall fidelity	88.4%	High adherence
Paired content-language objectives	94.4%	Consistently visible and used

Implementation indicator	Estimate	Interpretation
Representation linking	91.3%	Strong implementation
Structured scientific discourse	89.4%	Strong implementation
Planned scaffold fading	81.3%	Implemented with minor delay
Learner attendance	94.7%	High exposure
Acceptability - usefulness	4.49/5	Very high
Acceptability - clarity	4.41/5	Very high
Acceptability - engagement	4.36/5	Very high
Acceptability - manageability	4.18/5	High
Acceptability - overall	4.36/5	Very high

4. Discussion

GEN-LINK was developed as a curriculum-aligned, language-responsive strategy and obtained strong expert-validation evidence, acceptable instrument reliability, high implementation fidelity, and favorable outcome estimates. After baseline and section adjustment, learners exposed to GEN-LINK scored about three points higher on a 30-point conceptual test and 2.7 points higher on a 24-point scientific-communication rubric than learners who received time-matched comparison instruction. The pattern supported the central proposition of the study: disciplinary language activities can function as mechanisms for genetics learning when they are aligned to conceptual representations and evidence practices.

4.1 Improvement in conceptual understanding

The conceptual advantage was consistent with research describing genetics as a problem of coordinating levels and models.[11,12] GEN-LINK did not teach gene, DNA, chromosome, meiosis, and inheritance as a sequence of isolated definitions. Learners repeatedly placed the terms in relation to one another, moved between diagrams and verbal explanations, and used the relations to interpret evidence. This design resembled instructional approaches that had enabled younger learners to reason across inheritance, meiotic, and molecular models.[13-15]

The misconception-sensitive results clarified why the overall difference mattered. Gains were not limited to procedural Punnett-square performance. GEN-LINK learners more often recognized that dominance does not mean strength or frequency and that a probability model does not predict a particular family with certainty. These distinctions required careful language. Contrast frames such as 'dominant refers to expression in a heterozygote, whereas common refers to frequency in a population' forced learners to encode relations that a single-word response could conceal.

The mechanism and gene-environment domain remained relatively difficult. This result was unsurprising because mechanistic accounts require learners to connect information and physical processes across levels, a documented source of genetics difficulty.[11,16] The finding suggested that one Grade 8 unit could establish a foundation but should not be expected to produce expert molecular reasoning. Future revisions could add more explicit protein-mediated examples while guarding against cognitive overload and curriculum misalignment.

4.2 Improvement in scientific communication

The scientific-communication effect was larger than the conceptual effect. GEN-LINK provided repeated practice in selecting evidence, connecting it to a claim, and expressing causal or probabilistic warrants. Such practice reflected the view that reading, writing, and interpreting text are fundamental components of scientific literacy.[6] It also aligned with evidence that collaborative argumentation and explicit explanation scaffolds can improve the quality of reasoning made visible in student responses.[8,9]

Communication gains should not be interpreted as a separate English effect detached from science. The tasks asked learners to explain genetic phenomena accurately. A fluent but scientifically incorrect paragraph could not receive a high score, and correct vocabulary without evidence or reasoning was insufficient. The rubric therefore represented disciplinary communication: accuracy, evidence, reasoning, and language organization were related but distinct requirements.

The smaller difference in language and organization relative to accuracy, evidence, and reasoning may indicate that both groups benefited from ordinary classroom exposure to scientific terms and written tasks. GEN-LINK's added value appeared strongest where learners had to use language to perform epistemic work. This interpretation was consistent with the proposition that science practices create opportunities for language development when learners engage in explaining, arguing, and interpreting rather than completing decontextualized language exercises.[7]

4.3 Contribution of the integrated design

Three design features likely worked together. First, paired objectives made language functions visible without reducing science time. Second, multimodal linking reduced representational fragmentation. Third, scaffolded talk and writing required learners to externalize and revise their thinking. Integrated science-literacy studies have likewise found benefits when texts, investigation, writing, and discussion serve a common conceptual goal.[18-21] GEN-LINK extended this logic to a genetics unit and to the specific linguistic challenges of probability, causal chains, and multilevel biological explanation.

The fading component was theoretically important. Permanent sentence frames can increase completion while creating dependence. By moving from full frames to partial prompts and independent explanations, GEN-LINK aimed to support transfer, consistent with evidence that faded scaffolds can prepare learners for unprompted scientific explanation.[9] Fidelity data showed that this was also the hardest component to deliver. Teachers may need a simple decision rule: retain support when scientific reasoning is absent, but fade it when the learner can independently preserve claim, evidence, and warrant.

4.4 Implications for Philippine junior high school

The findings offered a practical response to the combined science and reading challenge reflected in national performance data.[3] The strategy did not require a new subject, additional weekly period, or English-only remediation. Instead, it reorganized existing Grade 8 genetics lessons so that reading, speaking, and writing produced evidence of science learning. This feature could be attractive in schools where curriculum time and specialist support are limited.

At the school level, implementation would require joint planning between Science and English teachers, but not necessarily co-teaching every lesson. English specialists could help identify language functions, text complexity, and feedback routines, while Science teachers retained responsibility for conceptual accuracy. A short calibration process using learner work would help both departments distinguish grammar correction from scientific-communication feedback.

At the division or regional level, GEN-LINK could be adapted as a lesson-study package. Teachers could agree on one target misconception, co-design a representation and explanation task, observe implementation, analyze learner responses, and revise the material. This process would preserve local ownership and generate evidence before wider adoption. The intervention should not be scaled solely on the basis of one school or one unit.

4.5 Practical Recommendations

Pilot before adoption. Implement the eight-week package with formal school authorization and collect authentic baseline, outcome, fidelity, and learner-feedback data.

Protect conceptual accuracy. Require joint material review by a Science specialist and a language specialist, especially for dominance, probability, and gene-to-trait explanations.

Calibrate scoring. Use anchor responses and blinded double scoring for at least 20%-30% of scientific explanations.

Fade supports deliberately. Move from full frames to partial prompts and independent explanations using learner evidence, not a fixed calendar alone.

Replicate across contexts. Test the strategy with multiple schools, teachers, and delayed outcomes before making system-level effectiveness claims.

References

[1] Department of Education, Philippines. Self-Learning Module - Quarter 4 - Science: Grade 8, Modules 1 to 11. DepEd Learning Resources Management and Development System; 2022. https://lrmds.deped.gov.ph/detail/21910

[2] Batasang Pambansa. Batas Pambansa Blg. 188: An Act Creating a National High School in the Municipality of Quezon, Province of Nueva Vizcaya, to be Known as the Quezon National High School, and Appropriating Funds Therefor. Supreme Court E-Library; 1982. https://elibrary.judiciary.gov.ph/thebookshelf/showdocs/2/14564

[3] Organisation for Economic Co-operation and Development. PISA 2022 Results: Country Note - Philippines. OECD Publishing; 2023. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/philippines_a0882a2d-en.html

[4] National Research Council. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. National Academies Press; 2012. https://doi.org/10.17226/13165

[5] NGSS Lead States. Next Generation Science Standards: For States, By States. National Academies Press; 2013. https://doi.org/10.17226/18290

[6] Norris SP, Phillips LM. How literacy in its fundamental sense is central to scientific literacy. Science Education. 2003;87(2):224-240. https://doi.org/10.1002/sce.10066

[7] Lee O, Quinn H, Valdés G. Science and language for English language learners in relation to Next Generation Science Standards. Educational Researcher. 2013;42(4):223-233. https://doi.org/10.3102/0013189X13480524

[8] Osborne J. Arguing to learn in science: the role of collaborative, critical discourse. Science. 2010;328(5977):463-466. https://doi.org/10.1126/science.1183944

[9] McNeill KL, Lizotte DJ, Krajcik J, Marx RW. Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. Journal of the Learning Sciences. 2006;15(2):153-191. https://doi.org/10.1207/s15327809jls1502_1

[10] McNeill KL, Krajcik J. Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing. Pearson; 2012. https://eric.ed.gov/?id=ED534533

[11] Duncan RG, Reiser BJ. Reasoning across ontologically distinct levels: students' understandings of molecular genetics. Journal of Research in Science Teaching. 2007;44(7):938-959. https://doi.org/10.1002/tea.20186

[12] Duncan RG, Rogat AD, Yarden A. A learning progression for deepening students' understandings of modern genetics across the 5th-10th grades. Journal of Research in Science Teaching. 2009;46(6):655-674. https://doi.org/10.1002/tea.20312

[13] Duncan RG, Tseng KA. Designing project-based instruction to foster generative and mechanistic understandings in genetics. Science Education. 2011;95(1):21-56. https://doi.org/10.1002/sce.20407

[14] Duncan RG, Freidenreich HB, Chinn CA, Bausch A. Promoting middle school students' understandings of molecular genetics. Research in Science Education. 2011;41(2):147-167. https://doi.org/10.1007/s11165-009-9150-0

[15] Freidenreich HB, Duncan RG, Shea N. Exploring middle school students' understanding of three conceptual models in genetics. International Journal of Science Education. 2011;33(17):2323-2349. https://doi.org/10.1080/09500693.2010.536997

[16] Haskel-Ittah M, Duncan RG, Chinn CA, Yarden A. Students' conception of genetic phenomena and its effect on their ability to understand the underlying mechanism. CBE-Life Sciences Education. 2018;17(3):ar49. https://doi.org/10.1187/cbe.18-01-0014

[17] Todd A, Kenyon L. The learning-loss effect in genetics: what ideas do students retain or lose after instruction? CBE-Life Sciences Education. 2018;17(4):ar55. https://doi.org/10.1187/cbe.16-10-0310

[18] Cervetti GN, Barber J, Dorph R, Pearson PD, Goldschmidt PG. The impact of an integrated approach to science and literacy in elementary school classrooms. Journal of Research in Science Teaching. 2012;49(5):631-658. https://doi.org/10.1002/tea.21015

[19] Tong F, Irby BJ, Lara-Alecio R, Guerrero C, Fan Y, Huerta M. A randomized study of a literacy-integrated science intervention for low-socio-economic-status middle school students. International Journal of Science Education. 2014;36(12):2083-2109. https://doi.org/10.1080/09500693.2014.883107

[20] Tong F, Irby BJ, Lara-Alecio R, Koch J. Integrating literacy and science for English language learners: from learning-to-read to reading-to-learn. Journal of Educational Research. 2014;107(5):410-426. https://doi.org/10.1080/00220671.2013.833072

[21] Lara-Alecio R, Tong F, Irby BJ, Guerrero C, Huerta M, Fan Y. Assessing conceptual understanding via literacy-infused, inquiry-based science among middle-school English learners and economically challenged students. Education Sciences. 2018;8(1):27. https://doi.org/10.3390/educsci8010027

[22] Fang Z. The language demands of science reading in middle school. International Journal of Science Education. 2006;28(5):491-520. https://doi.org/10.1080/09500690500339092

[23] Reynolds JA, Thaiss C, Katkin W, Thompson RJ Jr. Writing-to-learn in undergraduate science education: a community-based, conceptually driven approach. CBE-Life Sciences Education. 2012;11(1):17-25. https://doi.org/10.1187/cbe.11-08-0064

[24] 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

[25] Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health. 2007;30(4):459-467. https://doi.org/10.1002/nur.20199

[26] Kuder GF, Richardson MW. The theory of the estimation of test reliability. Psychometrika. 1937;2(3):151-160. https://doi.org/10.1007/BF02288391

[27] Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. But.... Communication Methods and Measures. 2020;14(1):1-24. https://doi.org/10.1080/19312458.2020.1718629

[28] Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. BMJ. 2001;323(7321):1123-1124. https://doi.org/10.1136/bmj.323.7321.1123

[29] Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t tests and ANOVAs. Frontiers in Psychology. 2013;4:863. https://doi.org/10.3389/fpsyg.2013.00863

[30] Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychology. 2004;23(5):443-451. https://doi.org/10.1037/0278-6133.23.5.443

[31] What Works Clearinghouse. Procedures and Standards Handbook, Version 5.0. Institute of Education Sciences, U.S. Department of Education; 2022. https://ies.ed.gov/ncee/wwc/handbooks

[32] Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behavior Research Methods. 2009;41(4):1149-1160. https://doi.org/10.3758/BRM.41.4.1149

[33] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine. 2016;15(2):155-163. https://doi.org/10.1016/j.jcm.2016.02.012