

Teaching Scientific Writing Through Multimodal Evidence: Changes in Students' Use of Evidence and Written Reasoning in Environmental Literacy Contexts

**Stella Talitha¹, Yeti Mulyati¹, Vismaia S. Damaianti¹, Khaerudin Kurniawan¹,
Ainiyah Ekowati², Helena Mudi²**

¹Indonesian Language Study Program

FPBS, Universitas Pendidikan Indonesia, Bandung, Indonesia

²Indonesian Language and Literature Study Program

FKIP, Universitas Pakuan, Bogor, Indonesia

Email: stellatalitha@upi.edu

Abstract

This study examined how multimodal evidence scaffolding supported undergraduate students' scientific writing in an environmental literacy context. Using a quasi-experimental pretest-posttest design with a comparison group, the study involved 60 Biology Education undergraduates from two intact classes. The experimental group received explicit scaffolding for interpreting visual, spatial, numerical, and observational evidence, while the comparison group received regular scientific report writing instruction. Students' scientific reports were assessed using an analytic rubric and coded for evidence use and written reasoning. ANCOVA results, controlling for pretest performance, showed higher adjusted posttest performance in the experimental group, particularly in written reasoning, $F(1, 57) = 42.37$, $p < .001$, partial $\eta^2 = .43$; multimodal interpretation, $F(1, 57) = 31.28$, $p < .001$, partial $\eta^2 = .35$; and evidence integration, $F(1, 57) = 25.14$, $p < .001$, partial $\eta^2 = .31$. Student-level analyses also showed higher adjusted posttest proportions of higher-level evidence use, $F(1, 57) = 137.29$, $p < .001$, partial $\eta^2 = .71$, and advanced written reasoning, $F(1, 57) = 71.47$, $p < .001$, partial $\eta^2 = .56$. Qualitative findings indicated that students increasingly connected maps, photographs, tables, field notes, and numerical data to scientific claims and causal explanations. The findings suggest that multimodal evidence can function as an epistemic scaffold when instruction explicitly supports students in transforming information into evidence and evidence into written reasoning.

Keywords: *scientific writing; multimodal evidence; written reasoning; evidence use; writing instruction*

INTRODUCTION

Scientific writing is not merely a means of reporting information; it is an epistemic practice through which observations, data, disciplinary concepts, and prior knowledge are transformed into claims that can be examined by a scholarly community. In higher education, students are expected not only to reproduce established knowledge but also to participate in disciplinary ways of knowing, arguing, and explaining. Scientific texts gain their force from how writers select evidence, situate claims, and make claim-evidence relations visible (Bazerman, 1988). Toulmin's model clarifies this relation by distinguishing claim, data, and the warrant that authorises the movement from data to claim (Toulmin, 2003). Written reasoning is therefore the explicit work through which writers explain why evidence supports a conclusion within a field.

This epistemic demand becomes especially visible when students write from multiple forms of evidence. Scientific explanations are rarely constructed from verbal information alone. They may draw on numerical measurements, tables, graphs, maps, photographs, diagrams, field notes, and other disciplinary representations. Each mode offers a different evidential affordance: maps foreground spatial relations, graphs reveal patterns of change, tables support comparison, photographs document visible conditions, and field notes preserve contextual observation. Yet these resources do not function as evidence simply because they appear in a report. A table remains a display until the writer identifies the relevant pattern and explains its significance; a map does not support a claim unless its spatial relation is interpreted; and a photograph does not contribute to an explanation unless the writer shows what it reveals and why it matters.

The distinction between information and evidence is therefore central. Evidence use is a constructive process in which writers select, organise, transform, and connect information for a particular claim, task, and audience (Spivey, 1997). It is also disciplinary because the adequacy of evidence depends on the concepts, methods, and forms of reasoning valued in a field. Students may collect substantial information and still produce weak scientific writing if they do not identify which details are relevant, interpret them through appropriate disciplinary knowledge, or organise them into a defensible explanation. Studies of college writing have long shown that students struggle to gather specific information, state a position, manage complexity, and use disciplinary methods to support that position (Walvoord & McCarthy, 1990). Thus,

stronger writing is not characterised by the quantity of data included, but by the writer's capacity to make data function as evidence within a coherent line of reasoning.

This problem persists despite the growing emphasis on academic writing in higher education. Writing programmes commonly address text structure, citation, research ethics, reference management, and preparation for publication (Anggaira & Wulantina, 2021; Faizah et al., 2023; Lida et al., 2024; Nurhasanah & Maspuroh, 2021; Rizqi & Purwanti, 2023; Tafonao & Widjaja, 2021; Wahyudin et al., 2023). These components are important, but they do not by themselves teach how evidence justifies a claim. Students may learn report conventions and still describe a graph without explaining its meaning, mention values without identifying their significance, or leave readers to infer relationships among representations. The central issue is whether they can transform observations and representations into evidence-based explanation.

Multimodality provides a useful basis for understanding this challenge. Meaning is produced through the selection and orchestration of modes rather than through language alone (Böck & Pachler, 2013). However, research on multimodal discourse often emphasises meaning-making, persuasion, identity, or ideological positioning, whereas scientific writing requires an additional epistemic operation: representations must be evaluated, interpreted, and connected to knowledge claims. Visual or spatial resources may influence how a text is perceived, but they do not establish epistemic justification unless the relation among representation, evidence, and claim is made explicit (Syah, 2024; Wuryaningrum et al., 2023). For scientific writing instruction, the question is therefore not simply how students use several modes, but how they reason with those modes as evidence.

Vygotsky's account of mediation helps explain why students need explicit support to perform this work. Signs, tools, social interaction, and culturally situated activity can reorganise attention and higher psychological functions (Vygotsky, 1978). In a scientific writing task, maps, graphs, photographs, tables, field notes, guiding questions, model paragraphs, and evidence-mapping worksheets can mediate what students notice and how they organise an explanation. Epistemic scaffolding gives this mediation a more specific function by helping learners evaluate evidence, relate data to disciplinary knowledge, and judge whether a claim is warranted (Boon et al., 2022; Liu et al., 2023). Multimodal evidence becomes an epistemic scaffold when instruction guides students

to identify salient patterns, select relevant information, connect evidence to claims, and articulate the reasoning that makes the connection defensible.

Such scaffolding also needs to be situated within an explicit process of writing and revision. Less experienced writers often rely on knowledge telling, whereas more developed writing requires knowledge transforming: writers set goals, examine problems, reorganise ideas, and refine understanding through composing (Bereiter & Scardamalia, 1987). Modelling, guided practice, prompts, discussion, feedback, and revision can make these tacit operations visible (Bereiter & Scardamalia, 1987; Graham et al., 2013). Revision allows students to reconsider claim precision, evidence relevance and specificity, interpretive accuracy, and the visibility of reasoning. Evidence-based revision is therefore epistemic rather than merely corrective.

Environmental literacy provides a productive context for examining these processes because environmental phenomena are complex, situated, and commonly represented through multiple modes. In watershed-based learning, for example, students may encounter photographs of river conditions, maps of land use, water-quality tables, graphs, field observations, and numerical measurements. These resources require students to compare locations, identify trends, interpret visible conditions, and connect human activity with environmental change. Environmental literacy tasks therefore make visible a broader writing problem: students must decide what counts as evidence, how evidence should be verbalised and ordered, and how several forms of evidence can be coordinated within a scientific explanation. The environmental context in this study is consequently treated as a disciplinary setting for investigating scientific writing rather than as a separate object of inquiry.

Recent writing research has provided substantial knowledge about composing behaviour, pre-task planning, metacognitive regulation, explicit text-structure instruction, and the design and implementation of writing interventions. Studies show variation in how writers pause, review, revise, and compose (Dux Speltz et al., 2026), the influence of planning on textual performance (Yoon et al., 2026), the role of metacognitive regulation in complex writing (Baillargeon, 2025), and the value of explicit text-structure instruction (Tesfa et al., 2025). Intervention research also stresses robust designs, implementation fidelity, and careful interpretation of outcomes (Wyse et al., 2025).

Nevertheless, the state of the art remains fragmented in three respects. First, research on writing processes and instructional interventions has explained how students plan, draft, and revise, but has less often examined the evidential work required to transform multimodal disciplinary representations into written reasoning. Second, multimodality research has demonstrated that modes shape meaning, but has not consistently distinguished the presence of multimodal resources from their disciplined use as evidence for scientific claims. Third, scientific literacy research has emphasised data interpretation and evidence evaluation (Thaariq et al., 2023; Vashti et al., 2021), yet the textual process through which students convert interpreted data into warrants and coherent scientific explanations remains insufficiently specified. What has been established is that explicit scaffolding, multimodal representation, evidence evaluation, and revision each matter; what remains to be investigated is how these elements can be integrated within one writing intervention and how that integration changes both writing performance and the evidential structure of students' texts.

The originality of the present research lies in bringing these strands together. It conceptualises multimodal evidence as an epistemic scaffold for scientific writing and operationalises that concept through a staged intervention that moves recursively from multimodal representation to guided interpretation, evidence selection, claim-evidence connection, written reasoning, scientific explanation, and evidence-based revision. The study goes beyond overall text quality by examining student-level evidence use and reasoning and by tracing movement from evidence mention to interpretation, isolated to converging evidence, and description to causal explanation. This perspective is significant across the curriculum because students increasingly write from visual, numerical, spatial, and observational sources.

Accordingly, the objective of this study is to determine how multimodal evidence scaffolding supports undergraduate students' scientific writing performance, higher-level evidence use, and advanced written reasoning in an environmental literacy context, and to identify the textual changes through which those outcomes are realised. To meet this objective, the study combines a quasi-experimental pretest-posttest design with a comparison group, ANCOVA that controls for initial performance, student-level analysis of evidence-use and reasoning proportions, and qualitative comparison of students' pretest and posttest texts. The study addresses the following research questions:

- 1) After controlling for pretest performance, how does posttest scientific writing performance differ between students receiving multimodal evidence scaffolding and those receiving regular scientific writing instruction?
- 2) After controlling for pretest proportions, how do the groups differ in higher-level evidence use and advanced written reasoning?
- 3) What textual changes characterise students' integration of visual, numerical, spatial, and observational evidence into scientific explanations?

METHOD

Research design

This study used a quasi-experimental pretest-posttest design with a comparison group, complemented by qualitative analysis of students' scientific texts, worksheets, and reflective notes. The design was selected because the study was conducted in two intact undergraduate classes where individual random assignment was not feasible. One class was assigned as the experimental group and received multimodal evidence scaffolding, while the other class served as the comparison group and received regular scientific writing instruction.

The quantitative component measured differences in scientific writing performance after instruction, while the qualitative component traced how students used evidence and constructed written reasoning. This combination was necessary because the study was concerned not only with whether writing scores differed, but also with how students transformed multimodal evidence into scientific explanation. Quasi-experimental designs are appropriate for educational intervention studies involving naturally occurring groups, provided that initial group differences and limits of causal inference are addressed (Shadish et al., 2002).

Participants and context

The participants were 60 undergraduate students enrolled in a Biology Education programme and drawn from two intact classes in an environmental literacy-oriented course. The experimental group consisted of 30 students, and the comparison group consisted of 30 students. All participants had prior experience with general academic writing, but their pretest texts indicated limited experience in using visual, numerical, spatial, and observational data as evidence in scientific writing.

The environmental literacy context was selected because it provided authentic materials for evidence-based scientific writing. The writing task was situated in

watershed-related issues and required students to examine environmental phenomena through field observations, maps, photographs, tables, graphs, and numerical data. The study followed the institution's procedures for classroom-based educational research. Participation was voluntary, students provided informed consent, and all written texts, worksheets, and reflective notes were anonymised before analysis.

Writing task

Students wrote a scientific report explaining a watershed-based environmental issue. The task required them to identify an environmental problem, formulate a scientific claim, select relevant evidence, interpret the evidence, explain how the evidence supported the claim, and organise the explanation into a coherent report. Each report was expected to use at least three types of evidence: observational evidence, visual or spatial evidence, and numerical evidence.

The pretest was administered before the intervention and the posttest after the intervention. The two tasks used comparable writing demands and the same assessment criteria, although the environmental dataset was adjusted to avoid direct repetition. The task was designed to represent the target construct of evidence-based scientific writing because writing assessment requires alignment among task, construct, scoring criteria, and score interpretation (Weigle, 2002).

Instructional intervention

The intervention consisted of seven stages designed to help students transform multimodal environmental data into written scientific reasoning. It was implemented across 12 instructional sessions and focused explicitly on evidence interpretation, claim formulation, written reasoning, drafting, feedback, and revision.

Table 1

Stages of multimodal evidence scaffolding

Stage	Instructional focus	Student activity	Construct targeted
1. Orienting to environmental evidence	Recognising problems and evidence sources	Examining maps, photographs, graphs, tables, and field notes	Topic awareness

Stage	Instructional focus	Student activity	Construct targeted
2. Interpreting multimodal evidence	Reading visual, spatial, numerical, and observational data	Identifying patterns, trends, anomalies, and relations	Evidence interpretation
3. Formulating claims	Turning interpreted data into claims	Writing specific scientific claims	Claim quality
4. Linking evidence to claims	Selecting relevant evidence	Matching claims with supporting evidence	Evidence relevance
5. Constructing written reasoning	Explaining the evidential relation	Writing claim-evidence-reasoning segments	Reasoning quality
6. Drafting scientific explanation	Organising the report	Writing a scientific report draft	Coherence
7. Evidence-based revision	Revising claim, evidence, and reasoning	Improving the report using feedback and the rubric	Revision quality

The intervention followed principles of explicit writing instruction in which students were supported through modelling, guided practice, feedback, and revision. This orientation treats planning, composing, evaluation, revision, and assessment as connected components of writing development (Graham et al., 2013).

Instruments

Four instruments were used. The first was an analytic scientific writing rubric used to score pretest and posttest reports. The second was an evidence-use coding scheme. The third was a written-reasoning coding scheme. The fourth source consisted of student worksheets and reflective notes used to examine students' interpretation of multimodal evidence during the intervention.

Table 2

Scientific writing rubric dimensions

Dimension	Score focus
Claim clarity	The claim is clear, specific, and defensible.
Evidence relevance	The selected evidence supports the claim.
Evidence specificity	The evidence is concrete, precise, and data-based.
Multimodal interpretation	Visual, numerical, spatial, and observational evidence is interpreted accurately.
Evidence integration	Evidence is embedded coherently into the paragraph or report.
Written reasoning	The relation between evidence and claim is explained logically.
Scientific coherence	The report is organised coherently and uses appropriate scientific terminology.

Table 3

Evidence-use coding scheme

Code	Description
E0 No evidence	A claim is made without evidence.
E1 Mentioned evidence	Evidence is mentioned but not explained.
E2 Descriptive evidence	Evidence is described but not interpreted.
E3 Interpretive evidence	Evidence is interpreted to show meaning or a pattern.
E4 Integrated evidence	Evidence is used to support a specific claim.
E5 Reasoned multimodal evidence	Several forms of evidence are connected to build an explanation.

Table 4

Written-reasoning coding scheme

Code	Description
R0 Unsupported assertion	The text states a claim without a reason.

Code	Description
R1 Simple explanation	The text gives a general explanation without evidence.
R2 Evidence-linked reasoning	The text begins to connect evidence to a claim.
R3 Causal reasoning	The text explains a cause-and-effect relation based on evidence.
R4 Integrative reasoning	The text connects several pieces of evidence.
R5 Scientific reasoning	The text builds a coherent explanation using multimodal evidence and scientific concepts.

Coding of evidence use and written reasoning

Students' pretest and posttest texts were analysed using qualitative content analysis. The unit of analysis was the sentence or paragraph segment in which a student stated a claim, referred to evidence, or explained the relationship between evidence and claim. Content analysis was appropriate because the study sought to make systematic inferences from students' texts about how evidence was used and how reasoning was constructed (Krippendorff, 2019).

Coding was conducted in two cycles. The first cycle identified evidence use and written reasoning in each text. The second cycle grouped coded segments into broader patterns of change across pretest and posttest texts. This procedure followed the logic of qualitative coding as an iterative movement from codes to categories and from categories to interpretive patterns (Saldaña, 2016).

Validity, reliability, and fidelity

The writing rubric and coding schemes were reviewed by experts in writing instruction, scientific writing, and environmental literacy before the main analysis. The review focused on construct representation, clarity of descriptors, applicability to student texts, and alignment with the intervention. Revisions were made before the instruments were applied to the full dataset.

Two independent raters scored all pretest and posttest reports after a training session using sample texts outside the main dataset. Inter-rater reliability for the analytic rubric was calculated using a two-way random-effects intraclass correlation coefficient for absolute agreement. Average-measure ICC values ranged from .87 to .91 across the seven dimensions, and the ICC for the total writing score was .98.

Discrepancies larger than one scale point were discussed after independent scoring and resolved through adjudication.

For qualitative coding, at least 25% of the texts were double-coded independently. Agreement was acceptable for evidence use, Cohen's $\kappa = .74$ and Krippendorff's $\alpha = .74$, and for written reasoning, Cohen's $\kappa = .69$ and Krippendorff's $\alpha = .69$. Coding definitions were refined when disagreement indicated ambiguity in category boundaries.

Implementation fidelity was monitored through lesson plans, instructional checklists, student worksheets, and observation notes. Mean fidelity was .90, with observed values ranging from .88 to 1.00, indicating that the intervention was delivered consistently enough for subsequent analysis.

Table 5

Reliability and implementation fidelity

Measure	Reliability index	Value
Claim clarity	ICC(A,2)	.87
Evidence relevance	ICC(A,2)	.91
Evidence specificity	ICC(A,2)	.89
Multimodal interpretation	ICC(A,2)	.89
Evidence integration	ICC(A,2)	.91
Written reasoning	ICC(A,2)	.90
Scientific coherence	ICC(A,2)	.90
Total writing score	ICC(A,2)	.98
Evidence-use coding	κ / Krippendorff's α	.74 / .74
Written-reasoning coding	κ / Krippendorff's α	.69 / .69
Implementation fidelity	Mean fidelity rate	.90

Data analysis

Quantitative data were analysed using descriptive and inferential statistics. Means, standard deviations, and confidence intervals were calculated for each rubric dimension in the pretest and posttest for both groups. Baseline equivalence was examined using pretest scores. Before inferential testing, normality, homogeneity of variance, and

homogeneity of regression slopes were checked. Because Shapiro-Wilk tests indicated non-normal residuals for several dimensions, stratified bootstrap confidence intervals with 3,000 resamples were computed as sensitivity checks for ANCOVA adjusted mean differences.

Between-group posttest differences were examined using ANCOVA, with posttest scores as dependent variables, instructional group as the fixed factor, and the corresponding pretest scores as covariates. Adjusted posttest means, adjusted mean differences, 95% confidence intervals, F values, p values, Holm-adjusted p values, and partial eta squared were reported. Where assumption checks suggested potential violations, the interpretation was compared with bootstrap confidence intervals.

Qualitative analysis focused on changes in evidence use and written reasoning. Frequencies were calculated for each evidence-use and reasoning code and compared across pretest, posttest, and group. To avoid treating coded segments as independent observations, student-level proportions were calculated. Higher-level evidence use was defined as E3-E5, while advanced written reasoning was defined as R3-R5. ANCOVA was conducted on posttest proportions, with pretest proportions as covariates.

Anonymised excerpts were selected to illustrate movement from visual mention to interpretation, from single evidence to converging evidence, from description to causal explanation, and from data insertion to argument construction. Quantitative and qualitative findings were integrated during interpretation to explain not only whether students' scientific writing differed, but also how their evidence use and written reasoning changed.

RESULTS AND DISCUSSION

Changes in overall scientific writing performance

Baseline comparisons indicated no statistically significant pretest differences between the experimental and comparison groups across the seven writing dimensions, p s = .125 to .639. Homogeneity of variance and homogeneity of regression slopes were met for all ANCOVA models. Shapiro-Wilk tests indicated non-normal residuals for evidence relevance, multimodal interpretation, evidence integration, and written reasoning; therefore, stratified bootstrap confidence intervals were computed as sensitivity checks. The bootstrap intervals were consistent with the parametric ANCOVA results.

Figure 1 presents the unadjusted pretest and posttest means. The experimental group improved across all seven dimensions, with especially visible gains in multimodal interpretation, evidence integration, and written reasoning. The comparison group also improved, but its posttest scores remained lower on the dimensions most directly related to evidence transformation.

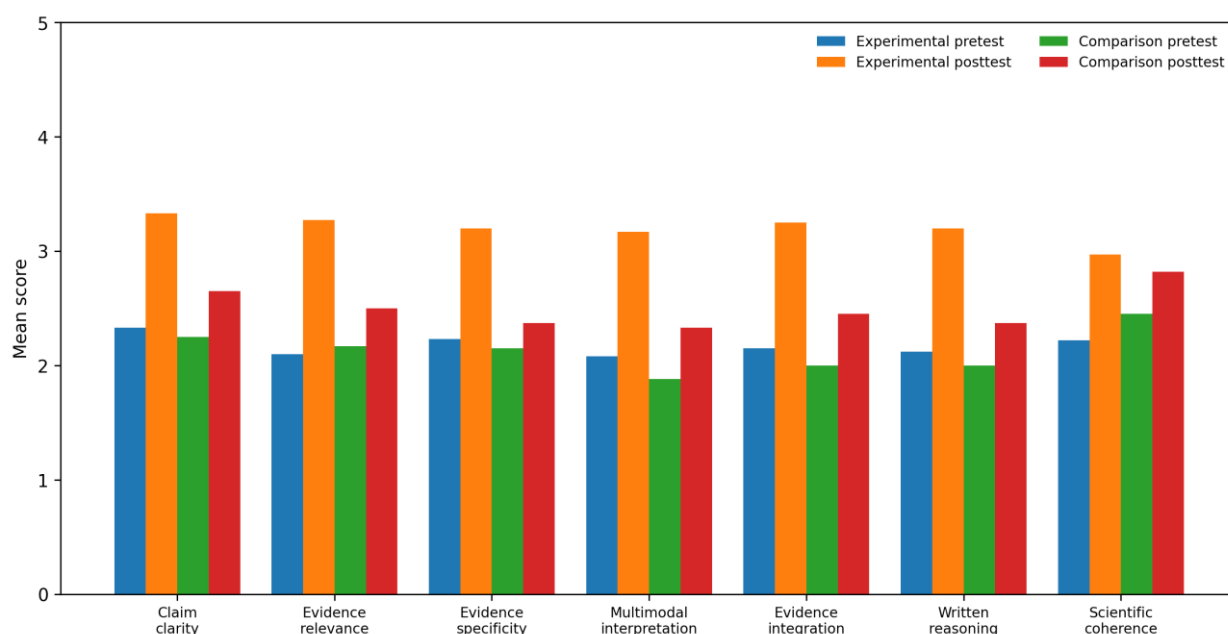


Figure 1. Pretest and posttest mean scores for scientific writing dimensions in the experimental and comparison groups. Means are unadjusted.

Table 6

Descriptive statistics for scientific writing performance

Dimension	Experimental pretest M (SD)	Experimental posttest M (SD)	Comparison pretest M (SD)	Comparison posttest M (SD)
Claim clarity	2.33 (.66)	3.33 (.72)	2.25 (.54)	2.65 (.53)
Evidence relevance	2.10 (.58)	3.27 (.50)	2.17 (.51)	2.50 (.59)
Evidence specificity	2.23 (.58)	3.20 (.68)	2.15 (.53)	2.37 (.59)
Multimodal interpretation	2.08 (.49)	3.17 (.50)	1.88 (.50)	2.33 (.59)

Dimension	Experimental pretest M (SD)	Experimental posttest M (SD)	Comparison pretest M (SD)	Comparison posttest M (SD)
Evidence integration	2.15 (.51)	3.25 (.60)	2.00 (.53)	2.45 (.63)
Written reasoning	2.12 (.63)	3.20 (.57)	2.00 (.45)	2.37 (.56)
Scientific coherence	2.22 (.61)	2.97 (.56)	2.45 (.69)	2.82 (.56)

Table 7

ANCOVA results for scientific writing dimensions

Dimension	Adjusted experimental posttest M	Adjusted comparison posttest M	Adjusted mean difference [95% CI]	F(1, 57)	p	Holm p	Partial η^2
Claim clarity	3.31	2.67	0.65 [0.35, 0.94]	18.74	< .001	< .001	.25
Evidence relevance	3.28	2.48	0.80 [0.56, 1.05]	42.85	< .001	< .001	.43
Evidence specificity	3.17	2.40	0.77 [0.53, 1.00]	42.92	< .001	< .001	.43
Multimodal interpretation	3.12	2.38	0.75 [0.48, 1.01]	31.28	< .001	< .001	.35
Evidence integration	3.21	2.49	0.71 [0.43, 1.00]	25.14	< .001	< .001	.31
Written reasoning	3.16	2.40	0.76 [0.53, 0.99]	42.37	< .001	< .001	.43
Scientific coherence	3.02	2.76	0.26 [0.01, 0.51]	4.47	.039	.039	.07

After controlling for pretest performance, adjusted posttest differences favoured the experimental group on all seven dimensions. The largest effects were found for evidence relevance, evidence specificity, and written reasoning, each with partial $\eta^2 = .43$. Multimodal interpretation and evidence integration also showed substantial

adjusted group differences, partial $\eta^2 = .35$ and $.31$, respectively. Scientific coherence showed a smaller effect, partial $\eta^2 = .07$. This pattern indicates that the intervention was associated most strongly with dimensions requiring students to interpret, specify, integrate, and reason with evidence.

Changes in students' use of evidence

The coding of evidence use showed a shift in the experimental group from surface-level evidence use toward more interpretive and integrated use. At pretest, students frequently made claims without evidence or mentioned evidence without explaining its relevance. At posttest, these lower-level categories decreased, while interpretive, integrated, and reasoned multimodal evidence increased.

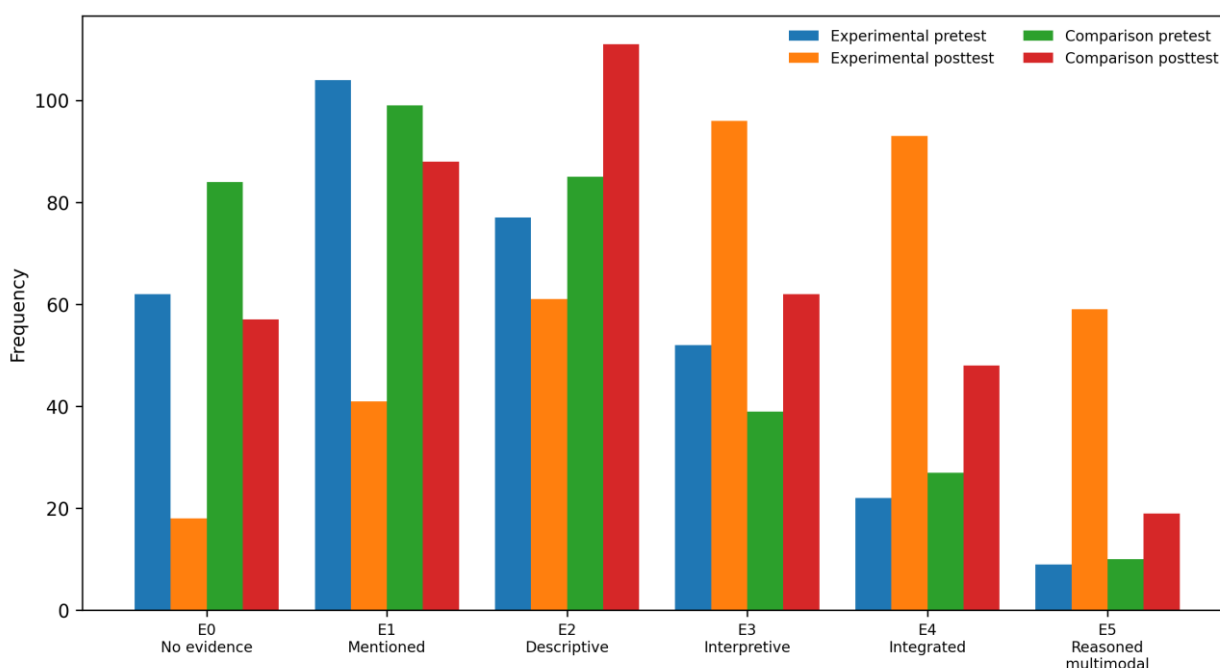


Figure 2. Distribution of evidence-use categories in pretest and posttest writing across groups.

Table 8

Changes in evidence-use frequency

Evidence-use category	Experimental pretest	Experimental posttest	Change	Comparison pretest	Comparison posttest	Change
E0 No evidence	62	18	-44	84	57	-27

Evidence-use category	Experimental pretest	Experimental posttest	Change	Comparison pretest	Comparison posttest	Change
E1 Mentioned evidence	104	41	-63	99	88	-11
E2 Descriptive evidence	77	61	-16	85	111	26
E3 Interpretive evidence	52	96	44	39	62	23
E4 Integrated evidence	22	93	71	27	48	21
E5 Reasoned multimodal evidence	9	59	50	10	19	9

Higher-level evidence use, consisting of interpretive, integrated, and reasoned multimodal evidence, increased in the experimental group from 83 of 326 coded evidence segments at pretest, or 25.5%, to 248 of 368 segments at posttest, or 67.4%. In the comparison group, the same categories increased from 76 of 344 segments, or 22.1%, to 129 of 385 segments, or 33.5%.

To avoid treating coded segments as independent observations, higher-level evidence use was also calculated as a student-level proportion. ANCOVA on posttest proportions, controlling for pretest proportions, showed a significant adjusted group difference, $F(1, 57) = 137.29$, $p < .001$, partial $\eta^2 = .71$. The experimental group increased from $M = .23$, $SD = .13$ at pretest to $M = .66$, $SD = .08$ at posttest, whereas the comparison group increased from $M = .21$, $SD = .15$ to $M = .31$, $SD = .14$. The result indicates that the change was not merely a consequence of producing more coded segments; it reflected a different distribution of evidence use within students' own texts.

Changes in written reasoning

The analysis of written reasoning showed a similar pattern. At pretest, students in both groups frequently relied on unsupported assertions or general explanations. After the intervention, students in the experimental group produced more causal, integrative, and scientific reasoning, while the comparison group remained concentrated in lower and middle reasoning categories.

Table 9

Changes in written-reasoning frequency

Reasoning level	Experimental pretest	Experimental posttest	Change	Comparison pretest	Comparison posttest	Change
R0 Unsupported assertion	59	15	-44	64	58	-6
R1 Simple explanation	84	34	-50	92	94	2
R2 Evidence-linked reasoning	57	61	4	55	78	23
R3 Causal reasoning	37	76	39	33	44	11
R4 Integrative reasoning	21	58	37	16	29	13
R5 Scientific reasoning	3	31	28	2	8	6

Advanced reasoning, consisting of causal, integrative, and scientific reasoning, increased in the experimental group from 61 of 261 coded reasoning segments, or 23.4%, to 165 of 275 segments, or 60.0%. In the comparison group, advanced reasoning increased from 51 of 262 segments, or 19.5%, to 81 of 311 segments, or 26.0%. The most visible change concerned the warranting function of writing: students increasingly explained how field observations, spatial patterns, and numerical data supported an environmental claim.

A parallel student-level analysis was conducted for advanced written reasoning. ANCOVA on posttest proportions, controlling for pretest proportions, showed a significant adjusted group difference, $F(1, 57) = 71.47$, $p < .001$, partial $\eta^2 = .56$. The experimental group increased from $M = .21$, $SD = .15$ to $M = .55$, $SD = .17$, while the comparison group increased from $M = .19$, $SD = .10$ to $M = .22$, $SD = .13$.

Patterns of multimodal evidence integration

The qualitative analysis identified four recurring patterns in students' posttest writing. These patterns show how students' use of multimodal evidence changed after the intervention.

Pattern 1: From visual mention to visual interpretation.

At pretest, several students referred to visuals without interpreting them. One student wrote, "The map shows the river area" (E03, pretest). At posttest, the same student wrote, "The map indicates that settlement density is highest near the riverbank,

which helps explain the accumulation of domestic waste in the downstream area" (E03, posttest). The posttest version interpreted the map as spatial evidence and connected it to a causal explanation.

Pattern 2: From single evidence to converging evidence.

Students also began to combine different types of evidence within one explanation. At pretest, one student wrote, "The water quality table shows pollution" (E11, pretest). At posttest, the student wrote, "The water quality table, field notes, and riverbank photograph together indicate that the most degraded section is located near dense housing and informal waste disposal points" (E11, posttest). Numerical, observational, and visual evidence converged to support a more specific claim.

Pattern 3: From description to causal explanation.

One student's pretest statement, "The river is polluted because it is dirty" (E20, pretest), relied on circular reasoning. At posttest, the same student wrote, "The increase in visible waste near dense settlements, combined with lower water clarity in the observation data, suggests that domestic activity contributes to declining river quality" (E20, posttest). The revised explanation linked visual evidence, settlement context, and observation data to build a plausible causal account.

Pattern 4: From data insertion to argument construction.

Several posttest texts used data as part of argument construction rather than as isolated information. Student E25 initially wrote, "Table 2 is about water quality." At posttest, the student wrote, "Table 2 shows that turbidity is highest at Station 3; because this station is located below the settlement area on the map, the data support the claim that land use near the river affects water condition." Numerical evidence was interpreted, spatial evidence was connected, and both were used to justify a claim. Together, these patterns suggest that multimodal evidence scaffolding helped students move from displaying evidence toward reasoning with evidence. The strongest posttest excerpts did not simply include more modes; they coordinated modes to support claims.

Remaining difficulties

Despite the overall improvement, several difficulties remained in posttest writing. These difficulties were more frequent in the comparison group, but they were not absent from the experimental group. A student was counted in a category when the difficulty occurred at least once in the posttest text.

Table 10

Remaining difficulties after posttest

Difficulty	Experimental count (%)	Comparison count (%)	Total count
Claim too broad	14 (46.7%)	28 (93.3%)	42
Evidence not specific	17 (56.7%)	29 (96.7%)	46
Visual not interpreted	13 (43.3%)	30 (100%)	43
Reasoning too short	13 (43.3%)	30 (100%)	43
Visual and numerical evidence disconnected	20 (66.7%)	30 (100%)	50
Causal link weak	6 (20.0%)	24 (80.0%)	30

The most persistent difficulty in the experimental group was coordinating visual and numerical evidence. Twenty students still showed at least one instance in which several modes appeared in the text but were not fully connected. Evidence specificity also remained challenging: some students referred to high data or poor water conditions without identifying the measured variable, threshold, or significance for the claim.

Some students continued to treat multimodal evidence as a list. For example, one student wrote, "The picture, graph, and map show that pollution is serious" (E18, posttest). Although several modes were mentioned, their relationship was not explained. The presence of multiple modes therefore did not necessarily indicate multimodal reasoning. These remaining difficulties suggest that cross-modal coordination is a demanding writing operation and may require longer, more sustained support.

Discussion

Multimodal evidence as an epistemic scaffold for scientific writing

After pretest performance was controlled, the experimental group achieved higher adjusted posttest scores on all seven writing dimensions. The intervention was especially strongly associated with evidence relevance, evidence specificity, and written

reasoning, each with partial $\eta^2 = .43$, as well as multimodal interpretation and evidence integration. Student-level analyses showed even larger adjusted group differences in higher-level evidence use, partial $\eta^2 = .71$, and advanced written reasoning, partial $\eta^2 = .56$. Taken together, the findings indicate that the central challenge in students' scientific writing was not the absence of information, but the difficulty of transforming available information into evidence.

Students in the experimental group did not merely add more maps, tables, photographs, or observation notes to their reports. Multimodal resources increasingly functioned as grounds for claims and as resources for explanation. This distinction matters because the presence of data in a text does not guarantee evidential use. Data become evidence only when writers select, interpret, and connect them to a claim in ways that are accountable to readers and disciplinary expectations.

This interpretation extends the view of scientific writing as an epistemic and rhetorical practice. Scientific writing is not a transparent record of facts, but a form of written knowledge in which observations, prior literature, representations, and claims are organised for assessment by a disciplinary community (Bazerman, 1988). The present findings add to this view by showing that multimodal resources can participate in epistemic work when students are taught to use them as evidence. Maps, photographs, tables, graphs, and field notes shape what students notice, what they treat as relevant, and how they justify environmental claims.

The results also support Toulmin's view that claims require more than data; they require warrants that make the relation between data and claim explicit (Toulmin, 2003). The strong effects for written reasoning, multimodal interpretation, and evidence integration suggest that the scaffold supported warranting work. Students increasingly explained why a spatial pattern, numerical difference, or field observation mattered for the claim being advanced. Multimodal evidence therefore functioned as an epistemic scaffold not because it made reports visually richer, but because it mediated the movement from perception to explanation.

From descriptive reporting to written reasoning

The most meaningful textual change was the movement from descriptive reporting to written reasoning. At pretest, many students stated what could be seen or generally known: the river was dirty, the map showed an area, the table contained data, or the photograph showed waste. These statements were not necessarily incorrect, but they

remained weak as scientific writing because they did not articulate the meaning of the evidence. At posttest, students in the experimental group more often explained what a pattern indicated, why a difference mattered, and how several pieces of evidence supported an environmental claim.

The student-level proportion analyses strengthen this interpretation because the pattern was not limited to a larger number of coded segments; it appeared in the relative distribution of evidence use and reasoning within students' own texts. Improvement in scientific writing therefore cannot be understood only as better structure, clearer language, or more complete content. The central developmental movement was epistemic: students learned to turn observation into explanation.

This movement aligns with the view that writing from sources and evidence involves selection, transformation, and connection rather than the simple transfer of information into a new text (Spivey, 1997). Reasoning emerged through intermediate textual moves: students first mentioned evidence, then described it, interpreted it, connected it with other evidence, and finally used it to construct an argument. This progression offers a useful analytic lens for examining how novice writers gradually learn to make evidence do explanatory work.

The findings also speak to a broader challenge in higher education. Across disciplines, students increasingly write from visual, numerical, spatial, digital, and textual resources, yet they are not always prepared to use these resources as evidence. The present study shows how explicit scaffolding can help students move from what is shown to what can be argued from what is shown.

Evidence integration as a writing problem

The remaining difficulties are especially revealing. Even after the intervention, some students struggled to connect visual and numerical evidence, specify the meaning of data, or sustain causal explanation. These difficulties should not be read simply as a failure to understand environmental content. They point to evidence integration as a writing problem. Students may understand a map or a table in isolation, yet still find it difficult to convert that understanding into coherent written explanation.

Data literacy and writing competence are often treated separately, but scientific writing requires their integration. Evidence must be interpreted, verbalised, sequenced, and positioned within discourse. A map has to be translated into a spatial claim, a table

into a comparison, a photograph into a documented condition, and a field note into a contextual explanation. These are compositional as well as analytical acts.

The findings therefore extend Walvoord and McCarthy's argument that student writing is shaped by disciplinary expectations (Walvoord & McCarthy, 1990). In environmental scientific writing, evidence is not limited to verbal citation; it may include spatial distribution, visual documentation, numerical measurement, and field observation. The persistence of integration problems also confirms the value of analytic assessment. A holistic score could obscure the fact that students may improve in claim clarity while still struggling with evidence integration or may include several modes without connecting them.

Planning, scaffolding, and revision as mechanisms of development

The intervention appears to have supported writing development because it reorganised what students did before, during, and after drafting. Students were guided to identify evidence, mark patterns, compare data, formulate claims, link evidence to claims, write reasoning, and revise their explanations. Planning therefore functioned as epistemic preparation rather than simply time allocated before writing.

This interpretation resonates with research showing that the value of planning depends on what writers do during planning, including how they organise content and explore relationships (Yoon et al., 2026). In the present study, students planned the epistemic logic of the report. The intervention also encouraged movement from knowledge telling toward knowledge transforming by requiring students to reconsider evidence, refine claims, and revise explanations through composing (Bereiter & Scardamalia, 1987).

Revision was central to this process. Evidence-based revision shifted attention away from surface correction toward the relation between claim and evidence. Students had to ask whether a claim was precise, whether evidence was relevant and specific, whether interpretation was adequate, and whether reasoning was visible. This extends research-based writing instruction by showing how revision can be directed specifically toward multimodal evidence and written reasoning (Graham et al., 2013).

Implications for writing instruction across the curriculum

Many disciplinary courses ask students to use data, support claims, or write scientific reports, but the steps required to transform data into evidence are often left implicit. The present study suggests that students need explicit support in reading evidence, selecting relevant details, interpreting representations, connecting evidence

to claims, and revising written explanations. Without such support, multimodal resources may remain present in learning activities but absent from the reasoning structure of student writing.

This implication is particularly important in data-rich learning environments, where scientific, environmental, technological, and social issues are represented through dashboards, maps, graphs, photographs, simulations, and numerical datasets. Writing instruction should therefore move beyond paragraph organisation and citation practice to teach students how to reason with multimodal evidence.

For teachers and lecturers, the findings suggest a practical sequence: guide students to interpret evidence, formulate a claim, connect evidence to the claim, write the reasoning, and revise the explanation. The sequence should not be treated as a rigid formula; its value lies in making visible the epistemic decisions that experienced writers often perform tacitly. The findings are also relevant to academic literacy and language education, where students can learn to verbalise visual and numerical information, build claims from evidence, and communicate disciplinary knowledge across modes.

Limitations and future research

Several limitations should be acknowledged. The study involved a limited number of undergraduates in one environmental literacy context and was implemented over a relatively short period. The findings therefore need to be interpreted cautiously when considering other disciplines, institutions, or student populations.

The quasi-experimental design included a comparison group, but full random assignment was not possible. Moreover, each instructional condition was represented by one intact class, so class-specific influences could not be fully separated from the instructional condition. The findings therefore support an association between multimodal evidence scaffolding and improved scientific writing rather than definitive causal proof. Larger multi-site studies with more classes per condition would strengthen causal and generalisability claims.

The study relied mainly on written products, worksheets, reflective notes, and coded textual evidence. These data did not capture all cognitive and behavioural processes involved in composing. Future research could use keystroke logging, screen recording, eye tracking, think-aloud protocols, or digital revision histories to examine how students attend to evidence while planning, drafting, and revising.

Future research should also examine transfer. It remains unclear whether students would apply similar evidence-use practices to research articles, laboratory reports, literature reviews, policy briefs, or argumentative essays. Comparative studies could investigate whether different modes support different forms of reasoning, such as spatial reasoning from maps, comparison from tables, trends from graphs, documentation from photographs, and contextual explanation from field notes.

Taken together, the findings contribute to writing research by showing that multimodal evidence can function as an epistemic scaffold when embedded in explicit instruction. The contribution is not simply that students improved after an intervention. More importantly, development can be understood as movement from information availability to evidence transformation, from evidence display to evidence integration, and from descriptive reporting to written reasoning.

CONCLUSION AND SUGGESTION

This study examined how multimodal evidence scaffolding shaped undergraduate students' scientific writing performance, evidence use, and written reasoning in an environmental literacy context. After initial performance was controlled, the experimental group achieved higher adjusted posttest scores across all writing dimensions, with the strongest effects concentrated in dimensions requiring students to select, interpret, integrate, and reason with evidence. Student-level analyses also showed substantial adjusted group differences in higher-level evidence use and advanced written reasoning.

The study contributes to writing research by reframing multimodal evidence as an epistemic scaffold for scientific writing. Maps, photographs, tables, graphs, field notes, and numerical data did not improve writing merely because they were present in the learning environment. Their value emerged when instruction guided students to identify what evidence showed, select what was relevant, interpret its meaning, connect it to a claim, and revise the explanation.

Qualitative findings showed visible textual movement from evidence mention to evidence interpretation, from single evidence to converging evidence, and from descriptive reporting to causal and integrative explanation. This progression reinforces the need to examine evidence relevance, specificity, interpretation, integration, and reasoning as distinct but connected dimensions of scientific writing.

For writing instruction across the curriculum, the findings indicate that lecturers should not only ask students to use data or support claims with evidence. Students need explicit support in reading multimodal representations, formulating claims, constructing warrants, coordinating evidence across modes, and revising explanations. Overall, scientific writing development depends not only on access to evidence, but on learning how to make evidence work epistemically within written reasoning.

REFERENCES

- Anggaira, A. S., & Wulantina, E. (2021). Pelatihan dan pendampingan penulisan artikel ilmiah dan manager reference bagi guru UPTD SMPN 2 Metro. *Dedikasi Jurnal Pengabdian Masyarakat*, 2(2), 258. <https://doi.org/10.32332/d.v2i2.3151>
- Araya Muñoz, I., & Majano Benavides, J. (2022). Didáctica universitaria en entornos virtuales. Experiencia en ciencias sociales. *Revista Electrónica Educare*, 26(3), 1–19. <https://doi.org/10.15359/ree.26-3.28>
- Baillargeon, K. (2025). Reflective orientations and genre navigation in dissertation writing: Insights from seven years of doctoral writing retreats. *Journal of Writing Research*.
- Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. University of Wisconsin Press.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.
- Boon, M., Orozco, M., & Sivakumar, K. (2022). Epistemological and educational issues in teaching practice oriented scientific research: Roles for philosophers of science. *European Journal for Philosophy of Science*, 12(1). <https://doi.org/10.1007/s13194-022-00447-z>
- Böck, M., & Pachler, N. (Eds.). (2013). *Multimodality and social semiosis: Communication, meaning making, and learning in the work of Gunther Kress*. Routledge.
- Cisneros, J. M. R., Pintado, I. L. U., Paredes, J. E. C., & Fernández, B. H. (2023). El portafolio como instrumento de evaluación y retroalimentación. Una revisión sistemática. *Revista De Climatología*, 23, 781–790. <https://doi.org/10.59427/rccli/2023/v23cs.781-790>

- Coronado, M. del C. R., & Martínez Álvarez, E. (2023). Entornos digitales para el fortalecimiento de las competencias escriturales en docentes de la básica secundaria. *Revista De Investigaciones De La Universidad Le Cordon Bleu*, 10(1), 36–45. <https://doi.org/10.36955/riulcb.2023v10n1.004>
- Dux Speltz, E., Roeser, J., Bouwman, W., Chukharev, E., & Torrance, M. (2026). Writing traits: Examining the consistency of behavioral patterns in writers' composing processes. *Journal of Writing Research*.
- Eliza, F., Sugirin, S., Triastuti, A., & Meutia, P. D. (2023). Strategi umpan balik langsung terhadap mahasiswa: Potret pada kelas menulis bahasa asing. *WL*, 3(1), 12. <https://doi.org/10.59562/wl.v3i1.50455>
- Faizah, S., Zaenudin, M., & Nugraha, D. (2023). Workshop penyusunan skripsi, publikasi ilmiah dan pembuatan poster ilmiah bagi mahasiswa tingkat akhir Jakarta Global University. *Jurnal Abdi Insani*, 10(3), 1543–1554. <https://doi.org/10.29303/abdiinsani.v10i3.1064>
- García Chitiva, M. del P. (2021). Aprendizaje colaborativo, mediado por internet, en procesos de educación superior. *Revista Electrónica Educare*, 25(2), 1–19. <https://doi.org/10.15359/ree.25-2.23>
- Graham, S., MacArthur, C. A., & Fitzgerald, J. (Eds.). (2013). *Best practices in writing instruction* (2nd ed.). The Guilford Press.
- Izah, I., Baalwi, M. A., & Wahyudi, A. (2022). Pengaruh model balanced literacy terhadap hasil belajar Bahasa Indonesia peserta didik kelas III SDN Sidoklumpuk Sidoarjo. *Jurnal Muassis Pendidikan Dasar*, 1(3), 227–236. <https://doi.org/10.55732/jmpd.v1i3.34>
- Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.
- Lida, U. M., Karsono, L. D. P., & Gunawan, K. (2024). Pengembangan modul menulis artikel ilmiah sebagai penunjang akselerasi publikasi ilmiah mahasiswa. *Alfabeta Jurnal Bahasa Sastra Dan Pembelajarannya*, 7(1), 1–11. <https://doi.org/10.33503/alfabeta.v7i1.4080>
- Liu, S., Liu, C., Samarapungavan, A., Gardner, S. M., Clase, K., & Pelaez, N. (2023). A framework for evidentiary reasoning in biology: Insights from laboratory courses focused on evolutionary tree thinking. *Science & Education*, 33(6), 1435–1466. <https://doi.org/10.1007/s11191-023-00435-6>

- Noa, S., Cruz, K. M. L. D. L., Apayco, L., Ramos, Z., & Lujano Ortega, Y. (2022). B learning en la enseñanza del idioma inglés en el nivel superior: una revisión sistemática. *Revista Innova Educación*, 4(2), 98–112. <https://doi.org/10.35622/j.rie.2022.02.006>
- Nurhasanah, E., & Maspuroh, U. (2021). Pelatihan menulis karya ilmiah bagi guru guru SDN Karyamakmur II dan SDN Segarjaya 1 Kecamatan Batujaya Kabupaten Karawang. *Kangmas Karya Ilmiah Pengabdian Masyarakat*, 2(2), 128–136. <https://doi.org/10.37010/kangmas.v2i2.303>
- Pratama, U. N., & Surahman, E. (2023). Desain online project based learning untuk pembelajaran menulis karya ilmiah pendidikan seni budaya. *Jurnal Ilmiah Pendidikan Citra Bakti*, 10(1), 184–198. <https://doi.org/10.38048/jipcb.v10i1.980>
- Rizqi, H. Y., & Purwanti, K. Y. (2023). Pelatihan penyusunan artikel ilmiah dalam jurnal terindeks Sinta untuk meningkatkan profesionalisme guru dan mahasiswa. *Intisari*, 1(1), 33–38. <https://doi.org/10.58227/intisari.v1i1.35>
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). SAGE.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Spivey, N. N. (1997). *The constructivist metaphor: Reading, writing, and the making of meaning*. Academic Press.
- Sugianto, A. (2022). [Book review] *Systemic functional linguistics (SFL) and teacher education in EFL context* by Wawan Gunawan. *JLA (Jurnal Lingua Applicata)*, 5(2), 145. <https://doi.org/10.22146/jla.68697>
- Syah, S. P. (2024). A multimodal critical discourse analysis on political speech by presidential candidate for the Republic of Indonesia 2024. *Suar Betang*, 19(1), 29–46. <https://doi.org/10.26499/surbet.v19i1.14721>
- Syarif, R., Arifin, M. B., & Siddik, M. (2021). Pengembangan bahan ajar menulis deskripsi menggunakan model circuit learning pada siswa kelas VII SMP di Samarinda. *Diglosia: Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, 4(2), 219–226. <https://doi.org/10.30872/diglosia.v4i2.70>
- Tafonao, T., & Widjaja, F. I. (2021). Pembinaan dan pelatihan penulisan karya ilmiah di kalangan dosen Sekolah Tinggi Teologi Kadesi Yogyakarta. *Yumary Jurnal*

- Pengabdian Kepada Masyarakat, 1(3), 127–134.
<https://doi.org/10.35912/yumary.v1i3.154>
- Tesfa, A., Amogne, D., Teklessellassie, Y., & Negede, D. (2025). Effects of text structure instruction on expository text writing: An intervention for grade 7 English as foreign language learners. *Journal of Writing Research*.
- Thaariq, Z. Z. A., Nurdyanto, R., & Sulfa, D. M. (2023). Masalah-masalah dalam paradigma pembelajaran sains modern. *JND*, 2(2), 112–121.
<https://doi.org/10.59291/jnd.v2i2.38>
- Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge University Press.
- Vashti, T. A., Hairida, H., & Hadi, L. (2021). Deskripsi kemampuan literasi sains siswa mata pelajaran ilmu pengetahuan alam. *EduChem*, 1(2).
<https://doi.org/10.26418/educchem.v1i2.41108>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- Wahyudin, A., Yanto, H., Sari, M. P., Nurkhin, A., & Budiantoro, R. A. (2023). Optimalisasi keterampilan menulis berbasis kecerdasan buatan bagi guru SMK Negeri 1 Demak. *PSNPM*, 1. <https://doi.org/10.61142/psnpm.v1.72>
- Walvoord, B. E., & McCarthy, L. P. (1990). *Thinking and writing in college: A naturalistic study of students in four disciplines*. National Council of Teachers of English.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Wuryaningrum, R., Muji, M., & Yolanda, Y. (2023). Representasi sosiolinguistik pada kognisi wacana persuasi kopi. *Indonesian Language Education and Literature*, 8(2), 362. <https://doi.org/10.24235/ileal.v8i2.11556>
- Wyse, D., Aarts, B., Anders, J., de Gennaro, A., Dockrell, J. E., Manyukhina, Y., Sing, S., & Torgerson, C. (2025). Teaching grammar and writing: A randomized controlled trial and implementation and process evaluation of Englicious. *Journal of Writing Research*.
- Yoon, H. J., Kim, S. H., & Joaquin, A. D. L. (2026). Planning to perform: Association of pre-task strategies with linguistic complexity and cohesion in argumentative essays. *Journal of Writing Research*.