Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika
https://journal-center.litpam.com/index.php/e-Saintika/index

July 2026 Vol. 10, No. 2
e-ISSN: 2615-6881
pp. 967-991

Learning Outcomes and Metacognitive Regulation after SLR-Oriented STEM-Integrated Research-Based Learning in Plant Metabolites

[*1a, 2] Ika Priantari, [1a] Suratno, [1a] Dwi Wahyuni, [1b] Dafik, [3] Yoga Dwi Windy Kusuma Ningtyas

[1a] Doctoral Program in Science Education; [b] Master's Program in Mathematics Education, Faculty of Teacher Training and Education, Jl. Kalimantan No. 37, Jember 68121, Indonesia.
[2] Biology Education Program, Faculty of Teacher Training and Education, Universitas Muhammadiyah Jember, Jl. Karimata No. 49, Jember 68121, Indonesia.
[3] International Education Development Program, Division of Educational Sciences, Graduate School of Humanities and Social Sciences, Hiroshima University, Japan 1 Chome-5-1, Kagamiyama, Higashihiroshima, Hiroshima Prefecture 739-0046, Japan.

*Corresponding Author e-mail: 200230104002@mail.unej.ac.id

Received: February 2026; Revised: June 2026; Published: July 2026

Abstract

The integration of Science, Technology, Engineering, and Mathematics (STEM) with Research-Based Learning (RBL) may provide an authentic context for students to engage with scientific evidence while regulating their learning processes. This study aimed to (1) determine the extent to which students' learning-outcome scores changed after participating in an SLR-oriented STEM-integrated RBL sequence and (2) describe students' post-intervention metacognitive-regulation profiles. Sixty-three undergraduate Biology Education students participated in a quantitative pre-experimental study using a one-group pretest-posttest design. The three-meeting intervention required students to formulate a research question, search and screen literature, organize and synthesize evidence on coffee metabolites and roasting conditions, prepare a scientific article, and reflect on their work. Learning outcomes were assessed using four constructed-response items with a maximum score of 100; the available 65-response instrument-testing dataset yielded item-total correlations of .254–.438 and Cronbach's α = .530. Metacognitive regulation was measured after the intervention using a 14-item questionnaire covering planning, monitoring, and evaluating (overall Cronbach's α = .923). Mean learning-outcome scores increased from 26.03 (SD = 9.77) at pretest to 65.92 (SD = 12.98) at posttest, with a mean difference of 39.89, $t(62) = 23.498$, $p < .001$, 95% CI [36.50, 43.28]. The normalized gain was .538 (moderate), and Cohen's dz was 2.960. The post-intervention metacognitive profile was good for planning (81.3%) and sufficient for monitoring (73.2%) and evaluating (69.7%). These findings provide evidence of substantial within-group learning-outcome change together with a differentiated post-intervention metacognitive-regulation profile. The design supports interpretation at the within-group and post-intervention levels, while causal effects and longitudinal metacognitive development remain questions for future comparative research.

Keywords: Integrated STEM; Metacognitive Regulation; Plant Metabolites; Research-Based Learning; Systematic Literature Review

How to Cite: Priantari, I., Suratno, S., Wahyuni, D., Dafik, D., & Ningtyas, Y. D. W. K. (2026). Learning Outcomes and Metacognitive Regulation after SLR-Oriented STEM-Integrated Research-Based Learning in Plant Metabolites. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika, 10*(2), 967-991. https://doi.org/10.36312/e-saintika.v10i2.5768

10.36312/e-saintika.v10i2.5768

Copyright© 2026, Priantari et al.
This is an open-access article under the CC-BY-SA License.

INTRODUCTION

Metacognition refers to an individual's awareness and control of cognitive processes used during learning. According to Schraw and Moshman (1995), metacognition comprises two interrelated dimensions: metacognitive knowledge and metacognitive regulation. Metacognitive knowledge concerns learners' knowledge of themselves, learning tasks, and available strategies, whereas metacognitive regulation concerns the processes used to plan, monitor, and evaluate learning activities (Stanton et al., 2019). This distinction remains relevant in contemporary biology education. Recent work with undergraduate life-science students shows that learners differ substantially in when and how they plan their learning, monitor their understanding, and evaluate the effectiveness of their strategies, indicating that metacognitive regulation remains an important instructional concern beyond simple awareness of learning strategies (Stanton et al., 2024).

Higher education increasingly expects students to interpret evidence, integrate knowledge across disciplinary boundaries, and address complex scientific problems. These demands are particularly relevant in biology, where students must connect conceptual knowledge with data interpretation, scientific reasoning, and decisions about appropriate learning strategies. However, knowing that a learning strategy exists does not ensure that students can select, monitor, and revise that strategy effectively. Studies of undergraduate biology and life-science students have shown that students may recognize useful approaches yet still experience difficulty accurately monitoring their understanding or following through on planned strategies (Stanton et al., 2019; Stanton et al., 2024). Consequently, biology education requires instructional environments that support not only conceptual understanding but also students' regulation of their own learning processes. Recent studies similarly highlight the relevance of metacognitive support for science learning and academic performance (Helendra et al., 2023; Sadykova et al., 2024; Stanton et al., 2019).

Metacognitive regulation is commonly characterized by three core processes: planning, monitoring, and evaluating. Planning involves identifying goals, selecting appropriate strategies, and organizing resources before or at the beginning of a task. Monitoring involves checking understanding, judging progress, and recognizing difficulties while learning is taking place. Evaluating involves judging the quality of the completed process or outcome and deciding whether strategies should be maintained or modified in future tasks (Zohar & Barzilai, 2015). Through repeated cycles of planning, monitoring, and evaluating, learners can adjust how they approach complex academic tasks (Schraw, 1998; Stanton et al., 2021). Recent research involving first-year life-science majors continues to support this three-process framework and further shows that students may require different forms of instructional guidance at different points in the regulatory cycle, particularly when monitoring their own understanding (Stanton et al., 2024).

Instructional tasks that require students to work with authentic scientific problems may provide useful contexts in which these regulatory processes become visible. Integrated Science, Technology, Engineering, and Mathematics (STEM) education seeks to connect disciplinary knowledge and practices around problems that cannot be addressed adequately through a single disciplinary perspective. Science contributes explanations of natural phenomena and evidence-based reasoning; technology involves tools, processes, and applications; engineering addresses design or operating constraints associated with practical problems; and mathematics supports quantitative representation, analysis, and interpretation (Brown et al., 2011; Marrero et al., 2014; White, 2014; Winarni et al., 2016; Xie et al., 2015). Contemporary integrated-STEM research emphasizes that meaningful integration depends not merely on naming the four disciplines but on making disciplinary connections explicit within authentic problem contexts and purposeful student activity (Hiwatig et al., 2024).

Research-Based Learning (RBL) provides one means of organizing such student activity. RBL integrates research processes and research findings into instruction, positioning students as active participants in inquiry rather than passive recipients of information (Sota & Peltzer, 2017). Previous work has associated RBL with academic achievement, independent learning, knowledge construction, and engagement with authentic research practices (Brew, 2007). RBL may incorporate contextual learning, collaborative activity, problem solving, inquiry, and hands-on or minds-on engagement, thereby creating opportunities for higher-order thinking (Monalisa et al., 2019). More recent higher-education studies likewise describe research-based curricula as requiring students to formulate questions, engage with research literature or evidence, analyze information, and communicate conclusions, while also showing that the quality of students' research experiences depends on how clearly those activities are structured and supported (Zou et al., 2024).

RBL is particularly relevant to science education because scientific learning involves identifying problems, collecting or locating evidence, analyzing information, interpreting findings, and drawing defensible conclusions. Through research-oriented activities, students may practice critical thinking, analytical reasoning, problem solving, and evidence-based decision making (Irwanto et al., 2017; Syofyan et al., 2019; Taleb & Chadwick, 2016). In biology education specifically, Daryanes and Sayuti (2023) reported positive student perceptions of RBL for developing critical-thinking practices, while more recent research-based learning and project studies in biology and plant physiology have extended this approach toward reflective thinking and broader meta-skills (Supit & Winardi, 2024; Yennita et al., 2025).

Suntusia et al. (2019) proposed a seven-stage RBL framework comprising (1) problem identification, (2) development of problem-solving strategies, (3) data collection and hypothesis formulation, (4) data analysis, (5) generalization and conjecture development, (6) focused group discussion, and (7) preparation of a research report. These stages create opportunities for students to participate actively throughout a research process. RBL has also been associated with learning motivation, problem-solving ability, collaborative learning, communication, and experience in managing research projects (Suntusia et al., 2019). Collaborative engagement is important because peer interaction can support knowledge sharing, feedback, and collective problem solving (Healey et al., 2014). These features make RBL potentially compatible with metacognitive regulation: students must decide how to approach a research problem, monitor whether selected evidence addresses that problem, and evaluate the adequacy of their conclusions and products.

The closest empirical literature indicates substantial overlap among RBL, STEM integration, biology learning, and metacognitive processes, but also reveals important differences from the present instructional design. Komaria et al. (2024), for example, examined STEM-oriented RBL materials with higher-education students and explicitly assessed planning, monitoring, and evaluation. This study provides close evidence that STEM-RBL and metacognition can be combined, but its instructional context, measurement approach, and analytical objectives differed from an SLR-based learning sequence in undergraduate biology. Yennita et al. (2025) investigated a research-based project in an undergraduate Plant Physiology course using a non-equivalent pretest-posttest control-group design. Their study focused on a broad meta-skills construct and reported differences between experimental and control groups, but it did not examine planning, monitoring, and evaluating as the specific post-intervention regulatory profile investigated here.

Daryanes and Sayuti (2023) examined RBL in undergraduate Biology Education but focused on students' perceptions of its contribution to critical-thinking skills rather than STEM integration or metacognitive regulation. At the broader higher-education level, Zou et al. (2024) showed that undergraduate experiences of research-based curricula vary according to how research opportunities are structured and communicated, while Hiwatig et al. (2024) examined the operationalization of integrated STEM and highlighted the importance of explicit disciplinary integration. Taken together, these studies delineate a specific research gap concerning how an SLR-oriented STEM-RBL sequence can integrate authentic biological inquiry with planning, monitoring, and evaluating in undergraduate biology learning. What remains less clearly documented is an instructional sequence in which undergraduate biology students use a Systematic Literature Review (SLR) as the research task, explicitly integrate STEM perspectives around an authentic plant-metabolite problem, and engage in activities aligned with planning, monitoring, and evaluating while learning outcomes and the post-intervention metacognitive profile are examined separately.

The present study addresses this gap through an SLR-oriented STEM-integrated RBL sequence focused on plant metabolites in coffee. Students formulated a research focus, identified and screened literature, extracted and organized evidence, synthesized findings, prepared a scientific article, presented their work, received feedback, and reflected on the quality of the process and product. The substantive contribution therefore lies not merely in using coffee roasting as the biological context, but in structuring an SLR as the central research-learning task and explicitly aligning its stages with STEM disciplinary perspectives and metacognitive-regulation opportunities.

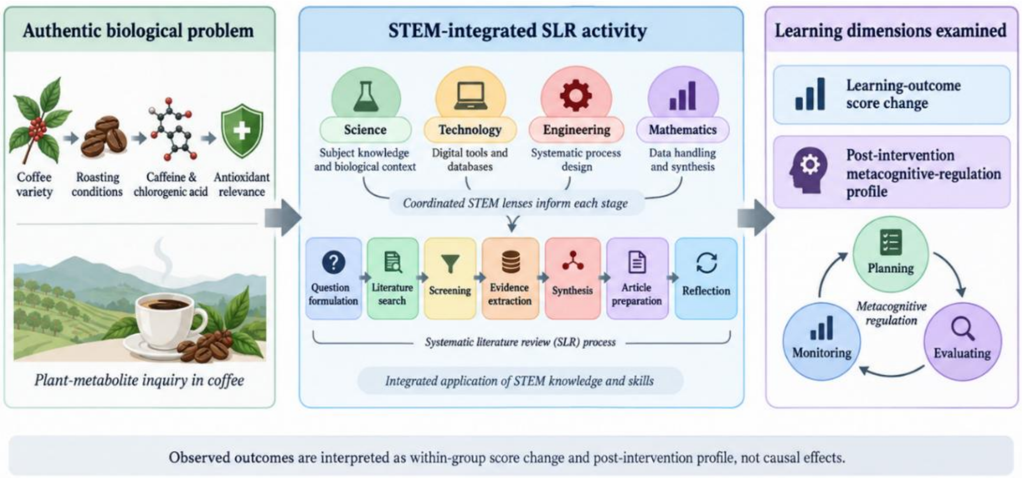

Figure 1. Conceptual framework of the SLR-oriented STEM-integrated Research-Based Learning sequence linking the authentic plant-metabolite problem, integrated STEM practices, research-based learning activities, and the outcomes examined in the study.

Figure 1 summarizes this conceptual relationship. It positions the authentic plant-metabolite problem as the instructional context, the SLR-oriented STEM-RBL sequence as the learning process, and learning-outcome score change and post-intervention metacognitive regulation as the two outcomes examined in the study. Accordingly, the framework positions the two outcomes as complementary indicators examined within the scope of the present study design. Planning, monitoring, and evaluating are treated as regulatory processes that can be engaged during different parts of the research task rather than as outcomes assumed to improve automatically because students participated in RBL. This distinction is important because research-oriented instruction can provide opportunities for regulation, but students may differ in how effectively they use those opportunities.

The study design defines two complementary levels of interpretation. The pretest-posttest component documents within-group learning-outcome change, whereas the post-intervention questionnaire characterizes students' metacognitive-regulation profile across planning, monitoring, and evaluating. Accordingly, the present study focuses on these two evidence streams and does not extend the analysis to correlational, regression, or predictive relationships between metacognitive regulation and achievement.

Accordingly, the research questions were: (1) To what extent did students' learning-outcome scores change after participating in STEM-integrated Research-Based Learning? and (2) What was the students' post-intervention metacognitive-regulation profile across planning, monitoring, and evaluating?

METHOD

Research Design

This study used a quantitative pre-experimental one-group pretest-posttest design, supplemented by descriptive analysis of post-intervention metacognitive regulation. The design was selected to examine within-group changes in learning-outcome scores following participation in the instructional sequence. Because the design did not include a concurrent comparison group, the observed pretest-posttest difference was interpreted as within-group change rather than as evidence of a causal treatment effect. One-group pretest-posttest designs remain vulnerable to alternative explanations such as testing, maturation, history, and regression to the mean (Marsden & Torgerson, 2012).

Metacognitive regulation was measured only after the intervention. It was therefore analyzed as a post-intervention profile rather than as a pretest-posttest outcome. Accordingly, the study addressed two empirically distinct outcomes: changes in learning-outcome scores and the post-intervention profile of planning, monitoring, and evaluating. Metacognitive regulation was interpreted as a post-intervention profile, consistent with the timing of its measurement.

Participants and Sampling

The participants were 63 undergraduate students enrolled in one intact Biology Education (Tadris Biologi) class at Universitas Islam Negeri Kiai Haji Achmad Siddiq Jember. The cohort comprised 59 women (93.7%) and four men (6.3%). Total sampling was used because all students in the intact class who met the eligibility criteria were included.

Eligibility required active enrolment, completion of the pretest and posttest, completion of the post-intervention metacognitive-regulation questionnaire, participation in the three instructional meetings, and submission of the assigned SLR product. Students with incomplete outcome data were excluded. Approximately 44 students (69.8%) reported an interest in plant-related courses. Only participant characteristics that could be verified from the available study records are reported.

STEM-Integrated Research-Based Learning Intervention

Intervention Schedule and Learning Organization

The intervention was conducted over three meetings. The instructional worksheet specified three instructional periods per meeting (3 meetings × 3 JP), with students working collaboratively in groups of five to six. Students investigated an authentic plant-metabolite problem concerning the relationships among coffee variety, roasting conditions, caffeine, and chlorogenic-acid content.

The principal learning product was an SLR-based scientific article. Through this task, students formulated a research focus, searched and screened literature, organized evidence, synthesized findings, prepared conclusions, presented their work, received feedback, and reflected on the learning process.

Seven-Stage Research-Based Learning Sequence

The intervention followed seven documented RBL stages. First, students formulated a research problem or question concerning coffee varieties, roasting processes, and relevant plant metabolites. Second, they planned the literature-based investigation and identified articles according to predetermined eligibility criteria. Third, they collected and screened potentially relevant literature using quality-assessment questions concerning article indexing, relevance to the research problem, and correspondence with the technological and engineering aspects of coffee roasting.

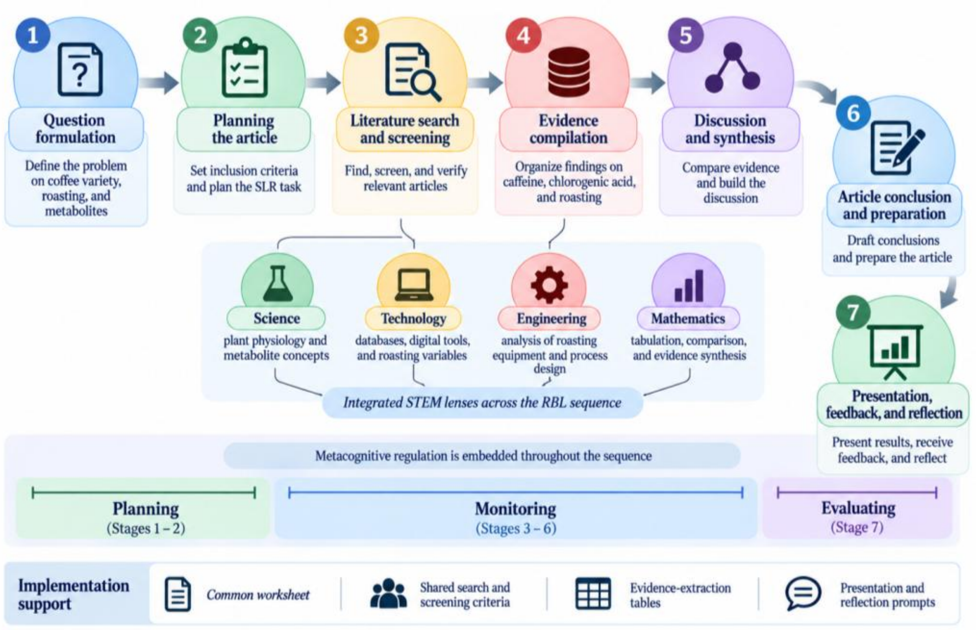

Figure 2. Seven-stage implementation of the STEM-integrated Research-Based Learning sequence across the three-meeting SLR project.

Fourth, students compiled evidence concerning coffee variety, roasting temperature and duration, processing equipment, caffeine, chlorogenic acid, and related outcomes. Fifth, they compared evidence across articles and constructed an evidence-based discussion. Sixth, students formulated conclusions that addressed the original research question and prepared their scientific article. Seventh, each group presented the article, responded to peer feedback, identified strengths and weaknesses of the project, and reflected on the benefits and challenges experienced during the learning process.

The sequence positioned students as active users of scientific evidence rather than solely as recipients of disciplinary content, consistent with the central characteristics of RBL described by Suntusia et al. (2019). The complete instructional progression is summarized in Figure 2. The figure shows how the seven RBL stages were implemented across the three-meeting project and where planning, monitoring, and evaluating were emphasized.

Operationalization of STEM Integration

STEM was integrated across the RBL sequence rather than taught as four independent disciplinary components. Science involved plant physiology, primary and secondary metabolites, antioxidant-related concepts, and interpretation of caffeine and chlorogenic-acid evidence. Technology involved the use of scholarly information sources and digital processes for locating, screening, organizing, and interpreting scientific literature, together with consideration of roasting temperature and duration as technological variables.

Mathematics involved the organization, comparison, and interpretation of quantitative information such as roasting temperature, duration, metabolite concentration, and percentage differences. Engineering was represented through analysis of coffee-roasting equipment, processing conditions, and operating variables rather than through construction and iterative testing of an engineering prototype. This interpretation follows the broader concept of integrated STEM in which disciplinary connections are made explicit around an authentic problem (Kelley & Knowles, 2016; Roehrig et al., 2021). More recent empirical work likewise emphasizes real-world contextualization, explicit disciplinary integration, evidence-based reasoning, technological practices, collaboration, and engineering-related problem solving as important characteristics of integrated STEM implementation (Hiwatig et al., 2024). Accordingly, the instructional sequence is described as STEM-integrated RBL rather than as a complete engineering-design intervention.

Implementation Consistency and Metacognitive Alignment

Implementation consistency was supported through a common instructional protocol. All groups received the same worksheet specifying the seven-stage RBL sequence, group size of five to six students, three-meeting schedule, article-selection criteria, quality-assessment questions, evidence-compilation tables, discussion and synthesis prompts, article-preparation requirements, presentation procedures, and final reflection activities.

Implementation consistency was documented through a standardized instructional protocol applied across groups. Fidelity was therefore characterized through protocol standardization rather than a numerical observational coefficient.

The sequence provided explicit opportunities for metacognitive regulation. Planning was emphasized when students formulated research questions and organized their literature-search strategy. Monitoring was embedded in literature screening, evidence checking, data compilation, interpretation, and manuscript development. Evaluating became particularly visible when students reviewed their conclusions, presented their work, considered peer feedback, and reflected on the adequacy of the final product. These dimensions are consistent with established conceptualizations of metacognitive regulation as planning, monitoring, and evaluating (Chatzipanteli et al., 2014; Dafik et al., 2019; Stanton et al., 2021). Recent research with undergraduate life-science students also identifies planning, monitoring, and evaluating as central regulatory processes that may be activated differently across stages of learning (Stanton et al., 2024).

Table 1. Conceptual Alignment of the STEM-RBL Sequence and Metacognitive Regulation

STEM-RBL stage	Student activity	Targeted metacognitive dimension	Observable activity	Documented instructional source
Research-question formulation	Define the coffee-metabolite problem and formulate the research focus	Planning	Identifies the problem, learning target, and required information	Research-question section of the worksheet
Research planning and article identification	Plan the literature-based investigation and identify potentially relevant articles	Planning	Organizes the search task and applies article-selection criteria	Article-search and planning table
Literature collection and screening	Retrieve and evaluate candidate studies	Monitoring	Checks indexing, relevance, and correspondence with the research focus	Quality-assessment and screening table
Evidence compilation	Organize findings concerning coffee variety, roasting conditions, caffeine, chlorogenic acid, and related outcomes	Monitoring	Checks completeness and relevance of extracted evidence	Evidence-compilation table
Discussion and evidence synthesis	Compare findings and construct an evidence-based explanation	Monitoring	Examines consistency among evidence and links findings with biological concepts	Discussion section of the worksheet and article draft
Conclusion and article preparation	Develop conclusions and prepare the scientific article	Monitoring and evaluating	Checks whether conclusions address the research question and are supported by evidence	Conclusion prompt and article draft
Presentation, feedback, and reflection	Present findings, respond to feedback, and identify strengths and weaknesses	Evaluating	Judges the adequacy of strategies and products and identifies improvements	Presentation and reflection section of the worksheet

Figure 2 provides a procedural overview of the intervention, whereas Table 1 provides a detailed mapping between each RBL stage and the corresponding metacognitive-regulation process. The stage-specific relationship between instructional activities and regulatory processes is presented in Table 1.

As shown in Table 1, metacognitive regulation was distributed across the instructional sequence rather than treated as a separate learning activity. Planning occurred primarily at the beginning of the research process, monitoring extended across literature searching, evidence handling, synthesis, and article development, while evaluating became most explicit when students judged the adequacy of their conclusions and responded to peer feedback.

Learning-Outcome Assessment

Learning outcomes were measured using four constructed-response items administered before and after the intervention. The blueprint aligned the assessment with the course learning outcome, “understanding the concepts, principles, and functions of plants for life.” Items 1–3 assessed analysis (C4) and carried maximum scores of 20 points each, whereas Item 4 assessed evidence-based conclusion drawing (C5) and carried a maximum score of 40 points. The total possible cognitive score was therefore 100.

Although the instructional blueprint also listed an article product as a fifth activity, this product was part of the learning project and was not included in the 0–100 pretest-posttest cognitive score. The learning-outcome analyses reported in this study therefore refer exclusively to Items 1–4.

Five biology-education experts reviewed the assessment for clarity of instructions, alignment with learning objectives, content accuracy, and language. Content-validity evidence was summarized using Aiken’s V (Aiken, 1985). The recorded coefficients ranged from .80 to 1.00. Aiken’s V remains widely used for quantifying expert judgments in contemporary educational instrument development (Kania et al., 2024).

In the 65-response instrument-testing dataset, item-total correlations ranged from .254 to .438, and all four scored items were recorded as meeting the validity criterion in the original analysis (p = .001–.041). Cronbach’s α for the four-item assessment was .530. Given the concise four-item constructed-response format, this coefficient is reported as the available index of internal consistency and considered alongside the instrument’s content-validity and scoring evidence. Responses were scored using the supplied analytic rubric. Independent double scoring and inter-rater agreement were not documented; therefore, no inter-rater reliability coefficient is reported.

Metacognitive-Regulation Instrument

Metacognitive regulation was assessed after the intervention using a 14-item questionnaire adapted from Altındağ and Senemoğlu (2013), Chatzipanteli et al. (2014), Dafik et al. (2019), Hindun et al. (2020), and Quigley et al. (2018). The questionnaire comprised planning (four items), monitoring (five items), and evaluating (five items). The three-dimensional structure is consistent with established metacognitive-regulation frameworks and with more recent biology-education research describing planning, monitoring, and evaluating as interrelated regulatory processes used by undergraduate learners (Stanton et al., 2021; Stanton et al., 2024).

Responses were recorded on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). All questionnaire statements were positively directed; therefore, reverse scoring was not required. Higher item scores represented stronger perceived metacognitive regulation.

Raw scores were converted to percentage scores using obtained score divided by maximum possible score multiplied by 100. For the complete scale, the maximum possible score was 70; the corresponding maxima were 20 for planning and 25 each for monitoring and evaluating. Following the classification used in the adapted instrument (Hindun et al., 2020), percentage scores were interpreted as very good (86–100), good (76–85), sufficient (60–75), low (55–59), or very low (0–54). These thresholds were applied consistently to the overall profile and the three dimensions.

Instrument Validity and Reliability

Five validators with expertise in biology education evaluated the learning-outcome assessment. For the metacognitive-regulation questionnaire, item-level Aiken's V coefficients ranged from .867 to 1.000, with a mean of .933. The use of Aiken's V preserves the original validation procedure while remaining consistent with contemporary applications of expert-based content-validity analysis (Aiken, 1985; Kania et al., 2024).

Cronbach's alpha coefficients for the questionnaire were .837 for planning, .880 for monitoring, .811 for evaluating, and .923 for the complete scale. The available reliability record comprised 65 responses used for instrument testing. Because cohort independence was not specified in the archived documentation, the dataset is reported as an instrument-testing dataset.

Data Collection Procedure

Data collection followed a fixed sequence. Before the first instructional meeting, participants completed the learning-outcome pretest under standardized classroom conditions. They then participated in the three-meeting STEM-integrated RBL sequence and prepared the required SLR article product.

After the final instructional meeting, participants completed the learning-outcome posttest using the same indicators, total score range, and scoring framework as the pretest. The metacognitive-regulation questionnaire was then administered once after completion of the posttest, consistent with the questionnaire instructions. Participant codes rather than names were used to link responses across instruments. Data were checked for completeness before analysis, and only students with complete pretest, posttest, and questionnaire records were included in the final dataset.

Data Analysis

Descriptive analyses included means, standard deviations, score ranges, frequencies, and percentages. A paired-samples t-test was used to examine the mean difference between pretest and posttest learning-outcome scores. The paired difference was additionally reported with its standard error and 95% confidence interval.

Normalized gain was calculated as (Mpost − Mpre)/(100 − Mpre). N-gain was interpreted as low (< .30), moderate (.30–.69), or high (≥ .70), following Hake (1998). The standardized magnitude of the paired difference was summarized using Cohen's d for dependent observations (dz), calculated as the mean paired difference divided by the standard deviation of the paired differences. Reporting the specific standardizer is important because several formulations of Cohen's d can be used in repeated-measures studies (Lakens, 2013).

Statistical significance was set at $\alpha = .05$, and exact p values were reported except when $p < .001$. The statistical analysis focused on the outcomes aligned directly with the two research questions: descriptive statistics, the paired-samples comparison, 95% confidence interval, normalized gain, and within-group standardized change. Participant-level raw data were not available for additional verified reanalysis; therefore, no unverified assumption diagnostics, correlation analyses, regression models, or predictor claims were included.

RESULTS AND DISCUSSION

Participant Characteristics

All 63 participants provided complete pretest, posttest, and post-intervention metacognitive-regulation data and were included in the final analyses. The cohort comprised 59 women (93.7%) and four men (6.3%). Approximately 44 students (69.8%) reported an interest in plant-related courses. No participant included in the analytic sample had missing data for the outcomes reported below.

Learning-Outcome Scores Before and After the Intervention

Descriptive statistics for the learning-outcome pretest, posttest, and paired-score difference are presented in Table 2. The mean pretest score was 26.03 (SD = 9.77), with observed scores ranging from 4 to 50. The mean posttest score was 65.92 (SD = 12.98), with observed scores ranging from 31 to 90. The mean within-participant increase from pretest to posttest was 39.89 points (SD of paired differences = 13.48).

Table 2. Descriptive Statistics for Pretest, Posttest, and Paired-Difference Scores

Measure	N	Mean	SD	Minimum	Maximum
Pretest	63	26.03	9.77	4	50
Posttest	63	65.92	12.98	31	90
Paired difference	63	39.89	13.48	—	—

Paired Pretest-Posttest Difference and Learning Gain

The statistical analysis of the pretest-posttest difference is summarized in Table 3. The mean paired difference was 39.89 points, with a standard error of 1.70. The 95% confidence interval for the mean difference ranged from 36.50 to 43.28 points. The paired-samples analysis showed a statistically significant difference between pretest and posttest scores, $t(62) = 23.498$, $p < .001$. The normalized gain was .538 and was classified as moderate. Cohen's dz was 2.960, representing the standardized magnitude of the within-participant pretest-posttest difference.

Table 3. Paired-Samples Test and Learning-Gain Statistics

Statistic	Value
Mean difference	39.89
SD of paired differences	13.48
Standard error	1.70
95% confidence interval	[36.50, 43.28]
t(df)	23.498 (62)
p	< .001
N-gain	.538 (moderate)
Cohen's dz	2.960

These results answer the first research question by documenting the magnitude of the within-group change in learning-outcome scores after participation in the STEM-integrated RBL sequence.

Measurement Evidence for the Metacognitive-Regulation Questionnaire

Before describing the post-intervention metacognitive profile, the validity and internal-consistency evidence for the questionnaire is summarized in Table 4. Item-level Aiken's V coefficients ranged from .867 to 1.000, with a mean of .933. Cronbach's alpha coefficients were .837 for planning, .880 for monitoring, .811 for evaluating, and .923 for the overall 14-item scale.

Table 4. Validity and Reliability Evidence for the Metacognitive-Regulation Questionnaire

Component	Evidence	Result	Interpretation
Content validity	Aiken's V	.867–1.000; mean = .933	High content validity
Planning (4 items)	Cronbach's α	.837	Good internal consistency
Monitoring (5 items)	Cronbach's α	.880	Good internal consistency
Evaluating (5 items)	Cronbach's α	.811	Good internal consistency
Overall scale (14 items)	Cronbach's α	.923	Excellent internal consistency

Overall Post-Intervention Metacognitive-Regulation Profile

The distribution of students across the overall metacognitive-regulation categories is presented in Table 5. Of the 63 students, eight (12.7%) were classified as very good, 23 (36.5%) as good, 30 (47.6%) as sufficient, and two (3.2%) as low. No student was classified as very low.

Table 5. Distribution of Students' Post-Intervention Metacognitive-Regulation Levels

Category	Number of students	Percentage (%)
Very good	8	12.7
Good	23	36.5
Sufficient	30	47.6
Low	2	3.2
Very low	0	0.0

These categories describe students' self-reported metacognitive regulation at the post-intervention measurement point. They do not represent change from a pre-intervention baseline because the questionnaire was administered only once.

Metacognitive Regulation Across Planning, Monitoring, and Evaluating

Dimension-specific percentage scores and classifications are presented in Table 6. Planning obtained the highest percentage score at 81.3% and was classified as good. Monitoring obtained 73.2%, while evaluating obtained 69.7%; both were classified as sufficient.

Table 6. Metacognitive Regulation by Dimension

Dimension	Percentage (%)	Category
Planning	81.3	Good
Monitoring	73.2	Sufficient
Evaluating	69.7	Sufficient

These results answer the second research question by characterizing the students' post-intervention metacognitive-regulation profile. Planning was the highest-scoring dimension, whereas monitoring and evaluating remained within the sufficient category. The dimension scores therefore provide a comparative post-intervention profile of planning, monitoring, and evaluating.

Discussion

Learning-Outcome Change after STEM-Integrated RBL

As reported in Tables 2 and 3, students' learning-outcome scores increased from a pretest mean of 26.03 to a posttest mean of 65.92, corresponding to a mean paired difference of 39.89 points. The normalized gain of .538 was classified as moderate, while Cohen's dz of 2.960 indicated a very large standardized within-participant difference. These results document a substantial change in learning-outcome scores within the cohort after the three-meeting STEM-integrated RBL sequence. The magnitude of this change provides strong descriptive evidence of learning progression within the cohort, while attribution of the change specifically to the intervention would require a comparative design.

The magnitude of the score change is educationally notable, but it should be interpreted in relation to the learning activities embedded in the intervention and the limitations of the study design. Active and inquiry-oriented learning commonly supports achievement by requiring students to retrieve, organize, explain, and apply information rather than receive it passively (Freeman et al., 2014; Theobald et al., 2020). Likewise, project-based learning has shown positive average effects on academic achievement, although the magnitude of those effects varies according to instructional design, duration, assessment, participant characteristics, and implementation quality (Chen & Yang, 2019).

The present SLR task incorporated several of these cognitively active processes. Students formulated a focused research question, searched and screened scientific literature, extracted information concerning coffee metabolites and roasting conditions, organized quantitative evidence, compared findings across studies, prepared an evidence-based synthesis, and revised their article following feedback. These activities required repeated engagement with course concepts rather than simple recall. Thus, the observed score increase is compatible with mechanisms proposed in the active-learning and inquiry literature. These mechanisms provide theoretically plausible explanations for the observed learning progression and offer directions for testing in future comparative studies.

The instructional sequence also incorporated characteristics commonly associated with integrated STEM, including an authentic problem context, explicit links among disciplinary perspectives, student-centered inquiry, collaboration, and purposeful use of disciplinary practices (Holmlund et al., 2018; Kelley & Knowles, 2016; Thibaut et al., 2018). More recent research on the operationalization of integrated STEM likewise emphasizes that meaningful integration depends on how disciplinary connections, authentic contexts, technology practices, and engineering-related reasoning are enacted within the learning activity rather than simply on labeling an activity as STEM (Hiwatig et al., 2024).

In the present intervention, engineering was intentionally operationalized through analysis of coffee-roasting equipment, processing conditions, and operating variables rather than through a full prototype-design cycle. This operationalization is consistent with the study's description of the sequence as STEM-integrated RBL rather than as a complete engineering-design intervention (Roehrig et al., 2021).

Interpreting the Magnitude of the Learning-Outcome Difference

The very large Cohen's dz highlights the magnitude of change observed within this cohort and warrants interpretation in relation to the within-participant design. A standardized within-participant difference is conceptually distinct from a between-group treatment effect, and the relatively low pretest mean provided substantial numerical scope for improvement. Repeated exposure to the same learning domain and assessment indicators may also have contributed to the observed difference.

As is typical of one-group pretest-posttest studies, alternative explanations such as testing, maturation, history, instructor-related influences, and regression to the mean remain relevant when interpreting the observed change. Consequently, the change documented in Tables 2 and 3 is most appropriately understood as substantial learning progression within this cohort.

This distinction is also important when comparing the present findings with previous studies. Yennita et al. (2025) used a non-equivalent pretest-posttest control-group design in an undergraduate Plant Physiology course and reported differences in a broad meta-skills construct following a research-based project. Because that study used a control group, a different intervention, and a different instrument, its effect estimates are not directly comparable with the present N-gain or dz. Likewise, Komaria et al. (2024) studied STEM-oriented RBL with higher-education students and assessed planning, monitoring, and evaluation in a contextual scientific problem. Differences in participants, instruments, analytical procedures, and instructional context prevent direct numerical comparison with the present learning-outcome results.

Accordingly, the N-gain of .538 and dz of 2.960 are most informative as indicators of learning progression within the design and measurement conditions of the present study, while comparisons with other studies should account for methodological differences.

Post-Intervention Metacognitive-Regulation Profile

The second research question concerned students' metacognitive regulation after the intervention. As shown in Tables 5 and 6, the overall profile was concentrated in the sufficient and good categories, while dimension-specific results showed planning at 81.3% (good), monitoring at 73.2% (sufficient), and evaluating at 69.7% (sufficient). Planning was therefore the strongest of the three measured dimensions. The interpretation of these profiles should also be read alongside the questionnaire validity and reliability evidence summarized in Table 4.

The relatively strong planning score is consistent with the structure of the intervention. Students were explicitly required to identify a research focus, determine what information was needed, identify relevant literature, organize the investigation, and follow a structured worksheet before they began evidence synthesis. Planning was therefore externally scaffolded and made visible through concrete task requirements.

Monitoring and evaluating showed comparatively greater room for further strengthening than planning. These processes required students to judge whether retrieved articles were sufficiently relevant, recognize inconsistencies or missing information, check whether extracted evidence supported emerging claims, reconsider previous decisions, and assess whether conclusions adequately answered the research question. Such regulatory judgments are less procedural than planning and require students to assess the quality of their own understanding and decision making while a task is underway or after it is completed.

This pattern is consistent with previous biology-education research showing that awareness of learning strategies does not necessarily translate into effective monitoring and control (Dye & Stanton, 2017; Stanton et al., 2019). Recent evidence from first-year life-science majors strengthens this interpretation. Stanton et al. (2024) found that students frequently sought guidance when planning but experienced particular difficulty accurately monitoring what they did and did not understand. A similar concern appears in developmental-biology research, where reflective assignments revealed continuing difficulty with higher-order application even when students possessed metacognitive knowledge (Aquino Nunez et al., 2025).

The pattern indicates that planning was the most strongly expressed dimension at the post-intervention measurement point, while longitudinal differences among dimensions remain to be examined in future studies.

Metacognitive Scaffolding and Instructional Implications

The monitoring and evaluating profiles highlight specific opportunities for further strengthening the instructional sequence. Rather than increasing the amount of general reflection, scaffolds can be positioned at the moments when students must make regulatory decisions. During literature screening, students could be required to justify inclusion or exclusion decisions. During evidence extraction, scheduled evidence-audit checkpoints could require students to verify values against source articles and identify contradictions. During synthesis, prompts could ask whether each claim is adequately supported. After presentation, students could use a structured reflection rubric to identify which strategies should be retained or changed.

These recommendations are consistent with earlier research showing that explicit prompts, exam wrappers, planning tools, and guided reflection can make otherwise implicit regulatory processes more visible and actionable (Colthorpe et al., 2019; Stanton et al., 2015; Tanner, 2012). Recent biology-education studies provide further support for targeted rather than generic metacognitive scaffolding. Angell et al. (2024) found that open-ended metacognitive exam-preparation assignments were associated with moderately higher exam scores for some lower-achieving introductory biology students compared with terminology-focused assignments. Ratnayake et al. (2024) similarly reported that structured planning and reflection tools in an undergraduate biology course helped students examine the effectiveness of their study strategies and identify future changes in learning behavior.

This interpretation extends the practical significance of the present findings. The central instructional issue is not simply whether students complete an SLR product, but whether the research sequence makes planning, monitoring, and evaluating sufficiently explicit. Integrated STEM connections should also be made visible rather than assumed. English (2016), Kelley and Knowles (2016), and Roehrig et al. (2021) emphasize the need for coherence among disciplinary components, while recent empirical work by Hiwatig et al. (2024) likewise indicates that the operational features of integrated STEM matter for how students experience an integrated curriculum.

A strengthened version of the present sequence could therefore combine a research-planning sheet before searching, mandatory search and decision logs during screening, evidence-audit checkpoints during data extraction, explicit prompts for disciplinary connections during synthesis, and a structured reflection rubric following peer review.

Integrated Interpretation and Boundaries of Evidence

The findings present a coherent pattern of substantial learning-outcome change and a distinct post-intervention metacognitive-regulation profile across planning, monitoring, and evaluating. Students demonstrated marked progression in learning-outcome scores, while their metacognitive profile showed particularly strong planning alongside opportunities to further strengthen monitoring and evaluating. These two findings provide complementary, but analytically distinct, perspectives on students' engagement with the SLR-oriented STEM-integrated RBL sequence.

The learning-outcome evidence captures change across two measurement points, whereas the metacognitive evidence characterizes students' regulatory profile after completing the intervention. Together, these evidence streams provide a focused account of learning progression and metacognitive regulation within the implemented instructional context. Questions concerning causal attribution, longitudinal metacognitive development, and predictive relationships can be addressed more directly through future comparative and repeated-measures research.

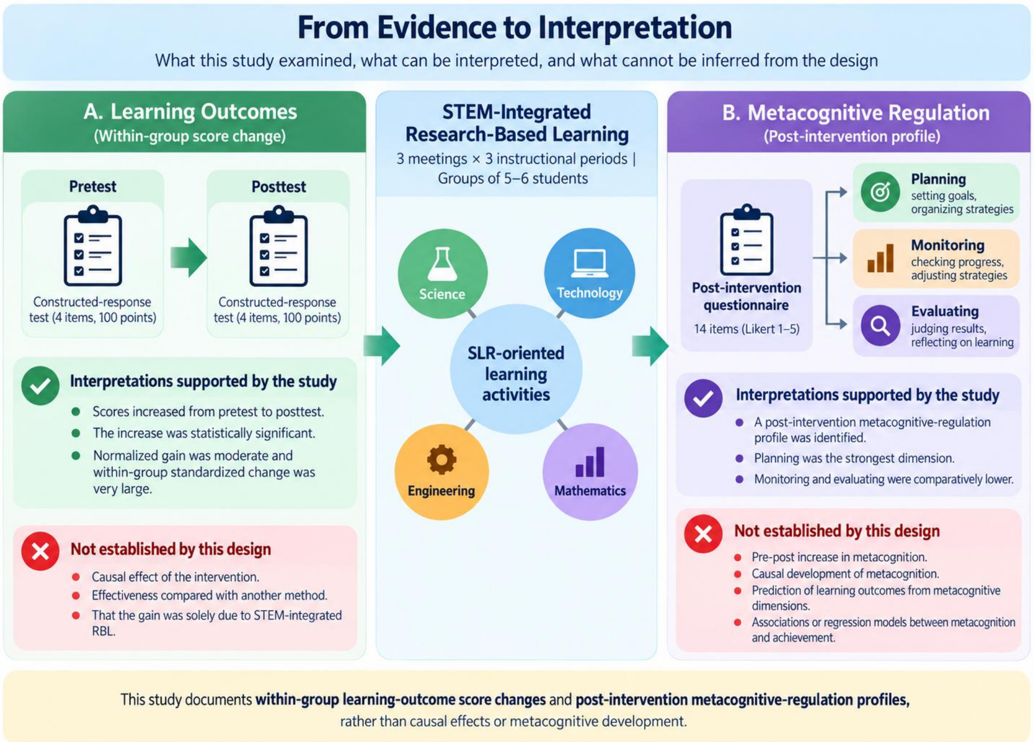

Figure 3. Evidence-to-interpretation framework distinguishing verified observations from causal, developmental, and predictive conclusions that are not supported by the study design.

Figure 3 summarizes these evidence boundaries. Solid pathways represent the observations directly supported by the study records, whereas unsupported causal, developmental, and predictive interpretations are separated from the verified findings. The pretest-posttest evidence supports a statement about within-group learning-outcome change, while the single questionnaire administration supports a statement about the post-intervention metacognitive profile. Neither evidence stream supports a conclusion about causal treatment effects, metacognitive development, or predictors of achievement.

Limitations and Directions for Future Research

Several design and measurement considerations help define the scope and transferability of the present findings. As is typical of a one-group pretest-posttest design, testing, maturation, history, instructor influences, and regression-to-the-mean effects remain relevant considerations. A concurrent comparison group in future research would allow the specific contribution of the instructional sequence to be examined more directly.

The sample came from one Biology Education program and was predominantly female. The findings are therefore most directly applicable to cohorts with similar disciplinary and institutional characteristics, while multisite research would help establish their broader transferability.

A further measurement consideration concerns the four-item constructed-response assessment. Although it was aligned with the course outcomes and supported by content-validity evidence, its α = .530 and the absence of documented inter-rater agreement indicate that future studies could strengthen score precision through a broader assessment and independent scoring procedures with documented inter-rater reliability.

The independence of the 65-response instrument-testing dataset was not documented. For this reason, the reliability coefficients are retained as instrument-testing evidence but should not be interpreted as estimates obtained from a clearly independent validation cohort.

The post-intervention self-report administration provides a clear profile of students' reported regulation at the end of the learning sequence. Future pretest-posttest and trace-based measures would extend this evidence by examining change over time and correspondence with observed regulatory behavior, using approaches such as think-aloud protocols, search logs, reflection records, decision logs, or observation of group activity.

Finally, participant-level raw data were unavailable for verification of additional statistical assumptions or secondary analyses. Accordingly, simulated diagnostic statistics and unsupported correlation and regression claims were removed rather than reconstructed. Future studies should preserve deidentified participant-level datasets and prespecify the intended analytic strategy so that assumption checks, sensitivity analyses, and any secondary relationships can be verified directly.

Controlled or quasi-experimental designs, multisite samples, longer intervention periods, pretest-posttest measures of metacognitive regulation, stronger constructed-response scoring procedures, direct measures of regulatory behavior, and formally documented implementation fidelity would allow more defensible conclusions about whether and how SLR-oriented STEM-integrated RBL contributes to biology learning and metacognitive regulation.

CONCLUSIONS

Students' learning-outcome scores increased substantially after participation in STEM-integrated Research-Based Learning, from a pretest mean of 26.03 to a posttest mean of 65.92. The change was statistically significant, with a moderate N-gain and a very large within-group standardized difference. These findings provide strong descriptive evidence of substantial learning progression within the participating cohort. Comparative designs would be valuable for establishing the specific contribution of STEM-integrated RBL to that progression.

Students' post-intervention metacognitive-regulation profile was good for planning (81.3%) and sufficient for monitoring (73.2%) and evaluating (69.7%). Planning was therefore the strongest metacognitive dimension observed after the intervention. Because metacognitive regulation was assessed after the intervention, these results characterize the regulatory profile achieved at the end of the learning sequence; longitudinal development remains an important direction for subsequent research.

Overall, this study contributes empirical evidence that an SLR-oriented STEM-integrated RBL sequence can provide a productive learning context for substantial learning-outcome progression while engaging undergraduate biology students in planning, monitoring, and evaluating during authentic plant-metabolite inquiry. The findings provide a foundation for future comparative and longitudinal studies to examine the mechanisms and longer-term outcomes of this instructional approach.

RECOMMENDATIONS

Future implementations should provide more explicit and stage-specific metacognitive scaffolding, including research-planning sheets before literature searching, search and decision logs during article screening, progress-check and evidence-verification prompts during data extraction, reflective journals during synthesis, and structured self-evaluation rubrics after presentation and peer feedback.

Future research should use experimental or quasi-experimental designs with appropriate comparison groups, larger and more diverse multisite samples, and longer intervention periods. Metacognitive regulation should be measured before and after the intervention to permit analysis of change over time. Future studies should also use broader and more reliable achievement measures, validated constructed-response scoring procedures with documented inter-rater agreement, and direct or trace-based measures of metacognitive behavior, such as think-aloud protocols, learning logs, search histories, decision records, or structured observations. Implementation fidelity should be documented systematically through attendance records, stage-completion records, observation checklists, and other protocol-adherence indicators. Participant-level deidentified data should also be retained to permit verification of statistical assumptions and any planned secondary analyses.

Acknowledgment

The authors gratefully acknowledge the support of the Biology Education Program (Tadris Biologi) at Universitas Islam Negeri Kiai Haji Achmad Siddiq Jember for facilitating this research. Sentence-level language editing were assisted by ChatGPT. All scientific interpretations, data checks, revisions, and final wording were reviewed and approved by the authors, who remain fully responsible for the manuscript.

Funding Information

This research received no external funding.

Conflict of Interest Statement

The authors declare no conflict of interest.

Ethical Considerations

Participation was voluntary. Students were informed about the study purpose, procedures, confidentiality protections, and their right to decline or withdraw without academic penalty. Written informed consent was obtained before data collection. Participation did not affect course grades, and participant codes rather than names were used in the analytic dataset.

Informed Consent
Informed consent was obtained from all individuals involved in this study.

Data Availability
The data that support the findings of this study are available from the first and corresponding author, Ika Priantari, upon reasonable request.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ika Priantari	✓	✓			✓	✓		✓	✓	✓	✓			
Suratno	✓	✓		✓						✓				
Dwi Wahyuni		✓		✓						✓				
Dafik	✓	✓		✓	✓					✓		✓		
YDWK Ningtyas				✓						✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

REFERENCES

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45(1), 131–142. https://doi.org/10.1177/0013164485451012

Altındağ, M., & Senemoğlu, N. (2013). Metacognitive skills scale. Hacettepe University Journal of Education, 28(1), 15–26. https://dergipark.org.tr/en/pub/hunefd/article/101825

Angell, D. K., Lane-Getaz, S., Okonek, T., & Smith, S. (2024). Metacognitive exam preparation assignments in an introductory biology course improve exam scores for lower ACT students compared with assignments that focus on terms. CBE—Life Sciences Education, 23(1), ar6. https://doi.org/10.1187/cbe.22-10-0212

Aquino Nunez, W., Scott, M. C. B., & Morgan, D. E. (2025). Metacognitive mapping: Developmental biology students struggle with higher-order cognitive skills in multiple course formats. Journal of Microbiology & Biology Education, 26(2), e00139-24. https://doi.org/10.1128/jmbe.00139-24

Brew, A. (2007, April 19–21). Research and teaching from the students' perspective. International Policies and Practices for Academic Enquiry: An International Colloquium, Winchester, United Kingdom. http://portal-live.solent.ac.uk/university/rtconference/2007/resources/angela_brew.pdf

Brown, R., Brown, J., Reardon, K., & Merrill, C. (2011). Understanding STEM: Current perceptions. Technology and Engineering Teacher, 70(6), 5–9. https://eric.ed.gov/?id=EJ918930

Chatzipanteli, A., Grammatikopoulos, V., & Gregoriadis, A. (2014). Development and evaluation of metacognition in early childhood education. Early Child Development and Care, 184(8), 1223–1232. https://doi.org/10.1080/03004430.2013.861456

Chen, C.-H., & Yang, Y.-C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. Educational Research Review, 26, 71–81. https://doi.org/10.1016/j.edurev.2018.11.001

Colthorpe, K., Ogiji, J., Ainscough, L., Zimbardi, K., & Anderson, S. (2019). Effect of metacognitive prompts on undergraduate pharmacy students' self-regulated learning behavior. American Journal of Pharmaceutical Education, 83(4), 6646. https://doi.org/10.5688/ajpe6646

Dafik, Sucianto, B., Irvan, M., & Rohim, M. A. (2019). The analysis of student metacognition skill in solving rainbow connection problem under the implementation of research-based learning model. International Journal of Instruction, 12(4), 593–610. https://doi.org/10.29333/iji.2019.12438a

Daryanes, F., & Sayuti, I. (2023). Research-based learning in biology courses to train students critical thinking skills: Student's perception. Biosfer: Jurnal Pendidikan Biologi, 16(1), 124–137. https://doi.org/10.21009/biosferjpb.23160

Dye, K. M., & Stanton, J. D. (2017). Metacognition in upper-division biology students: Awareness does not always lead to control. CBE—Life Sciences Education, 16(2), ar31. https://doi.org/10.1187/cbe.16-09-0286

English, L. D. (2016). STEM education K–12: Perspectives on integration. International Journal of STEM Education, 3, Article 3. https://doi.org/10.1186/s40594-016-0036-1

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1), 64–74. https://doi.org/10.1119/1.18809

Healey, M., Jenkins, A., & Lea, J. (2014). Developing research-based curricula in college-based higher education. Higher Education Academy. https://eric.ed.gov/?id=ED561351

Helendra, H., Lufri, L., & Elizar, E. (2023). Students' metacognitive skill level of Biology Education Study Program. In Proceedings of the 3rd International Conference on Biology, Science and Education (IcoBioSE 2021) (pp. 76–83). Atlantis Press. https://doi.org/10.2991/978-94-6463-166-1_11

Hindun, I., Nurwidodo, N., & Wicaksono, A. G. C. (2020). Metacognitive awareness components of high-academic ability students in biology hybrid learning: Profile and correlation. JPBI (Jurnal Pendidikan Biologi Indonesia), 6(1), 31–38. https://doi.org/10.22219/jpbi.v6i1.11097

Hiwatig, B. M. R., Roehrig, G. H., & Rouleau, M. D. (2024). Unpacking the nuances: An exploratory multilevel analysis on the operationalization of integrated STEM education and student attitudinal change. Disciplinary and Interdisciplinary Science Education Research, 6(1), Article 18. https://doi.org/10.1186/s43031-024-00108-6

Holmlund, T. D., Lesseig, K., & Slavit, D. (2018). Making sense of "STEM education" in K–12 contexts. International Journal of STEM Education, 5, Article 32. https://doi.org/10.1186/s40594-018-0127-2

Irwanto, Rohaeti, E., Widjajanti, E., & Suyanta. (2017). Students' science process skill and analytical thinking ability in chemistry learning. AIP Conference Proceedings, 1868, 030001. https://doi.org/10.1063/1.4995100

Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., Gürbüz, F., & Bonyah, E. (2024). Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills. REID (Research and Evaluation in Education), 10(1), 64–79. https://doi.org/10.21831/reid.v10i1.71032

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. International Journal of STEM Education, 3, Article 11. https://doi.org/10.1186/s40594-016-0046-z

Komaria, N., Suratno, Sudarti, & Dafik. (2024). Analysis of the implementation of research-based learning material with STEM approach in improving students' metacognitive skill. International Journal of Instruction, 17(2), 199–218. https://doi.org/10.29333/iji.2024.17212a

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, 863. https://doi.org/10.3389/fpsyg.2013.00863

Marrero, M. E., Gunning, A. M., & Germain-Williams, T. (2014). What is STEM education? Global Education Review, 1(4), 1–6. https://ger.mercy.edu/index.php/ger/article/view/135

Marsden, E. J., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. Oxford Review of Education, 38(5), 583–616. https://doi.org/10.1080/03054985.2012.731208

Monalisa, L. A., Dafik, Hastuti, Y., Hussen, S., & Oktavianingtyas, E. (2019). The implementation of research based learning in developing the students mathematical generalization thinking skills in solving a paving blocks design problem. IOP Conference Series: Earth and Environmental Science, 243, 012168. https://doi.org/10.1088/1755-1315/243/1/012168

Quigley, A., Muijs, D., & Stringer, E. (2018). Metacognition and self-regulated learning: Guidance report. Education Endowment Foundation. https://dera.ioe.ac.uk/id/eprint/31617/

Ratnayake, A., Bansal, A., Wong, N., Saseetharan, T., Prompiengchai, S., Jenne, A., Thiagavel, J., & Ashok, A. (2024). All "wrapped" up in reflection: Supporting metacognitive awareness to promote students' self-regulated learning. Journal of Microbiology & Biology Education, 25(1), e00103-23. https://doi.org/10.1128/jmbe.00103-23

Roehrig, G. H., Dare, E. A., Ring-Whalen, E., & Wieselmann, J. R. (2021). Understanding coherence and integration in integrated STEM curriculum. International Journal of STEM Education, 8, Article 2. https://doi.org/10.1186/s40594-020-00259-8

Sadykova, A., Iskakova, M., Ismailova, G., Ishmukhametova, A., Sovetova, A., & Mukasheva, K. (2024). The impact of a metacognition-based course on school students' metacognitive skills and biology comprehension. Frontiers in Education, 9, 1460496. https://doi.org/10.3389/feduc.2024.1460496

Schraw, G. (1998). Promoting general metacognitive awareness. Instructional Science, 26(1–2), 113–125. https://doi.org/10.1023/A:1003044231033

Schraw, G., & Moshman, D. (1995). Metacognitive theories. Educational Psychology Review, 7(4), 351–371. https://doi.org/10.1007/BF02212307

Sota, C., & Peltzer, K. (2017). The effectiveness of research based learning among Master Degree student for health promotion and preventable disease, Faculty of Public Health, Khon Kaen University, Thailand. Procedia - Social and Behavioral Sciences, 237, 1359–1365. https://doi.org/10.1016/j.sbspro.2017.02.226

Stanton, J. D., Dye, K. M., & Johnson, M. S. (2019). Knowledge of learning makes a difference: A comparison of metacognition in introductory and senior-level biology students. CBE—Life Sciences Education, 18(2), ar24. https://doi.org/10.1187/cbe.18-12-0239

Stanton, J. D., Halmo, S. M., Carter, R. J., Yamini, K. A., & Ososanya, D. (2024). Opportunities for guiding development: Insights from first-year life science majors' use of metacognition. Journal of Microbiology & Biology Education, 25(3), e00053-24. https://doi.org/10.1128/jmbe.00053-24

Stanton, J. D., Neider, X. N., Gallegos, I. J., & Clark, N. C. (2015). Differences in metacognitive regulation in introductory biology students: When prompts are not enough. CBE—Life Sciences Education, 14(2), ar15. https://doi.org/10.1187/cbe.14-08-0135

Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. CBE—Life Sciences Education, 20(2), fe3. https://doi.org/10.1187/cbe.20-12-0289

Suntusia, Dafik, & Hobri. (2019). The effectiveness of research based learning in improving students' achievement in solving two-dimensional arithmetic sequence problems. International Journal of Instruction, 12(1), 17–32. https://doi.org/10.29333/iji.2019.1212a

Supit, P. G. Y., & Winardi, Y. (2024). Pembelajaran berbasis riset (Research-Based Learning) untuk meningkatkan kemampuan berpikir kritis, berpikir kreatif dan berpikir reflektif siswa dalam pembelajaran biologi. Polyglot: Jurnal Ilmiah, 20(2), 115–135. https://doi.org/10.19166/pji.v20i2.8355

Syofyan, H., MS, Z., & Sumantri, M. S. (2019). Use of integrated thematic teaching materials based on problem solving in natural science learning in elementary schools. In Proceedings of the First International Conference on Technology and Educational Science (ICTES 2018). EAI. https://doi.org/10.4108/eai.21-11-2018.2282034

Taleb, H. M., & Chadwick, C. (2016). Enhancing student critical and analytical thinking skills at a higher education level in developing countries: Case study of the British University in Dubai. Journal of Educational and Instructional Studies in the World, 6(1), 67–77. https://www.arastirmax.com/en/system/files/dergiler/116392/makaleler/6/1/arastirmax-enhancing-student-critical-and-analytical-thinking-skills-higher-education-level-developing-countries-case-study-british-university-dubai.pdf

Tanner, K. D. (2012). Promoting student metacognition. CBE—Life Sciences Education, 11(2), 113–120. https://doi.org/10.1187/cbe.12-03-0033

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., Lowe, A., Newman, S., Okolo, V., Olroyd, S., Peecook, B. R., Pickett, S. B., Slager, D. L., Caviedes-Solis, I. W., Stanchak, K. E., Sundaravardan, V., Valdebenito, C., Williams, C. R., Zinsli, K., & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. Proceedings of the National Academy of Sciences, 117(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., Boeve-de Pauw, J., Dehaene, W., Deprez, J., De Cock, M., Hellinckx, L., Knipprath, H., Langie, G., Struyven, K., Van de Velde, D., Van Petegem, P., & Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. European Journal of STEM Education, 3(1), Article 2. https://doi.org/10.20897/ejsteme/85525

White, D. W. (2014). What is STEM education and why is it important? Florida Association of Teacher Educators Journal, 1(14), 1–9. http://www.fate1.org/journals/2014/white.pdf

Winarni, J., Zubaidah, S., & Koes, S. (2016). STEM: Apa, mengapa, dan bagaimana. In Prosiding Seminar Nasional Pendidikan IPA Pascasarjana UM (Vol. 1, pp. 976–984). https://pasca.um.ac.id/wp-content/uploads/2017/02/Juniaty-Winarni-976-984.pdf

Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. Annual Review of Sociology, 41, 331–357. https://doi.org/10.1146/annurev-soc-071312-145659

Yennita, Y., Astriawati, F., Jumiarni, D., & Irawati, S. (2025). The impact of Research-Based Projects (RBP) in Plant Physiology course on undergraduate student meta-skills. Bioedusiana: Jurnal Pendidikan Biologi, 10(1), 34–45. https://doi.org/10.37058/bioed.v10i1.12922

Zohar, A., & Barzilai, S. (2015). Metacognition and teaching higher order thinking (HOT) in science education: Students' learning, teachers' knowledge and instructional practices. In R. Wegerif, L. Li, & J. C. Kaufman (Eds.), The Routledge international handbook of research on teaching thinking (pp. 229–242). Routledge. https://doi.org/10.4324/9781315797021

Zou, T. X. P., Lee, J. C. Y., Yu, K. S., Chow, K. L., Barry, T. J., Leung, L. Y. Y., & Brew, A. (2024). Faculty members' perceptions and students' experiences of research-based curricula: A multiple case study of four undergraduate programmes. Higher Education, 88(3), 1205–1225. https://doi.org/10.1007/s10734-023-01166-x

APPENDIX A

Learning-Outcome Assessment Blueprint and Scoring Structure

The scored pretest-posttest assessment comprised four constructed-response items. A fifth article-product activity in the instructional blueprint was part of the project and was not included in the 0–100 cognitive score.

Table A1. Learning-Outcome Assessment Blueprint

Item	Learning indicator	Representative task	Maximum score
1	Analyze the roles of plants for humans (C4)	Identify and explain plant roles, scientific/common name, part used, field of use, active/metabolite compounds, and application.	20
2	Analyze primary and secondary metabolites functioning as antioxidants (C4)	Identify plants and relevant primary/secondary metabolites with antioxidant roles.	20
3	Analyze primary and secondary metabolites in coffee (C4)	Identify two primary and two secondary metabolites in coffee and explain antioxidant mechanisms.	20
4	Draw an evidence-based conclusion from a coffee-metabolite abstract (C5)	Conclude how coffee type and processing relate to caffeine, chlorogenic acid, and antioxidant properties.	40

The original analytic scoring rubric allocated increasing scores according to the completeness and accuracy of each response. The total scored range for Items 1–4 was 0–100.

APPENDIX B

Representative STEM-RBL Worksheet Structure

Table B1. Seven-Stage STEM-RBL Worksheet

Stage	RBL activity	Representative worksheet requirement	Metacognitive emphasis
1	Formulate a research problem/question	Coffee variety, roasting process/equipment, temperature/time, caffeine, chlorogenic acid	Planning
2	Plan the literature-based investigation	Identify relevant full-text indexed articles and organize search information	Planning
3	Collect and screen literature	Apply quality/relevance questions to candidate articles	Monitoring
4	Compile evidence	Tabulate findings on roasting conditions and metabolite outcomes	Monitoring
5	Develop discussion	Compare and synthesize evidence across articles	Monitoring
6	Prepare conclusion/article	Align conclusions with the research question and article template	Monitoring and evaluating
7	Present, receive feedback, and reflect	Present article, identify strengths/weaknesses, reflect on benefits/challenges	Evaluating

APPENDIX C

Metacognitive-Regulation Questionnaire

Table C1. Metacognitive-Regulation Questionnaire

Dimension	Item	Statement
Planning	1	When I first received the article-project task, I understood the main problem to be investigated.
Planning	2	I had a structured strategy or plan for completing the article project.
Planning	3	I had a specific strategy for connecting Plant Physiology concepts with roasting processes that affect secondary metabolites.
Planning	4	I understood the basic concepts required to complete the article project.
Monitoring	5	I monitored whether my ideas or solutions were consistent with the research question.
Monitoring	6	I had several strategies for overcoming difficulties when completing the article project.
Monitoring	7	I attempted to complete the article project according to the research question and article template by following the specified stages.
Monitoring	8	I used strategies to verify whether each stage of the project was appropriate.
Monitoring	9	I evaluated or reflected on the completed article project.
Evaluating	10	I could identify parts of the project article that were not yet optimal.
Evaluating	11	I considered several alternative solutions before selecting the best option.
Evaluating	12	I used strategies I had mastered to address similar new problems.
Evaluating	13	I evaluated or reflected on the strategies used to complete the article project.
Evaluating	14	I evaluated or reflected on whether the article project was consistent with the research question or objective established at the beginning.

Response options: 1 = strongly disagree, 2 = disagree, 3 = neutral/uncertain, 4 = agree, and 5 = strongly agree. All statements were positively worded. Percentage scores were categorized as very good (86–100), good (76–85), sufficient (60–75), low (55–59), or very low (0–54).