INTERNATIONAL JOURNAL OF ENVIRONMENT, ENGINEERING & EDUCATION

ISSN 2656-8039 (ONLINE)
VOLUME 8 • ISSUE 3 • SEPTEMBER 2026

Received: 23 June 2026 | Accepted: 23 September 2026 | Published: 28 September 2026

Research Article

Mathematical Critical Thinking in Trigonometry Under E-Module Supported Project-Based Learning: An Explanatory Sequential Study Across Self-Reported Modality Preferences

Epih Purnamasari • Stevanus Budi Waluya * • Arief Agoestanto • Emi Pujiastuti

Department of Mathematics Education, Semarang State University, Indonesia

ABSTRACT

Mathematical critical thinking (MCT) involves interpreting information, analyzing relationships, evaluating procedures, inferring conclusions, explaining reasoning, and regulating problem solving. This study examined pretest-posttest changes in MCT during e-module-supported Project-Based Learning (PjBL) in trigonometry and explored variation across self-reported Visual, Auditory, and Kinaesthetic (VAK) preferences. A quantitatively driven explanatory sequential mixed-methods design was used with Grade XI students at a public senior high school in Cirebon Regency, West Java, Indonesia. From 180 students screened, 72 were included in the quantitative sample ($n = 24$ per VAK group). Students completed an MCT pretest and posttest around six 90-minute PjBL sessions. Qualitative follow-up used interviews with 12 purposively selected students, classroom observations, and student artifacts. Mean MCT scores increased from 43.95 ($SD = 6.26$) to 78.18 ($SD = 5.77$); the mean within-student change was 34.23 points, 95% CI [32.26, 36.20], $t(71) = 34.66$, $p < 0.001$, Cohen's $d_z = 4.08$. All six MCT dimension indicators showed statistically significant pretest-posttest changes after false-discovery-rate adjustment. No statistically significant differences were detected among VAK groups for posttest MCT or normalized gain, and ANCOVA controlling for pretest performance yielded the same inferential conclusion. Qualitative analysis identified overlapping patterns of visualization, verbalization, digital manipulation, trial-and-check reasoning, peer discussion, and reflection. The integrated findings document marked within-student change alongside heterogeneous reasoning processes, but the single-group design precludes causal attribution of the observed changes to PjBL or the e-module.

KEYWORDS constructed-response assessment • digital scaffolding • GeoGebra • higher-order thinking skills • mixed-methods research • mathematical reasoning • VARK framework

ARTICLE CITATION

E. Purnamasari, S. B. Waluya, A. Agoestanto, E. Pujiastuti, "Mathematical Critical Thinking in Trigonometry Under E-Module Supported Project-Based Learning: An Explanatory Sequential Study Across Self-Reported Modality Preferences," International Journal of Environment, Engineering and Education, Vol. 8, No. 3, pp. 500-524, 2026. https://doi.org/10.55151/ijeedu.v8i3.466

*CORRESPONDENCE

Stevanus Budi Waluya budiw@mail.unnes.ac.id Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, Semarang State University, Jl. Raya Sekaran, Gunungpati, Semarang 50229, Central Java, Indonesia https://orcid.org/0000-0002-8834-1138

1. INTRODUCTION

Critical thinking is a central objective of mathematics education because mathematical proficiency requires more than procedural accuracy. Students must interpret information, identify relevant relationships, evaluate strategies, justify conclusions, and monitor the coherence of their own reasoning. Meta-analytic evidence indicates that critical-thinking instruction is most effective when learners engage explicitly with authentic problems, dialogue, and guided reflection rather than encounter critical thinking only as an implicit by-product of subject teaching [1].

In mathematics, these demands are particularly consequential because reasoning must coordinate symbolic, graphical, numerical, and contextual representations. Recent research likewise conceptualizes critical thinking in mathematics as a domain-embedded set of reasoning practices rather than a generic skill detached from mathematical content [2], [3]. The Indonesian context gives this problem additional urgency. In PISA 2022, only 18% of Indonesian students reached at least Level 2 proficiency in mathematics, compared with an OECD average of 69% [4]. Although PISA does not directly measure mathematical critical thinking (MCT), the result indicates persistent difficulty with mathematical representation, reasoning, and non-routine problem solving.

MCT can be understood as the disciplined application of critical-thinking processes within mathematical activity. The present study operationalizes MCT through six dimensions: interpretation, analysis, evaluation, inference, explanation, and self-regulation, consistent with Facione's framework and recent studies [3], [5]. Mathematical performance involves more than correct answers; it requires interpreting information, analyzing relationships, evaluating procedures, inferring conclusions, explaining reasoning, and monitoring errors.

This multidimensional view reflects evidence that mathematical critical thinking integrates reasoning, judgment, justification, and reflective regulation rather than representing a single undifferentiated ability [2]. Recent evidence on mathematical literacy also distinguishes critical and reflective thinking as related but non-identical processes, reinforcing the value of examining variation within higher-order mathematical reasoning rather than collapsing it into an overall achievement measure [6]. Such distinctions are methodologically important because equivalent test scores may conceal different reasoning processes, levels of justification, or patterns of self-monitoring.

Trigonometry provides a particularly demanding setting in which these processes can be observed. Understanding trigonometric relationships requires students to coordinate triangle-based ratios, unit-circle interpretations, algebraic transformations, functional representations, and graphical behavior. Research has documented persistent difficulties in moving between triangle and circle representations, understanding periodicity and radian measure, and connecting conceptual knowledge with procedural operations [7].

Experimental work further shows that trigonometric problem solving becomes more difficult when algebraic transformations increase in complexity and that instructional guidance can materially influence conceptual and procedural performance [8]. Contextualized instruction has also been associated with stronger conceptual understanding of trigonometry, suggesting that students benefit when abstract relationships are connected to interpretable mathematical situations [9]. Trigonometry is therefore not merely a technically difficult topic; it is a domain in which interpretation, strategy selection, verification, explanation, and self-regulation are repeatedly required.

Project-Based Learning (PjBL) offers an instructional structure capable of eliciting these forms of reasoning when projects are organized around mathematically substantive inquiry. PjBL typically requires learners to investigate meaningful problems, make decisions, collaborate, construct products or solutions, communicate results, and reflect on the process [10]. These activities align conceptually with MCT: problem formulation requires interpretation, investigation requires analysis and inference, solution construction requires evaluation, presentation requires explanation, and reflection creates conditions for self-regulation.

Meta-analytic evidence has generally found positive effects of PjBL on academic achievement, although the magnitude of effects varies across contexts and implementation features [11]. A recent umbrella review reached a similarly qualified conclusion: PjBL effects are directionally positive across academic achievement, higher-order thinking, and 21st-century skills, but the methodological quality of the underlying meta-analyses is uneven [12]. PjBL should therefore be treated as an instructional environment that can create opportunities for critical reasoning, not as a mechanism that automatically produces MCT improvement.

Digital resources may strengthen those opportunities when their functions are aligned with the cognitive demands of project work. Research syntheses on technology-enhanced mathematics instruction show positive but heterogeneous effects, with outcomes depending on how technology is used rather than on technology exposure alone [13], [14]. This distinction is central to the role of the e-module in the present study. Interactive tasks, digital worksheets, formative feedback, and dynamic representations can provide scaffolds for exploring mathematical relationships, checking conjectures, and revising solutions.

Evidence from GeoGebra-supported mathematics instruction indicates that dynamic representations can improve conceptual understanding and retention [15]. A recent systematic review further suggests that GeoGebra can support metacognitive monitoring and iterative adjustment by making mathematical relationships dynamically observable and by providing feedback during problem solving [16]. At the same time, digital environments can impose unnecessary cognitive demands when representations and task requirements are poorly coordinated [17]. The e-module is therefore conceptualized as a digital cognitive scaffold embedded within PjBL rather than as an independent treatment assumed to improve MCT simply because it is interactive.

Learner variability requires similar conceptual caution. Students may report preferences for visual, auditory, or kinaesthetic modes of engagement, but such preferences do not establish stable cognitive types or justify matching instruction to a presumed learning style. The meshing hypothesis predicts that learning should improve when instructional modality corresponds to an identified sensory preference; rigorous reviews have not found replicable evidence for the required interaction between learner preference and instructional modality [18], [19].

Historical analysis further shows that visual-auditory-kinaesthetic classifications developed through heterogeneous educational and psychological traditions rather than from a unified cognitive theory [20]. In this study, VAK is therefore used strictly as a self-reported modality-preference variable. The categories describe reported tendencies in engagement and are not used to prescribe matched instruction, infer fixed cognitive capacities, or imply superiority of one group over another.

Four related gaps motivate the study. First, research on critical thinking in mathematics often emphasizes aggregate achievement or total critical-thinking outcomes, although MCT is multidimensional and may be expressed differently during problem solving [2]. Second, PjBL research more often establishes whether outcomes differ across instructional conditions than how specific MCT processes are enacted during project activity. This distinction matters because a change in an overall score does not identify whether students changed primarily in interpreting information, evaluating procedures, explaining reasoning, or regulating their own problem solving.

Third, research on educational technology shows that digital tools can support mathematical learning, but their contribution depends on instructional function. This leaves a need to examine how an e-module operates as a scaffold within a project sequence rather than as a separate technological treatment [14], [16]. Fourth, although learning-styles research provides little support for modality matching, self-reported preferences may retain descriptive value for examining whether similar performance outcomes coexist with different forms of engagement. The unresolved issue is therefore not only whether MCT performance changes during e-module-supported PjBL, but also how its constituent processes are expressed within a common instructional environment.

The problem also has a methodological dimension. Quantitative pretest-posttest analysis can estimate within-student change and compare outcomes across preference groups, but it cannot by itself explain the reasoning processes associated with those patterns. Interviews, observations, and student work can provide complementary evidence about how students construct strategies, coordinate representations, verify results, justify conclusions, and regulate their thinking. Mixed-methods integration is most informative when these strands are connected at the design, analysis, and interpretation stages rather than presented as parallel datasets [21].

The explanatory sequential design therefore uses qualitative evidence to explain, complement, or qualify quantitative patterns across the six MCT dimensions. The study examines MCT simultaneously as a multidimensional performance outcome and as an enacted reasoning process, while treating VAK as a descriptive preference variable rather than a causal mechanism.

Because the quantitative strand uses a single-group pretest-posttest structure, changes observed during the instructional period cannot be attributed independently to PjBL, the e-module, GeoGebra, or VAK preference. The analyses therefore estimate within-student differences observed during participation in the instructional environment rather than definitive intervention effects. This inferential boundary is important because PjBL research frequently relies on quasi-experimental or other non-randomized designs, which constrain causal attribution [10]. Against this background, the study addresses four research questions:

- RQ1: To what extent does students' mathematical critical-thinking performance change from pretest to posttest after participation in the e-module-assisted Project-Based Learning environment?
- RQ2: Are there significant differences in posttest mathematical critical-thinking performance among students with Visual, Auditory, and Kinesthetic self-reported modality preferences after participation in the same instructional environment?
- RQ3: How do students with different self-reported modality preferences enact interpretation, analysis, evaluation, inference, explanation, and self-regulation during e-module-assisted Project-Based Learning?
- RQ4: How do the qualitative findings explain, complement, or contextualize the quantitative findings concerning changes in MCT performance and differences across modality-preference groups?

2. LITERATURE REVIEW

2.1. Mathematical Critical Thinking (MCT)

Mathematical critical thinking (MCT) refers to purposeful, reflective, and self-regulatory judgment enacted in mathematical activity. It extends beyond obtaining correct answers or reproducing algorithms because students must interpret information, identify mathematical relationships, evaluate procedures and claims, draw warranted conclusions, explain reasoning, and monitor their own thinking. This conception aligns with the broader critical-thinking framework of interpretation, analysis, evaluation, inference, explanation, and self-regulation [22], while recognizing that mathematical thinking is discipline-specific and constrained by definitions, representations, logical relations, and standards of justification. Accordingly, MCT concerns the quality of mathematical reasoning rather than procedural accuracy alone [2].

This distinction is important because successful computation does not necessarily constitute critical reasoning. Mathematical competence also involves strategic control, metacognition, and sense-making [23]. Similarly, Lithner [24] distinguished imitative reasoning from mathematically grounded reasoning in which arguments are constructed and justified through relevant mathematical properties. Thus, MCT is demonstrated when students question assumptions, compare strategies, coordinate representations, evaluate transformations, and justify conclusions. Systematic evidence further suggests that critical thinking in mathematics should be defined through explicit cognitive objectives, criteria, and standards rather than treated as an undifferentiated higher-order skill [3].

In this study, MCT is operationalized through six interrelated dimensions: interpretation, analysis, evaluation, inference, explanation, and self-regulation. These dimensions are analytically distinguishable but cognitively recursive rather than strictly sequential. Evaluation may prompt reinterpretation, explanation may expose weaknesses in analysis, and self-regulation may initiate revision at any stage. Self-regulation therefore functions as a cross-cutting mechanism for monitoring and refining mathematical reasoning.

MCT is particularly evident in non-routine, open-ended, and representation-rich tasks that require students to select relevant information, compare strategies, evaluate evidence, and justify conclusions [25], [26]. Such demands are especially salient in trigonometry, where students must coordinate symbolic identities, algebraic transformations, graphical and geometric representations, function values, and domain restrictions [8].

Taken together, MCT is best understood as an interconnected and recursive reasoning system rather than a collection of isolated cognitive skills. Figure 1 synthesizes these relationships by positioning the six dimensions as mutually interacting processes supported by continuous self-regulation [22]–[24].

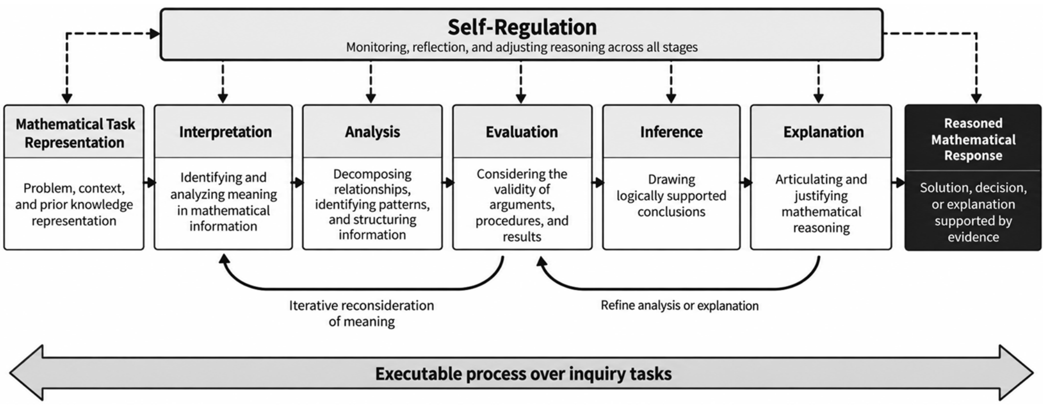


Figure 1. Conceptual architecture of mathematical critical thinking as a recursive and self-regulated reasoning process. Developed synthesis based on Facione [22], Schoenfeld [23], Lithner [24], and Monteleone et al. [2].

2.2. Project-Based Learning and Mathematical Critical Thinking

Project-Based Learning (PjBL) is a student-centered approach in which learning is organized around meaningful problems, sustained inquiry, collaborative investigation, product development, presentation, and reflection. Rather than positioning projects as supplementary activities, PjBL embeds conceptual learning within the process of investigating and responding to authentic or complex problems [10], [27]. This structure is particularly relevant to mathematics because students must move beyond procedural reproduction to determine what information is relevant, select mathematical representations and strategies, justify decisions, and communicate defensible conclusions.

The potential contribution of PjBL to mathematical critical thinking lies primarily in its cognitive demands rather than in the project format itself. Problem orientation requires students to interpret contexts and identify relevant mathematical information; planning and investigation involve analyzing relationships, generating strategies, and examining alternatives; product development requires testing and evaluating the validity of mathematical solutions; presentation makes explanation and justification explicit; and reflection encourages students to evaluate and revise their reasoning. These activities correspond closely to the MCT dimensions of interpretation, analysis, evaluation, inference, explanation, and self-regulation. Evidence from authentic mathematics learning similarly indicates that tasks requiring investigation, argumentation, decision-making, and reflection can provide productive conditions for critical-thinking processes [25].

However, PjBL should not be assumed to activate critical thinking automatically. A project may remain procedurally or cognitively superficial when students focus primarily on completing a product without sustained mathematical reasoning. Reviews of PjBL emphasize that productive learning depends on the quality of the driving problem, structured inquiry, teacher guidance, collaboration, reflection, and appropriate scaffolding [10]. Meta-analytic evidence also suggests that PjBL outcomes vary according to instructional and contextual characteristics, including subject domain, instructional duration, and technological support [11].

PjBL is consequently conceptualized here as an instructional structure that creates recurring opportunities for MCT to be enacted, not as a direct or deterministic cause of higher MCT performance. Its theoretical role is to generate cognitively demanding mathematical activity in which students repeatedly interpret problems, analyze relationships, evaluate strategies, formulate conclusions, justify reasoning, and reflect on their decisions. The e-module, discussed in the next section, is positioned as a digital scaffold for these activities within the PjBL cycle.

2.3. E-Modules as Cognitive Scaffolds

E-modules are structured digital learning resources that integrate content, multiple representations, interactive activities, feedback, and learner-controlled navigation within a coherent instructional environment. Their educational value therefore lies not simply in digitizing printed materials but in how their affordances organize and support cognitive activity. From the perspective of multimedia learning, meaningful learning requires learners to select, organize, and integrate verbal and visual information while operating within limited cognitive-processing capacity [28]. Consequently, effective e-module design should support relevant processing while minimizing unnecessary cognitive demands through principles such as coherence, signaling, and appropriate integration of representations [29]–[31].

For mathematical learning, these affordances are particularly relevant because students frequently need to coordinate symbolic, graphical, numerical, and contextual representations. Interactive multimedia environments can support this process through guided activity, learner control, feedback, and reflection rather than through multimedia presentation alone [32]. Thus, dynamic representations and interactive tasks within an e-module can function as cognitive scaffolds by enabling students to inspect relationships, manipulate parameters, test conjectures, compare representations, and reconsider initial solutions. In the present study, these functions are operationalized through instructional videos, GeoGebra-based simulations, digital worksheets, interactive exercises, and project activities embedded within the e-module.

Feedback constitutes another important mechanism. In digital environments, feedback can move learning beyond answer verification by providing information that prompts learners to diagnose errors, reconsider strategies, and revise reasoning. Meta-analytic evidence indicates that elaborated feedback in computer-based environments is more strongly associated with higher-order learning outcomes than feedback limited to correctness information [33]. Such feedback is theoretically relevant to MCT because evaluation and self-regulation require students to examine the validity of their procedures, identify weaknesses, and modify their reasoning when necessary.

Table 1. Alignment of E-module affordances with MCT processes.

E-module affordance	Cognitive function	Primary MCT processes
Multimedia representations	Integrating symbolic, visual, and contextual information	Interpretation, analysis
GeoGebra/dynamic simulations	Exploring relationships and testing conjectures	Analysis, inference, evaluation
Digital worksheets/prompts	Structuring inquiry and strategy development	Interpretation, analysis, self-regulation
Interactive exercises	Testing and revising solutions	Evaluation, inference
Elaborated formative feedback	Detecting errors and reconsidering strategies	Evaluation, self-regulation
Reflection/self-assessment	Monitoring and revising reasoning	Evaluation, self-regulation
Digital portfolio	Articulating and defending mathematical reasoning	Explanation, evaluation

E-modules may also support self-regulated learning by allowing students to revisit explanations, monitor progress, respond to formative prompts, and regulate the pace and sequence of learning. Research on hypermedia environments shows that digital resources are most productive when learners are supported in planning, monitoring, and regulating their learning rather than being left to navigate complex information without guidance [34], [35]. Accordingly, reflection prompts and self-assessment features can make monitoring and revision explicit, creating conditions in which self-regulation becomes integrated with mathematical problem solving.

These theoretical mechanisms do not imply that digital technology itself improves MCT. Learning outcomes depend more strongly on instructional design and the cognitive activity elicited than on the delivery medium alone (Clark, 2021). In the present framework, the e-module is therefore positioned as a digital scaffold embedded within PjBL: PjBL supplies the problems, inquiry, decision making, and justification demands, whereas the e-module provides representations, guidance, interaction, feedback, and opportunities for monitoring and revision. Its proposed contribution is process-oriented and conditional rather than an independent causal effect.

2.4. E-Integrative Conceptual Framework

The preceding literature suggests that MCT, PjBL, e-modules, and learning preferences should be understood as components of an integrated instructional–cognitive framework rather than as independent variables operating in isolation. In this framework, MCT constitutes the core cognitive outcome, represented by interpretation, analysis, evaluation, inference, explanation, and self-regulation [2], [22]. PjBL provides the instructional structure through which students encounter complex problems, conduct inquiry, make decisions, justify mathematical choices, and reflect on their reasoning [10], [27]. The e-module, in turn, functions as a cognitive scaffold through multiple representations, interactive exploration, formative feedback, and opportunities for monitoring and revision [36]–[38].

The relationship among these components is therefore functional rather than additive. PjBL generates the cognitive demands that require mathematical reasoning, whereas the e-module provides digital affordances that support students in responding to those demands. For example, problem orientation may stimulate interpretation, investigation may activate analysis and inference, product development may require evaluation and revision, and presentation may make explanation and justification explicit. Interactive representations, feedback, and reflective prompts can further support these processes by enabling students to test conjectures, identify inconsistencies, and reconsider strategies. Thus, the proposed mechanism is not that PjBL or digital technology automatically produces stronger MCT, but that their coordinated design creates repeated opportunities for critical mathematical reasoning to be enacted.

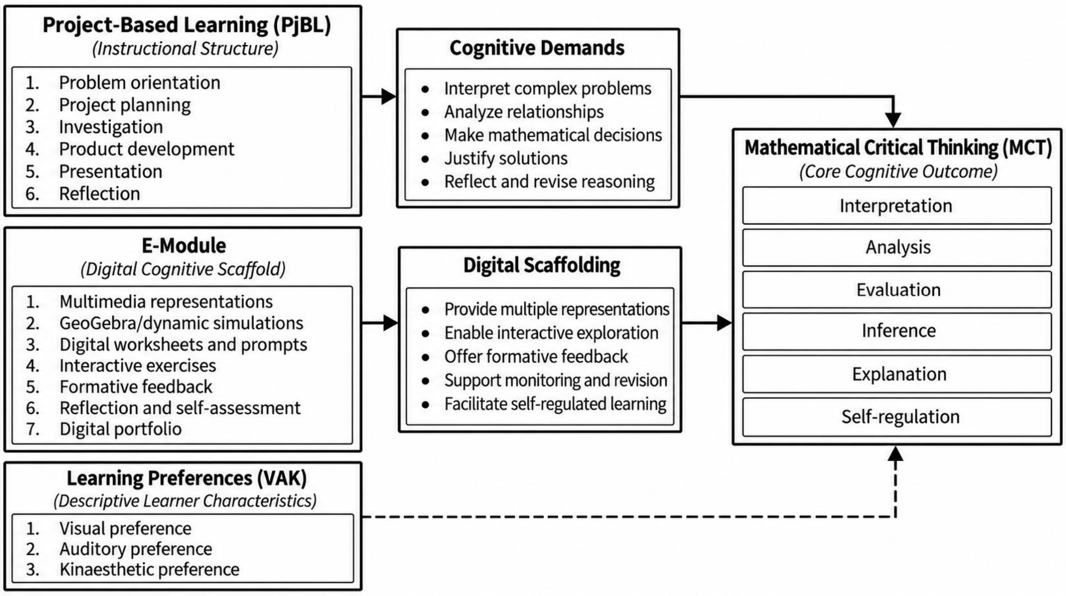


Figure 2. Integrative conceptual framework.

Learning preferences occupy a different conceptual position. They are treated as descriptive learner characteristics rather than determinants of cognitive ability or prescriptions for instructional matching [18], [39]. Students with different self-reported preferences may emphasize different modes of engagement—such as visualization, verbalization, or interactive manipulation—while participating in the same learning environment. These variations may influence how MCT processes become observable without necessarily producing differences in overall MCT performance.

The conceptual framework assumes neither a simple linear causal chain nor preference-specific instructional effects. Instead, it proposes that cognitively demanding project activities, supported by appropriately designed digital scaffolds, create conditions in which MCT processes can be repeatedly enacted, monitored, and refined. Learning preferences are examined as possible sources of variation in how reasoning is expressed rather than as explanations of superior or inferior mathematical performance. This framework provides the basis for integrating quantitative evidence of MCT performance with qualitative evidence of reasoning processes within the same PjBL-e-module environment.

3. MATERIALS AND METHODS

3.1. Research Design

This study employed a quantitatively driven explanatory sequential mixed-methods design with embedded qualitative process data [40], [41]. The quantitative strand used a single-group pretest-posttest design to examine changes in mathematical critical-thinking (MCT) performance and differences across self-reported Visual, Auditory, and Kinaesthetic (VAK) preference groups. Students completed the MCT pretest and VARK questionnaire before six sessions of e-module-supported Project-Based Learning (PjBL), followed by the MCT posttest.

Structured classroom observations and student artifacts provided process evidence during the instructional period, whereas semi-structured interviews were conducted after the quantitative analysis with purposively selected students representing different VAK preferences and performance patterns. Quantitative findings informed the interview sampling and follow-up inquiry; the strands were then compared to identify convergence, complementarity, qualification, or divergence [21]. Because no control group was included, pretest-posttest differences were interpreted as observed changes during the instructional period rather than definitive causal effects of PjBL or the e-module [42].

3.2. Research Setting and Participants

The study was conducted at SMAN 1 Dukupuntang, a public senior high school in Cirebon Regency, West Java, Indonesia, during the first semester of the 2025-2026 academic year. The study cohort comprised 180 Grade XI students enrolled across six classes. All students initially completed the learning-preference questionnaire as part of the screening procedure used to establish eligibility for the Visual, Auditory, and Kinaesthetic (VAK) analytical groups.

Screening followed predefined eligibility criteria. Students were excluded when questionnaire responses were incomplete or invalid, when profiles were multimodal, when two or more preference dimensions produced tied dominant scores, or when Read/Write was the dominant preference. After these exclusions, 104 students constituted the eligible VAK pool: 38 Visual (36.5%), 31 Auditory (29.8%), and 35 Kinaesthetic (33.7%). Relative to the initial cohort of 180 students, these groups represented 21.1%, 17.2%, and 19.4%, respectively. The complete screening and selection pathway is presented in Figure 3.

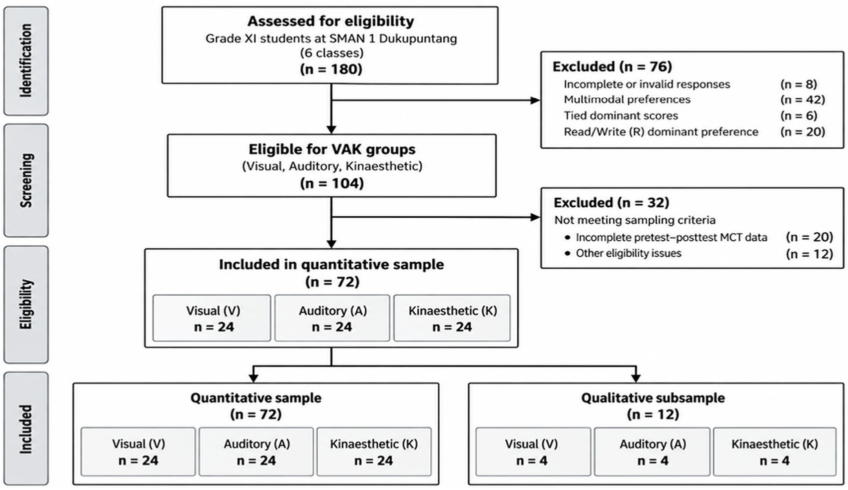


Figure 3. Participant screening and sampling flow diagram.

A criterion-based purposive sampling strategy was subsequently applied to obtain comparable representation across the three modality-preference categories [43]. From the 104 eligible students, 72 participants with complete preference classifications and paired pretest-posttest MCT data were included in the quantitative phase. The final analytical sample contained 24 students in each category: Visual (n = 24; 33.3%), Auditory (n = 24; 33.3%), and Kinaesthetic (n = 24; 33.3%). This balanced structure facilitated between-group comparison but should be interpreted as an analytical sampling decision rather than as an estimate of preference prevalence in the source cohort.

After quantitative analysis, 12 participants were purposively selected for the explanatory qualitative phase. The subsample comprised four students from each preference category - Visual (n = 4), Auditory (n = 4), and Kinaesthetic (n = 4) - while also incorporating variation in MCT performance. Selection was intended to obtain information-rich cases capable of clarifying, contextualizing, or elaborating quantitative patterns, consistent with explanatory mixed-methods sampling [43].

3.3. Learning Intervention

The intervention comprised six 90-minute sessions of e-module-supported Project-Based Learning (PjBL) organized around a contextual trigonometry project. Each session emphasized one principal phase of the PjBL cycle: problem orientation, project planning, investigation, product development, presentation, and reflection/evaluation. Students worked in groups of four to five and developed their mathematical responses progressively across the six sessions, from problem identification and investigation planning to solution testing, presentation, and evaluation of the reasoning used. The teacher structured tasks, monitored progress, posed guiding questions, and provided formative feedback without directly supplying solutions.

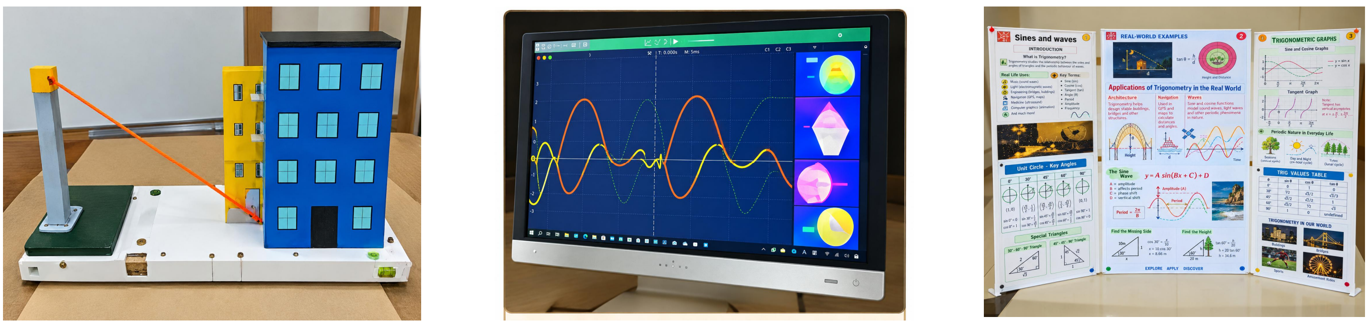
The e-module was integrated into the project sequence as a digital instructional scaffold rather than implemented as a separate treatment. Its functions were aligned with the cognitive demands of each PjBL phase. Contextual materials and instructional videos supported initial problem interpretation; digital worksheets structured project planning; GeoGebra-based dynamic representations supported mathematical exploration; interactive exercises and formative feedback were used during solution development; digital portfolio facilities supported presentation; and self-assessment activities were used during reflection. This alignment provided repeated opportunities to interpret mathematical information, identify relationships, evaluate procedures, formulate inferences, justify conclusions, and monitor reasoning across the project cycle.

The project addressed trigonometric identities, function values, and graphical representations. Its final output consisted of a mathematically justified solution or model supported by calculations, representations, and documentation compiled in a digital project portfolio. The e-module was developed using the ADDIE framework (Analysis, Design, Development, Implementation, and Evaluation) and was reviewed and revised by mathematics-education experts before classroom implementation [44]. Components included trigonometric materials, instructional videos, GeoGebra simulations, interactive exercises, formative assessments, digital worksheets, reflection activities, and portfolio submission facilities.

Table 2 summarizes the intervention by aligning each PjBL phase with its principal mathematical activity, digital support, and expected output. Supplementary Table S1 provides the corresponding session-level detail on teacher facilitation, student activities, project tasks, expected outputs, and e-module features.

Table 2. Alignment of PjBL phases, mathematical activities, and e-module support.

Session	PjBL phase	Core mathematical activity	E-Module support	Expected output
1	Problem orientation	Interpreting the contextual problem and identifying relevant mathematical information	Instructional video and contextual materials	Initial problem formulation
2	Project planning	Planning the investigation and selecting mathematical strategies	Digital worksheet and planning prompts	Project plan
3	Investigation	Exploring relationships, testing representations, and generating evidence	GeoGebra simulations and interactive representations	Mathematical evidence and investigation results
4	Product development	Constructing, testing, evaluating, and revising the solution	Interactive exercises and formative feedback	Mathematical solution or model
5	Presentation	Explaining and justifying mathematical reasoning	Digital portfolio and presentation resources	Presented solution and project portfolio
6	Reflection and evaluation	Evaluating reasoning, identifying errors, revising conclusions, and monitoring performance	Self-assessment and reflection activities	Revised portfolio and individual reflection



(a) (b) (c)

Figure 4. Student project outputs demonstrating the application of trigonometric concepts in Project-Based Learning (PjBL): (a) Angle of elevation; (b) Sinusoidal wave modeling; (c) Trigonometry posters.

Figure 4 presents representative student project outputs produced during the PjBL sequence: an angle-of-elevation model, sinusoidal-wave modeling, and trigonometry posters. These artifacts illustrate how trigonometric concepts were represented and communicated across physical, digital, and visual formats during the project.

The three examples in Figure 4 document project products rather than isolated effects of individual e-module features. The angle-of-elevation model externalizes geometric relationships in a contextual form; the sinusoidal-wave display represents functional behavior dynamically; and the posters consolidate symbolic, graphical, and explanatory content for communication. Their evidential role is descriptive: they show the forms of mathematical representation produced during the instructional sequence and do not establish the independent contribution of any single digital component.

3.4. Instrument

3.4.1. VARK Learning Preference Questionnaire

Students' self-reported modality preferences were assessed using a 16-item VARK learning preference questionnaire adapted for the Indonesian secondary-school context. Each item contained response alternatives corresponding to Visual (V), Auditory/Aural (A), Read/Write (R), and Kinaesthetic (K) modalities. Participants were classified according to their predominant modality score. In this study, the questionnaire served only as a descriptive grouping instrument; the resulting categories were not treated as fixed cognitive types and were not used to match students with modality-specific instruction. This restriction is consistent with evidence that instructional matching to a preferred sensory modality does not reliably improve learning outcomes [18].

The adapted questionnaire yielded Cronbach's alpha = 0.83, while expert review produced a reported Content Validity Ratio (CVR) > 0.80. These indices were used as evidence supporting the instrument's use for preference classification in the study sample. Previous psychometric work on VARK has provided preliminary support for a multidimensional structure while also identifying limitations in item wording and scoring, warranting cautious interpretation of modality classifications [45].

Table 3. Characteristics of the VARK learning preference questionnaire.

Component	Specification
Construct	Self-reported learning modality preference
Instrument	Adapted VARK learning preference questionnaire
Number of items	16
Modalities	Visual, Auditory/Aural, Read/Write, Kinaesthetic
Response structure	Four modality-based alternatives per item
Classification rule	Predominant modality score
Analytical groups retained	Visual, Auditory, Kinaesthetic
Excluded profiles	Incomplete, multimodal, tied dominant scores, Read/Write dominant
Internal consistency	Cronbach's (α = 0.83)
Content validity	CVR > 0.80
Analytical role	Descriptive grouping variable; not used for instructional matching

3.4.2. Mathematical Critical Thinking Ability Test

Mathematical critical thinking (MCT) was assessed using a six-item open-ended test developed for the trigonometry content addressed in the intervention. The instrument operationalized MCT through interpretation, analysis, evaluation, inference, explanation, and self-regulation, consistent with established conceptions of critical thinking as coordinated interpretive, evaluative, inferential, explanatory, and metacognitive judgment [22]. In mathematics, these processes are expressed through interpreting representations, analyzing relationships, evaluating procedures and arguments, drawing warranted conclusions, justifying reasoning, and monitoring one's own solution process [2]. Each dimension was represented by one constructed-response item that required students to make their reasoning explicit. The items addressed trigonometric identities, function values, mathematical relationships, graphical representations, and problem-solving situations.

Before administration in the main study, the test was piloted with 30 students outside the analytical sample. Three mathematics-education specialists evaluated content relevance and construct alignment, producing an average content validity index (CVI) of 0.89. The CVI was treated as evidence of expert agreement on item relevance to the intended construct [46], [47]. The six-item instrument yielded Cronbach's alpha = 0.87. Because the items represented conceptually distinct MCT dimensions, this coefficient was interpreted at the overall-test level and not as evidence that the six dimensions constituted interchangeable or independently reliable subscales [48].

Item-level performance was examined using corrected item-total correlations, difficulty indices, and discrimination indices. Retention criteria were a corrected item-total correlation of at least 0.30, a difficulty index from 0.30 to 0.70, and a discrimination index of at least 0.40. All six items met these criteria. Corrected item-total correlations ranged from 0.60 to 0.68, difficulty indices from 0.52 to 0.62, and discrimination indices from 0.46 to 0.54 (Table 4).

Table 4. Item-Level psychometric properties of the mathematical critical thinking instrument.

Item	MCT dimension	Corrected item–total correlation	Difficulty index (P)	Discrimination index (D)	Decision
1	Interpretation	0.68	0.62	0.54	Retained
2	Analysis	0.65	0.58	0.51	Retained
3	Evaluation	0.61	0.52	0.48	Retained
4	Inference	0.64	0.60	0.52	Retained
5	Explanation	0.63	0.56	0.49	Retained
6	Self-regulation	0.60	0.54	0.46	Retained

Note. Item-retention criteria were corrected item–total correlation ≥ 0.30, difficulty index = 0.30 – 0.70, and discrimination index ≥ 0.40.

Table 5. Characteristics of the VARK Learning Preference Questionnaire.

Item	MCT dimension	Trigonometric content	Reasoning process assessed	Raw score
1	Interpretation	Trigonometric identity	Identifying and interpreting relevant mathematical information, conditions, symbols, and representations	0–4
2	Analysis	Trigonometric identity	Identifying mathematical relationships and selecting an appropriate solution strategy	0–4
3	Evaluation	Trigonometric function value	Examining the correctness and validity of a mathematical procedure or result	0–4
4	Inference	Trigonometric relationship	Drawing a logically supported conclusion from available mathematical evidence	0–4
5	Explanation	Graphical representation	Communicating and justifying mathematical reasoning and conclusions	0–4
6	Self-regulation	Trigonometric problem/representation	Monitoring, checking, evaluating, and revising the solution process	0–4
Total	**MCT dimensions**	**Trigonometry**	**Overall MCT performance**	**0–24**

Table 5 summarizes the item-construct alignment used in the MCT test. Each item was linked explicitly to trigonometric content and to the reasoning process it was intended to elicit, providing content-based evidence for construct representation [49], [50]. Interpretation referred to identifying and making meaning from relevant mathematical information and representations; analysis to identifying relationships and selecting an appropriate strategy; evaluation to examining the validity of procedures and results; inference to deriving conclusions supported by available evidence; explanation to communicating and justifying mathematical reasoning; and self-regulation to monitoring, checking, and revising the solution process.

These processes are consistent with frameworks that distinguish the quality of mathematical reasoning from the correctness of the final answer alone [2], [51]. MCT was therefore operationalized as a set of interrelated reasoning processes enacted within mathematical activity.

Each response was evaluated using a five-level analytic rubric ranging from 0 to 4. A score of 0 indicated no relevant evidence of the targeted reasoning process, whereas a score of 4 represented a mathematically accurate, complete, and adequately justified response. Analytic scoring was used because constructed-response tasks require reasoning quality to be judged against explicit performance criteria rather than inferred solely from answer correctness [52]. The six items produced a maximum raw score of 24. Because each MCT dimension was represented by one item, dimension-level scores were treated as item-specific performance indicators rather than independently validated multi-item subscales. Supplementary Table S2 provides the dimension-specific scoring descriptors. For reporting and comparison, total raw scores were transformed linearly to a 0-100 scale.

$$MCT_{0-1\ 00}=\frac{Raw\ MCT\ score}{24}\times 100 \tag{1}$$

Under this transformation, a raw item score of 4 contributed 16.67 points to the 0-100 total scale, and the six items together yielded a maximum transformed score of 100. The transformation changed only the reporting metric; it did not alter rank ordering or relative distances among total scores. Unless otherwise stated, descriptive and inferential analyses used the transformed scores.

3.4.3. Double-Scoring Procedure and Reliability Analysis

All pretest and posttest responses to the six open-ended MCT items were independently scored by two raters using the same dimension-specific analytic rubric. Inter-rater agreement was assessed with the intraclass correlation coefficient (ICC). A two-way random-effects, absolute-agreement, single-measure model, ICC(2,1), was used in accordance with the stated objective of evaluating agreement between independently assigned numerical ratings [53].

ICC estimates and 95% confidence intervals were calculated for each MCT dimension and for the total score at pretest and posttest. ICC was selected because it incorporates both between-subject variability and rater disagreement [54], [55]. Following Koo and Li [56], values were interpreted as poor (< 0.50), moderate (0.50-0.75), good (0.75-0.90), or excellent (> 0.90). Inter-rater agreement was evaluated before the subsequent quantitative analyses.

3.5. Data Collection Procedure

Data collection followed the explanatory sequential mixed-methods design, with quantitative assessment preceding the principal interview follow-up [57], [58]. At baseline, students completed the VARK learning-preference questionnaire and the MCT pretest. They then participated in six e-module-supported PjBL sessions. Structured classroom observations documented mathematical reasoning and engagement during these sessions, and student work artifacts were retained as process evidence. The MCT posttest was administered immediately after the final session.

After quantitative analysis, 12 students were purposively selected to represent the three VAK preference groups and variation in MCT performance. Each completed an individual semi-structured interview lasting approximately 40-45 minutes. Interview accounts were examined alongside classroom observations and student work artifacts to contextualize the quantitative patterns. Evidence was compared across the three sources for each participant and MCT dimension, and convergent, complementary, and discrepant evidence was documented in a triangulation matrix.

3.6. Data Analysis

3.6.1. Quantitative Analysis

Descriptive statistics included the mean, standard deviation, minimum, and maximum for total and dimension-specific MCT scores. Pretest-posttest change was examined using paired-samples t-tests, with mean differences, 95% confidence intervals, and Cohen's dz reported as the within-participant standardized effect size [59]. For the six dimension-level comparisons, p-values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) procedure to limit inflation of Type I error across multiple tests [60].

Differences across Visual, Auditory, and Kinaesthetic preference groups were examined using one-way ANOVA for posttest scores and normalized gain. Normalized gain was calculated as g = (posttest - pretest) / (100 - pretest), representing the proportion of the maximum attainable improvement achieved from baseline (Hake, 1998). Dimension-specific pretest-posttest change scores were also compared by one-way ANOVA with FDR adjustment. An ANCOVA was additionally conducted with posttest MCT as the dependent variable, pretest MCT as the covariate, and VAK preference as the fixed factor. ANOVA effects were reported as eta squared (η^2), and ANCOVA effects as partial eta squared (partial η^2). Statistical significance was evaluated at $\alpha = 0.05$.

3.6.2. Qualitative Analysis

Interview transcripts, classroom observation records, and student work artifacts were analyzed using a theory-informed thematic coding framework organized around the six MCT dimensions. Codes captured observable reasoning processes related to interpretation, analysis, evaluation, inference, explanation, and self-regulation, as specified in Table 6. Evidence was compared within and across participants to identify recurrent, complementary, and discrepant patterns. Negative cases were retained and examined against the corresponding interviews, observations, and artifacts so that variation capable of qualifying dominant interpretations was not removed from the analysis.

Table 6. Qualitative coding framework for the Six MCT dimensions.

Main Category	Sub-Codes	Indicators	Data Sources
Interpretation	Understanding problem context, identifying information	Students explain known and unknown variables and problem requirements	Written solutions, interviews
Analysis	Identifying relationships, selecting strategies	Students connect mathematical concepts and determine solution procedures	Written solutions, interviews
Evaluation	Checking accuracy, providing justification	Students evaluate the validity of their solutions and reasoning	Written solutions, interviews
Inference	Drawing conclusions, generating alternatives	Students formulate conclusions and alternative approaches	Written solutions, interviews
Explanation	Structural justification, presenting arguments, articulating reasoning	Students communicate step-by-step mathematical arguments clearly, justify identity transformations, and translate visual/algebraic concepts	Written solutions, interviews, observation notes
Self-Regulation	Procedural monitoring, self-correction, domain verification	Students monitor their solution process, verify domain restrictions (e.g., $x \neq \frac{\pi}{2} + k\pi$), and independently correct mathematical errors	Written solutions, interviews, reflection logs

3.6.3. Mixed-Methods Integration

Quantitative and qualitative findings were integrated through a joint display that linked statistical outcomes with corresponding evidence from interviews, observations, and student artifacts. Integration examined whether the qualitative evidence converged with, complemented, qualified, or diverged from the quantitative patterns. This approach follows mixed-methods integration through merging and joint display, in which datasets are compared to develop integrated interpretations rather than reported as independent findings [21]. The resulting meta-inferences addressed both observed changes in MCT performance and the reasoning processes through which those patterns were expressed across VAK preference groups.

4. RESULTS

4.1. Distribution of Learning Preferences and Baseline Characteristics

The final analytical sample comprised 72 students, with equal representation across the three self-reported modality-preference groups: Visual (n = 24, 33.3%), Auditory (n = 24, 33.3%), and Kinaesthetic (n = 24, 33.3%). This balanced composition resulted from the sampling procedure described in the Methods and represents the structure of the analytical sample rather than the prevalence of these preferences in the source cohort.

Baseline MCT scores were descriptively similar across the three groups. The overall transformed pretest mean was 43.95 (SD = 6.26), with scores ranging from 31.00 to 58.00 on the 0-100 scale. Group means were 44.12 (SD = 6.18) for Visual, 43.50 (SD = 6.45) for Auditory, and 44.17 (SD = 6.31) for Kinaesthetic students. These descriptive values show limited initial separation but are not evidence of statistical equivalence.

Table 7. Baseline distribution of learning preferences and mathematical critical thinking scores.

Learning preference	n	%	Pretest MCT, Mean (SD)
Visual	24	33.3	44.12 (6.18)
Auditory	24	33.3	43.50 (6.45)
Kinaesthetic	24	33.3	44.17 (6.31)
Total	**72**	**100.0**	**43.95 (6.26)**

4.2. Overall Change in Mathematical Critical Thinking

Inter-rater agreement was good to excellent for the dimension-specific MCT scores and excellent for the total score. Dimension-specific ICCs ranged from 0.818 to 0.878 at pretest and from 0.892 to 0.924 at posttest. Full ICC estimates, 95% confidence intervals, and their interpretations are reported in Supplementary Table S3.

The transformed MCT mean was 43.95 (SD = 6.26) at pretest and 78.18 (SD = 5.77) at posttest. Observed scores ranged from 31.00 to 58.00 at pretest and from 64.00 to 88.00 at posttest. The mean within-student difference was 34.23 points (SD = 8.38), 95% CI [32.26, 36.20], $t(71) = 34.66$, $p < 0.001$, corresponding to Cohen's d_z = 4.08. Because the study used a single-group design, this estimate describes change during the instructional period and does not isolate a causal intervention effect.

Table 8. Pretest–Posttest changes in overall mathematical critical thinking (MCT) scores

Outcome / Statistic	Total MCT
Pretest Mean (SD)	43.95 (6.26)
Posttest Mean (SD)	78.18 (5.77)
Mean change (SD)	34.23 (8.38)
95% CI for mean change	[32.26, 36.20]
Paired-samples t-test	$t(71) = 34.66$, $p < 0.001$
Cohen's dz	4.08

4.3. Dimension-Specific Changes in MCT

Post-test means exceeded pretest means across all six MCT dimension indicators. Mean transformed scores changed from 7.64 to 13.52 for Interpretation, 7.21 to 13.08 for Analysis, 6.85 to 12.38 for Evaluation, 7.42 to 13.19 for Inference, 7.50 to 13.01 for Explanation, and 7.33 to 13.00 for Self-regulation. Mean differences ranged from 5.51 to 5.88 points, and all six paired comparisons remained statistically significant after Benjamini-Hochberg FDR adjustment (all adjusted $p < 0.001$).

The posttest profile was relatively similar across the six item-level indicators. Interpretation had the highest transformed mean (13.52), whereas Evaluation had the lowest (12.38). These contrasts are descriptive only: because each dimension was represented by a different constructed-response item, they should not be interpreted as evidence that one MCT dimension was intrinsically easier, more difficult, or more developed than another.

Table 9. Pre- and posttest change across mathematical critical thinking dimensions

MCT dimension	Pretest Mean (SD)	Posttest Mean (SD)	Mean change	Posttest raw mean (0-4)	FDR-adjusted p
Interpretation	7.64 (1.42)	13.52 (1.28)	5.88	3.24	< 0.001
Analysis	7.21 (1.35)	13.08 (1.21)	5.87	3.14	< 0.001
Evaluation	6.85 (1.28)	12.38 (1.15)	5.53	2.97	< 0.001
Inference	7.42 (1.38)	13.19 (1.24)	5.77	3.16	< 0.001

MCT dimension	Pretest Mean (SD)	Posttest Mean (SD)	Mean change	Posttest raw mean (0-4)	FDR-adjusted p
Explanation	7.50 (1.31)	13.01 (1.19)	5.51	3.12	< 0.001
Self-regulation	7.33 (1.30)	13.00 (1.22)	5.67	3.12	< 0.001

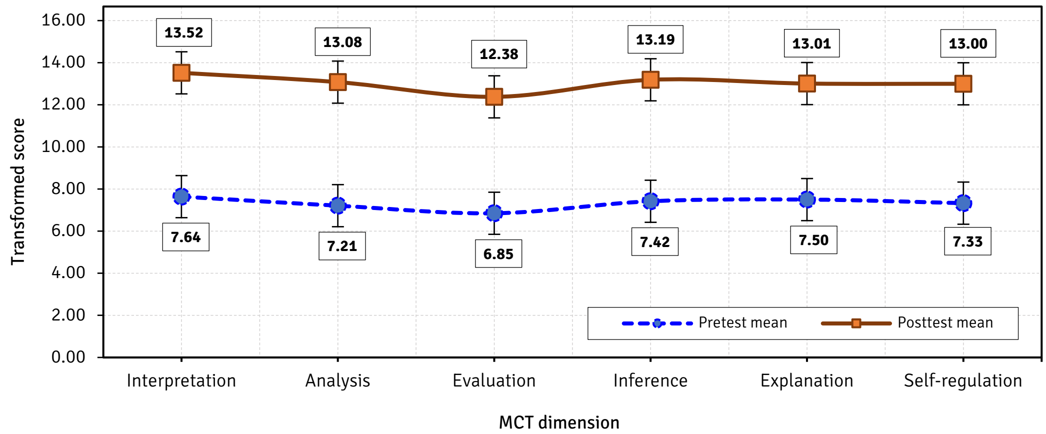


Figure 5. Mean pre- and posttest scores across the six mathematical critical thinking dimensions on the transformed scale.

4.4. MCT Performance Across VAK Preference Groups

Post-test MCT scores were closely clustered across the three self-reported preference groups. Mean transformed scores were 78.42 (SD = 5.81) for Visual, 77.85 (SD = 5.92) for Auditory, and 78.25 (SD = 5.68) for Kinaesthetic students; the largest difference between group means was 0.57 points on the 0-100 scale. Mean normalized gains were 0.61 (SD = 0.09), 0.60 (SD = 0.09), and 0.61 (SD = 0.09), respectively, giving a maximum between-group difference of 0.01.

One-way ANOVA detected no statistically significant between-group difference in posttest MCT, $F(2, 69) = 0.062$, $p = 0.940$, $\eta^2 = 0.002$, or normalized gain, $F(2, 69) = 0.071$, $p = 0.931$, $\eta^2 = 0.002$. Adjustment for baseline performance produced the same inferential conclusion: VAK preference was not significantly associated with posttest MCT, $F(2, 68) = 1.60$, $p = 0.209$, partial $\eta^2 = 0.045$. The pretest covariate was also not significant, $F(1, 68) = 0.021$, $p = 0.885$, partial $\eta^2 < 0.001$. The full ANCOVA decomposition is reported in Supplementary Table S4.

Dimension-specific change scores likewise showed no statistically significant differences across the three preference groups after FDR adjustment; adjusted p-values ranged from 0.713 to 0.972. These analyses did not detect systematic VAK-related differences in MCT outcomes within this sample. Because non-significant tests do not establish statistical equivalence, the result is interpreted as an absence of detected between-group differences rather than evidence that the groups performed identically.

Table 10. Mathematical critical thinking outcomes across VAK preference groups.

VAK preference	n	Pretest Mean (SD)	Posttest Mean (SD)	Normalized gain, g Mean (SD)
Visual	24	44.12 (6.18)	78.42 (5.81)	0.61 (0.09)
Auditory	24	43.50 (6.45)	77.85 (5.92)	0.60 (0.09)
Kinaesthetic	24	44.17 (6.31)	78.25 (5.68)	0.61 (0.09)
Total	**72**	**43.95 (6.26)**	**78.18 (5.77)**	**0.61 (0.09)**

Table 11. Between-group comparisons of mathematical critical thinking outcomes.

Outcome	Statistical test	Test statistic	p	Effect size
Posttest MCT	One-way ANOVA	$F(2, 69) = 0.062$	0.940	$\eta^2 = 0.002$
Normalized gain	One-way ANOVA	$F(2, 69) = 0.071$	0.931	$\eta^2 = 0.002$
Adjusted posttest MCT	ANCOVA	$F(2, 68) = 1.600$	0.209	partial $\eta^2 = 0.045$

The ANCOVA yielded the same inferential conclusion after baseline adjustment: VAK preference was not significantly associated with posttest MCT, $F(2, 68) = 1.60$, $p = 0.209$, partial $\eta^2 = 0.045$. Supplementary Table S4 provides the complete ANCOVA decomposition. Dimension-specific change scores also showed no statistically significant between-group differences after FDR adjustment; group-specific mean changes, ANOVA statistics, and adjusted p-values are reported in Supplementary Table S5.

4.5. Qualitative Findings on MCT Processes

Qualitative analysis of interviews (n = 12), classroom observations, and student artifacts identified six recurring modes through which mathematical critical thinking became visible: visualization, verbalization, digital manipulation, trial-and-check reasoning, peer discussion, and reflection. Representative evidence for each mode was present in at least two qualitative source types, providing cross-source corroboration while preserving differences in what interviews, observations, and artifacts made visible. These modes were treated as overlapping forms of mathematical engagement rather than as exclusive properties of particular VAK categories.

Visualization and representational coordination. Dynamic representations supported coordination between symbolic trigonometric expressions and spatial relationships. Within the purposive qualitative subsample, Visual-preference participants often referred to graphs, diagrams, and the unit-circle representation when explaining how mathematical relationships became interpretable.

> *"When I opened the dynamic unit circle in the E-Module and saw the right triangle transform visually ... the identity immediately made sense."* (P02-V, INT-02; OBS-S3)

Verbalization and error detection. Verbal explanation was used to sequence proof steps and expose gaps in reasoning. This pattern was especially visible in interviews with Auditory-preference participants, although it also appeared in peer interaction across preference groups.

> *"Verbalizing the logic helped me spot errors in my own reasoning."* (P06-A, INT-06; OBS-S2)

Digital manipulation and iterative testing. Manipulation of digital parameters and trial-and-check activity supported exploratory analysis, hypothesis testing, and revision. Participants used interactive representations and scratchpads to examine whether transformations remained valid and to abandon strategies that could not be sustained.

> *"If an approach collapsed into an unsolvable expression, I checked where it broke down and restarted."* (P10-K, INT-10; OBS-S3)

Peer discussion and reflective monitoring. Collaborative exchanges enabled participants to compare representations and justify alternative strategies, while embedded reflection prompts supported procedural checking, error detection, and attention to domain restrictions.

> *"Combining our methods made the final solution much tighter."* (P05-A, INT-05; OBS-S5)

> *"I realized I kept forgetting domain restrictions for tangent functions, so I added a checking step to my habit routine."* (P11-K, INT-11; ART-S5)

The six modes overlapped across participants and were not treated as exclusive properties of particular VAK categories. Supplementary Table S6 links these qualitative patterns with the corresponding quantitative findings at the MCT-dimension level. Supplementary Table S7 provides additional representative participant evidence, and Supplementary Table S8 documents cross-source coverage across interviews, classroom observations, and student artifacts.

4.6. Quantitative and Qualitative Findings

Quantitative and qualitative findings were integrated by linking statistical outcomes with evidence from interviews, classroom observations, and student artifacts. The comparison examined whether qualitative evidence converged with, complemented, qualified, or diverged from the quantitative patterns. Table 12 presents the compact joint display, and Supplementary Table S6 provides a dimension-level version.

At the outcome level, the marked pretest-posttest increase in MCT was accompanied by qualitative evidence of explicit interpretation, strategy testing, justification, error checking, and reflection during the project. Dimension-specific evidence followed a similar pattern, although evaluation was qualified by instances in which some students still required prompts to verify procedures or domain restrictions.

Table 12. Joint display of quantitative and qualitative findings.

Quantitative finding	Qualitative evidence	Integration	Meta-inference
Overall MCT increased from 43.95 to 78.18; mean change = 34.23; $p < .001$.	More explicit interpretation, strategy testing, explanation, monitoring, and revision were observed across data sources.	Convergence	Higher MCT scores were accompanied by observable changes in mathematical reasoning during the instructional period.
All six MCT dimensions improved after FDR adjustment.	Visualization, verbalization, manipulation, trial-and-check,	Convergence + complementarity	MCT change was multidimensional and expressed through multiple forms of engagement.

Quantitative finding	Qualitative evidence	Integration	Meta-inference
	discussion, and reflection supported different reasoning processes.		
Evaluation improved but remained descriptively lower than Interpretation at posttest.	Some students required prompts to verify procedures or domain restrictions.	Qualification	Improved evaluation did not imply that verification had become uniformly spontaneous.
No significant VAK effect on posttest MCT or normalized gain.	Preference groups differed in recurring emphases, but substantial cross-group overlap was observed.	Complementarity + qualification	Similar performance outcomes coexisted with heterogeneous modes of reasoning.
Qualitative evidence varied in visibility across interviews, observations, and artifacts.	Some reasoning was clearer in one source than another.	Convergence + divergence	Triangulation captured aspects of MCT not fully represented by any single qualitative source.

The VAK comparison produced a different pattern of integration. Quantitative analyses did not detect between-group differences in MCT performance, whereas qualitative evidence showed variation in how reasoning was expressed. Visual-preference participants more often emphasized graphical representation, Auditory-preference participants verbalization and discussion, and Kinaesthetic-preference participants manipulation and trial-and-check; however, these processes overlapped substantially across groups. The integrated finding is therefore one of outcome-level similarity in the observed sample alongside process-level heterogeneity, without implying that preference categories determine distinct cognitive pathways.

5. DISCUSSION

5.1. Changes in MCT Performance

The most prominent quantitative result was a large pretest-posttest difference in MCT during the six-session e-module-supported PjBL period. The overall mean increased from 43.95 (SD = 6.26) to 78.18 (SD = 5.77), and all six item-level MCT indicators showed statistically significant change after FDR adjustment. Because the design lacked a comparison group, these findings document within-student change over the instructional period rather than a causal effect attributable uniquely to PjBL, the e-module, or GeoGebra. The multidimensional pattern is nevertheless consistent with conceptualizations of MCT as coordinated interpretive, analytical, evaluative, inferential, explanatory, and self-regulatory reasoning [61].

Mathematics-education research similarly distinguishes critical mathematical reasoning from procedural reproduction by emphasizing analysis, evaluation, justification, and reflection [2], [51]. The present pattern is therefore more informative when interpreted across the six targeted reasoning processes than as a change in a single undifferentiated score. At the same time, each dimension was represented by one item, so dimension-level results remain item-specific indicators rather than independently reliable subscales.

A plausible process-level interpretation is that the PjBL sequence repeatedly elicited the forms of reasoning represented in the MCT framework. Problem orientation required interpretation; planning and investigation required analysis, strategy selection, and evaluation of evidence; product development involved testing and revision; presentation required explanation and justification; and reflection foregrounded monitoring of reasoning. These demands are consistent with literature linking problem- and project-based environments to opportunities for higher-order thinking through inquiry, collaboration, and knowledge construction [62], [63].

This interpretation does not imply that project completion itself caused the observed score change. Loyens et al. [64] showed that higher-order and critical thinking are conceptualized and measured differently across problem- and project-based learning studies, underscoring the importance of alignment between tasks and targeted cognitive processes. Guo et al. [65] likewise identify learning processes and artifact production as central features of PjBL. Umbrella evidence reports generally positive directions for higher-order outcomes but also substantial heterogeneity and uneven methodological quality [12]. The present findings are therefore most consistent with repeated engagement in cognitively demanding, reflective, and evidence-based activity rather than with a simple project-format effect.

The e-module could have provided additional scaffolding by coordinating symbolic, graphical, verbal, and interactive representations. Multiple external representations can support STEM learning when learners receive sufficient assistance in translating and integrating information across formats [66], and their value depends on representational support rather than on the number of representations alone. Meta-analytic evidence also reports positive associations between dynamic mathematical software and mathematics performance [67], including GeoGebra-specific effects on achievement [68]. In the present study, graphs, simulations, worksheets, feedback, and interactive tools are therefore interpreted as potential supports for conjecture testing, representational comparison, error detection, and revision, not as independently isolated treatment components.

Dimension-level variation adds a more limited descriptive observation. Interpretation had the highest posttest item mean, whereas Evaluation had the lowest. Because these dimensions were represented by different single items, the contrast cannot establish that evaluation was intrinsically weaker than interpretation. The qualitative evidence is nevertheless compatible with a narrower process interpretation: some participants still required prompts to check procedural validity or domain restrictions, suggesting that evaluative checking was not uniformly spontaneous within the observed activities.

Reflection prompts and formative feedback may likewise have made monitoring and revision more explicit. Meta-analytic evidence indicates that regulated-learning scaffolds can improve regulation strategies and academic performance [69], while metacognitive approaches have been associated with mathematical critical thinking [61]. These external findings provide a theoretically plausible account of the observed process patterns, but the present single-group design cannot estimate the independent contribution of reflection prompts, feedback, or other e-module features.

5.2. MCT Performance Across VAK Preference Groups

The between-group analyses qualify the overall pretest-posttest pattern. Posttest means and normalized gains were closely clustered across Visual, Auditory, and Kinaesthetic groups, and the ANCOVA did not detect a statistically significant association between VAK preference and posttest MCT after adjustment for pretest performance. Dimension-specific change scores also showed no statistically significant VAK-related differences after FDR adjustment. Within this sample and instructional context, the analyses therefore did not provide statistically detectable evidence that self-reported VAK preference differentiated MCT performance.

These results do not establish that the groups were equivalent or that all between-group differences were absent. A non-significant test is not an equivalence test. The findings instead provide limited evidence for using VAK preference as an explanation of MCT performance in this sample. This interpretation is consistent with contemporary evidence questioning modality-matching assumptions. Clinton-Lisell and Litzinger [70] found that although a small average matching effect appeared across 21 studies, the crossover interaction required by the learning-styles matching hypothesis occurred in only a minority of outcomes. Melzner and Kappes [71] likewise reported no significant interaction between visual/verbal preferences and matched presentation modes for learning performance or metacognitive judgments.

The findings are therefore consistent with treating VAK as a descriptive preference variable rather than a fixed cognitive type or prescription for differentiated instruction. Students with different reported preferences reached similar observed MCT outcomes within the same multimodal environment. This interpretation also accords with evidence that the usefulness of multiple representations depends primarily on their informational function, coordination, and instructional support rather than on matching one representation to one category of learner [66].

5.3. MCT Processes Across VAK Preference Groups

The qualitative findings clarify why similar quantitative outcomes should not be equated with identical reasoning processes. Participants enacted MCT through overlapping combinations of visualization, verbalization, digital manipulation, trial-and-check reasoning, peer discussion, and reflection. Within the purposive subsample, Visual-preference participants often emphasized dynamic graphs and geometric representations, Auditory-preference participants frequently articulated solution steps and used dialogue, and Kinaesthetic-preference participants commonly manipulated digital parameters and tested conjectures interactively. These tendencies were not exclusive, and negative cases showed substantial cross-group use of the same processes.

Across the six MCT dimensions, the qualitative evidence suggests an interconnected reasoning cycle. Interpretation involved coordinating contextual, symbolic, and graphical information; analysis involved identifying relationships and generating strategies; evaluation required checking procedures and validity; inference transformed tested relationships into defensible conclusions; explanation made reasoning visible to others; and self-regulation connected monitoring, error detection, and revision. This process-oriented account is consistent with research emphasizing explicit opportunities to analyze, justify, evaluate, and reflect on mathematical ideas [2], [3]. The broader PjBL literature similarly identifies authentic problems, inquiry, dialogue, problem solving, and evaluation as conditions capable of eliciting critical-thinking processes [64].

Digital manipulation and peer interaction appeared to serve complementary process functions in the qualitative data. Dynamic tools allowed participants to alter parameters, inspect consequences, test tentative relationships, and revise unsuccessful approaches. This interpretation is consistent with evidence that dynamic mathematical software can support mathematics performance when digital affordances are embedded within meaningful mathematical activity [67], [68]. Peer discussion, in contrast, required students to externalize reasoning, compare alternative interpretations, and defend mathematical choices, linking explanation with evaluation in ways also emphasized in critical-thinking-oriented PjBL [64].

These processes should not be assigned exclusively to particular VAK groups. A student classified as Visual could reason verbally, an Auditory-preference student could benefit from graphs, and a Kinaesthetic-preference student could construct symbolic justification. This cross-modal flexibility is consistent with evidence that students do not learn optimally only when presentation mode matches a preferred modality [70], [71]. The qualitative findings therefore position the preference categories as descriptive indicators of recurring emphases in engagement rather than as predictors of distinct reasoning mechanisms or MCT performance.

5.4. Integration of Quantitative and Qualitative

Mixed-methods integration adds process context to the statistical findings. Quantitatively, MCT scores increased from pretest to posttest across the total score and all six item-level indicators, while statistically significant differences across VAK groups were not detected. Qualitatively, the same MCT dimensions became visible through partially different combinations of representation, dialogue, manipulation, experimentation, and reflection. The resulting meta-inference is one of similar observed outcomes across VAK groups alongside heterogeneous and overlapping forms of reasoning.

Three relationships between the strands are evident. Convergence occurred where higher posttest performance was accompanied by explicit qualitative evidence of interpretation, strategy testing, error checking, explanation, and revision during the instructional period. Complementarity occurred because interviews, observations, and artifacts revealed processes that test scores alone could not identify, such as visualization, verbal reconstruction, peer critique, digital manipulation, and iterative testing. Qualification occurred where quantitative score increases did not imply fully autonomous reasoning: evaluation scores increased, yet some participants still required prompts to verify assumptions or domain restrictions. Similar VAK outcomes likewise coexisted with different emphases in engagement.

This form of integration reflects the analytical purpose of explanatory sequential mixed methods. Peters and Fàbregues [72] argue that joint displays should move beyond parallel presentation and generate higher-order meta-inferences by systematically examining convergence, complementarity, and divergence across strands. In the present study, triangulation across interviews, observations, and artifacts was analytically useful because each source represented a different aspect of reasoning: written work recorded products, interviews elicited retrospective articulation, and observations documented behavior in situ. Discrepancies were retained because no single source fully represented students' mathematical thinking.

The integrated evidence is more compatible with multimodal access than with VAK-based instructional matching. Within a common learning environment, students used visual, verbal, symbolic, interactive, and collaborative resources in overlapping ways. A defensible pedagogical implication is therefore to provide coordinated opportunities to visualize relationships, verbalize reasoning, manipulate representations, justify procedures, evaluate alternatives, and reflect on errors, rather than restricting access according to a presumed modality category. This interpretation is consistent with evidence on multiple representations, dynamic mathematics environments, regulated-learning scaffolding, and the limited empirical basis for modality matching [66], [67], [69], [70].

5.5. Implications, Boundaries and Future Research

Two instructional implications follow within the scope of the study. First, MCT-oriented activities can be distributed across a project sequence so that students repeatedly interpret information, compare representations, justify transformations, evaluate conclusions, communicate reasoning, and revise errors. Second, digital affordances are most defensible when aligned with specific reasoning demands: dynamic graphs can support exploration, interactive workspaces can support conjecture testing, formative feedback can prompt error detection, and reflection prompts can make monitoring explicit. This interpretation is consistent with meta-analytic evidence on computer-assisted mathematics integrated with PjBL, while recognizing that outcomes vary by educational level, mathematical content, and technology [73].

Interpretation is limited by the single-group design, the one-school setting, the balanced purposive analytical sample, exclusion of multimodal and Read/Write profiles from the VAK comparison, the small purposive qualitative subsample, and the immediate posttest. In addition, each MCT dimension was represented by one constructed-response item, so dimension-level findings should remain item-specific. These constraints limit causal inference, population generalization, and the strength of claims about individual MCT dimensions. Future studies could use controlled or randomized designs, broader learner profiles, delayed retention measures, and multi-item measurement of each MCT dimension. Equivalence testing or Bayesian analysis would provide a stronger basis for interpreting non-significant group differences, while process tracing, screen recording, and direct measures of self-regulation or cognitive load could test the proposed process explanations more directly.

6. CONCLUSION

This study provides a multidimensional account of mathematical critical thinking (MCT) within an e-module-supported Project-Based Learning (PjBL) environment. Overall MCT scores and all six item-level dimension indicators were higher at posttest than at pretest. Posttest performance and normalized gain did not differ significantly among students classified by Visual, Auditory, and Kinaesthetic self-reported preferences, and baseline-adjusted analysis yielded the same inferential conclusion. Qualitative evidence showed that MCT was expressed through diverse but overlapping forms of engagement, including visualization, verbalization, digital manipulation, trial-and-check reasoning, peer discussion, and reflection. Similar observed outcomes across preference groups therefore coexisted with heterogeneous reasoning processes.

By integrating outcome- and process-level evidence, the study shows how a common multimodal instructional environment can be associated with marked pretest-posttest change while supporting varied forms of mathematical reasoning. The findings are consistent with providing all students opportunities to interpret, analyze, evaluate, infer, justify, and monitor reasoning rather than rigidly matching instruction to VAK categories. Because the quantitative strand used a single-group pretest-posttest design, the observed changes should not be interpreted causally. Further research with controlled designs, broader learner profiles, delayed assessments, and process-sensitive measures is needed to evaluate mechanisms, transferability, and persistence of the observed patterns.

ACKNOWLEDGMENTS

The authors would like to express their deepest gratitude to the Semarang State University, participants and colleagues for their exceptional support and resources, which were critical in facilitating this research.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests or conflicts of interest related to this study.

DATA AVAILABILITY

The datasets utilized and analyzed during this research are available from the corresponding author upon reasonable request.

ETHICAL STATEMENTS

The authors confirm that the study complied with all applicable local laws, ethical standards, and institutional guidelines, including obtaining approval from relevant ethics committees.

FUNDING

This research was conducted without financial support. The authors confirm that no funding was received for this study's research, analysis, or publication.

REFERENCES

[1] P. C. Abrami, R. M. Bernard, E. Borokhovski, D. I. Waddington, C. A. Wade, and T. Persson, “Strategies for Teaching Students to Think Critically: A Meta-Analysis,” Rev. Educ. Res., vol. 85, no. 2, pp. 275–314, 2015, https://doi.org/10.3102/0034654314551063

[2] C. Monteleone, J. Miller, and E. Warren, “Conceptualising critical mathematical thinking in young students,” Math. Educ. Res. J., vol. 35, no. 2, pp. 339–359, 2023, https://doi.org/10.1007/s13394-023-00445-1

[3] Q. Wang and A. H. Abdullah, “Enhancing Students’ Critical Thinking Through Mathematics in Higher Education: A Systemic Review,” SAGE Open, vol. 14, no. 3, 2024, https://doi.org/10.1177/21582440241275651

[4] OECD, “PISA 2022 Results (Volume I): The State of Learning and Equity in Education,” OECD Publishing, Paris, 2023. https://doi.org/10.1787/53f23881-en.

[5] A. Putri, T. Nusantara, Purwanto, and A. R. As’ari, “The contribution of critical thinking skills in rich mathematical problem completion: Insights from pre-service mathematics teachers,” Eurasia J. Math. Sci. Technol. Educ., vol. 21, no. 2, p. em2581, 2025, https://doi.org/10.29333/ejmste/15931

[6] K. L. Yang, C. H. Chen, and W. R. Wu, “Exploring critical and reflective thinking in mathematical literacy through variable- and person-centered approaches,” Think. Ski. Creat., vol. 60, p. 102127, 2026, https://doi.org/10.1016/j.tsc.2026.102127

[7] D. Kamber and D. Takaci, “On problematic aspects in learning trigonometry,” Int. J. Math. Educ. Sci. Technol., vol. 49, no. 2, pp. 161–175, 2018, https://doi.org/10.1080/0020739X.2017.1357846

[8] B. H. Ngu and H. P. Phan, “Differential instructional effectiveness: overcoming the challenge of learning to solve trigonometry problems that involved algebraic transformation skills,” Eur. J. Psychol. Educ., vol. 38, no. 4, pp. 1505–1525, 2023, https://doi.org/10.1007/s10212-022-00670-5

[9] A. T. Ferede, M. A. Ayele, A. A. Mihrka, and A. A. Arara, “The impact of contextualized teaching and learning approach on students’ conceptual understanding of trigonometry,” Teach. Math. its Appl., vol. 45, no. 2, pp. 161–180, 2026, https://doi.org/10.1093/teamat/hraf008

[10] D. Kokotsaki, V. Menzies, and A. Wiggins, “Project-based learning: A review of the literature,” Improv. Sch., vol. 19, no. 3, pp. 267–277, 2016. https://doi.org/10.1177/1365480216659733

[11] C. H. Chen and Y. C. Yang, “Revisiting the effects of project-based learning on students’ academic achievement: A meta-analysis investigating moderators,” Educ. Res. Rev., vol. 26, pp. 71–81, 2019, https://doi.org/10.1016/j.edurev.2018.11.001

[12] S. Farshad and C. Fortin, “An umbrella review of meta-analyses on project-based learning: Effects on academic achievement, higher-order thinking, and 21st-century skills,” Educ. Res. Rev., vol. 52, p. 100809, 2026, https://doi.org/10.1016/j.edurev.2026.100809

[13] J. Young, “Technology-enhanced mathematics instruction: A second-order meta-analysis of 30 years of research,” Educ. Res. Rev., vol. 22, pp. 19–33, 2017, https://doi.org/10.1016/j.edurev.2017.07.001

[14] C. R. Rakes, R. N. Ronau, S. B. Bush, S. O. Driskell, M. L. Niess, and D. K. Pugalee, “Mathematics achievement and orientation: A systematic review and meta-analysis of education technology,” Educ. Res. Rev., vol. 31, p. 100337, 2020, https://doi.org/10.1016/j.edurev.2020.100337

[15] O. Birgin and K. Uzun Yazıcı, “The effect of GeoGebra software–supported mathematics instruction on eighth-grade students’ conceptual understanding and retention,” J. Comput. Assist. Learn., vol. 37, no. 4, pp. 925–939, 2021, https://doi.org/10.1111/jcal.12532

[16] A. M. K. Mannan, M. R. Bajuri, S. S. Abdul Rahim, and N. A. Abdullah, “How do digital tools support metacognition in mathematics learning? A systematic review of research on GeoGebra and problem-solving,” Educ. Res.

Rev., vol. 52, p. 100827, 2026, https://doi.org/10.1016/j.edurev.2026.100827

[17] A. Skulmowski and K. M. Xu, "Understanding Cognitive Load in Digital and Online Learning: a New Perspective on Extraneous Cognitive Load," Educ. Psychol. Rev., vol. 34, no. 1, pp. 171–196, 2022, https://doi.org/10.1007/s10648-021-09624-7

[18] H. Pashler, M. McDaniel, D. Rohrer, and R. Bjork, "Learning styles concepts and evidence," Psychol. Sci. Public Interes. Suppl., vol. 9, no. 3, pp. 105–119, 2008, https://doi.org/10.1111/j.1539-6053.2009.01038.x

[19] K. Aslaksen and H. Lorås, "The modality-specific learning style hypothesis: A mini-review," Front. Psychol., vol. 9, no. Aug, p. 1538, 2018, https://doi.org/10.3389/fpsyg.2018.01538

[20] T. Fallace, "The Long Origins of the Visual, Auditory, and Kinesthetic Learning Style Typology, 1921–2001," Hist. Psychol., vol. 26, no. 4, pp. 334–354, 2023, https://doi.org/10.1037/hop0000240

[21] M. D. Fetters, L. A. Curry, and J. W. Creswell, "Achieving integration in mixed methods designs - Principles and practices," Health Serv. Res., vol. 48, no. 6 PART2, pp. 2134–2156, 2013, https://doi.org/10.1111/1475-6773.12117

[22] P. A. Facione, "Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations," American Philosophical Association, Newark, DE, 1990. [Online]. Available: https://eric.ed.gov/?id=ED315423

[23] A. H. Schoenfeld, "Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics," in Handbook for Research on Mathematics Teaching and Learning, D. A. Grouws, Ed. New York, NY: Macmillan, 1992, pp. 334–370.

[24] J. Lithner, "A research framework for creative and imitative reasoning," Educ. Stud. Math., vol. 67, pp. 255–276, 2008, https://doi.org/10.1007/s10649-007-9104-2

[25] S. Dolapcioglu and A. Doğanay, "Development of critical thinking in mathematics classes via authentic learning: an action research," Int. J. Math. Educ. Sci. Technol., vol. 53, no. 6, pp. 1363–1386, 2022, https://doi.org/10.1080/0020739X.2020.1819573

[26] N. Monrat, M. Phaksunchai, and R. Chonchaiya, "Developing Students' Mathematical Critical Thinking Skills Using Open-Ended Questions and Activities Based on Student Learning Preferences," Educ. Res. Int., vol. 2022, p. 3300363, 2022, https://doi.org/10.1155/2022/3300363

[27] P. C. Blumenfeld, E. Soloway, R. W. Marx, J. S. Krajcik, M. Guzdial, and A. Palincsar, "Motivating project-based learning: Sustaining the doing, supporting the learning," Educ. Psychol., vol. 26, no. 3–4, pp. 369–398, 1991, https://doi.org/10.1080/00461520.1991.9653139

[28] R. E. Mayer, Multimedia Learning, 3rd ed. Cambridge, UK: Cambridge University Press, 2020. https://doi.org/10.1017/9781316941355

[29] R. E. Mayer, "Cognitive Theory of Multimedia Learning," in The Cambridge Handbook of Multimedia Learning, 3rd ed., R. E. Mayer and L. Fiorella, Eds. Cambridge University Press, 2021, pp. 57–72. https://doi.org/10.1017/9781108894333.008

[30] L. Fiorella and R. E. Mayer, "Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles," in The Cambridge Handbook of Multimedia Learning, 3rd ed., R. E. Mayer and L. Fiorella, Eds. Cambridge, UK: Cambridge University Press, 2021, pp. 185–198. https://doi.org/10.1017/9781108894333.019

[31] R. C. Clark, "Multimedia learning in e-courses," in The Cambridge Handbook of Multimedia Learning, 3rd ed., R. E. Mayer and L. Fiorella, Eds. Cambridge, UK: Cambridge University Press, 2021, pp. 537–551. https://doi.org/10.1017/9781108894333.054

[32] R. Moreno and R. E. Mayer, "Interactive multimodal learning environments," Educ. Psychol. Rev., vol. 19, pp. 309–326, 2007, https://doi.org/10.1007/s10648-007-9047-2

[33] F. M. Van der Kleij, R. C. W. Feskens, and T. J. H. M. Eggen, "Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis," Rev. Educ. Res., vol. 85, no. 4, pp. 475–511, 2015, https://doi.org/10.3102/0034654314564881

[34] R. Azevedo, "Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning," Educ. Psychol., vol. 40, no. 4, pp. 199–209, 2005, https://doi.org/10.1207/s15326985ep4004_2

[35] R. Azevedo and J. G. Cromley, "Does training on self-regulated learning facilitate students' learning with hypermedia?," J. Educ. Psychol., vol. 96, no. 3, pp. 523–535, 2004, https://doi.org/10.1037/0022-0663.96.3.523

[36] F. Reinhold, T. Leuders, K. Loibl, M. Nückles, M. Beege, and J. M. Boelmann, "Learning Mechanisms Explaining Learning With Digital Tools in Educational Settings: a Cognitive Process Framework," Educ. Psychol. Rev., vol. 36, no. 1, p. 14, 2024, https://doi.org/10.1007/s10648-024-09845-6

[37] L. Brummer, H. de Boer, J. M. Mouw, and J. W. Strijbos, "A meta-analysis of the effects of context, content, and task factors of digitally delivered instructional feedback on learning performance," Learn. Environ. Res., vol. 27, no. 3, pp. 453–476, 2024, https://doi.org/10.1007/s10984-024-09501-4

[38] R. E. Mayer, "The Past, Present, and Future of the Cognitive Theory of Multimedia Learning," Educ. Psychol. Rev., vol. 36, no. 1, p. 8, 2024, https://doi.org/10.1007/s10648-023-09842-1

[39] B. A. Rogowsky, B. M. Calhoun, and P. Tallal, "Matching learning style to instructional method: Effects on comprehension," J. Educ. Psychol., vol. 107, no. 1, pp. 64–78, 2015, https://doi.org/10.1037/a0037478

[40] J. W. Creswell and V. L. Plano Clark, Designing and Conducting Mixed Methods Research, 3rd ed. Thousand Oaks, CA: SAGE Publications, 2018.

[41] N. V. Ivankova, J. W. Creswell, and S. L. Stick, "Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice," Field methods, vol. 18, no. 1, pp. 3–20, 2006, https://doi.org/10.1177/1525822X05282260

[42] W. R. Shadish, T. D. Cook, and D. T. Campbell, Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA: Houghton Mifflin, 2002.

[43] L. A. Palinkas, S. M. Horwitz, C. A. Green, J. P. Wisdom, N. Duan, and K. Hoagwood, "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research," Adm. Policy Ment.

Heal. Ment. Heal. Serv. Res., vol. 42, no. 5, pp. 533–544, 2015, https://doi.org/10.1007/s10488-013-0528-y
[44] R. M. Branch, Instructional Design: The ADDIE Approach. New York, NY: Springer, 2009. https://doi.org/10.1007/978-0-387-09506-6
[45] W. L. Leite, M. Svinicki, and Y. Shi, "Attempted validation of the scores of the VARK: Learning Styles Inventory with multitrait–multimethod confirmatory factor analysis models," Educ. Psychol. Meas., vol. 70, no. 2, pp. 323–339, 2010, https://doi.org/10.1177/0013164409344507
[46] K. Warner, M. Hamza, A. Oliver-Smith, F. Renaud, and A. Julca, "Climate change, environmental degradation and migration," Nat. Hazards, vol. 55, no. 3, pp. 689–715, 2010, https://doi.org/10.1007/s11069-009-9419-7
[47] D. F. Polit and C. T. Beck, "The content validity index: Are you sure you know what's being reported? Critique and recommendations," Res. Nurs. Heal., vol. 29, no. 5, pp. 489–497, 2006, https://doi.org/10.1002/nur.20147
[48] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," Int. J. Med. Educ., vol. 2, pp. 53–55, 2011, https://doi.org/10.5116/ijme.4dfb.8dfd
[49] S. Sireci and M. Faulkner-Bond, "Validity evidence based on test content," Psicothema, vol. 1, no. 26, pp. 100–107, 2014, https://doi.org/10.7334/psicothema2013.256
[50] S. Messick, "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning," Am. Psychol., vol. 50, no. 9, pp. 741–749, 1995, https://doi.org/10.1037/0003-066X.50.9.741
[51] P. A. Facione, "Critical Thinking : A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction Executive Summary " The Delphi Report," Calif. Acad. Press, vol. 423, no. c, pp. 1–19, 1990, [Online]. Available: http://www.insightassessment.com/pdf_files/DEXadobe.PDF
[52] A. Jönsson and G. Svingby, "The use of scoring rubrics: Reliability, validity and educational consequences," Educ. Res. Rev., vol. 2, no. 2, pp. 130–144, 2007, https://doi.org/10.1016/j.edurev.2007.05.002
[53] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," Psychol. Bull., vol. 86, no. 2, pp. 420–428, 1979, https://doi.org/10.1037/0033-2909.86.2.420
[54] K. A. Hallgren, "Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial," Tutor. Quant. Methods Psychol., vol. 8, no. 1, pp. 23–34, 2012, https://doi.org/10.20982/tqmp.08.1.p023
[55] K. O. McGraw and S. P. Wong, "Forming Inferences about Some Intraclass Correlation Coefficients," Psychol. Methods, vol. 1, no. 1, pp. 30–46, 1996, https://doi.org/10.1037/1082-989X.1.1.30
[56] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," J. Chiropr. Med., vol. 15, no. 2, pp. 155–163, 2016, https://doi.org/10.1016/j.jcm.2016.02.012
[57] J. P. G. Kasakewitch, H. G. B. Coelho, and B. B. Lima, "Mixed methods research," Transl. Cardiol., vol. 12, no. 3, pp. 469–473, 2025, https://doi.org/10.1016/B978-0-323-91790-2.00070-8
[58] V. L. Plano Clark, "Mixed methods research," J. Posit. Psychol., vol. 12, no. 3, pp. 305–306, 2017, https://doi.org/10.1080/17439760.2016.1262619
[59] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," Front. Psychol., vol. 4, no. NOV, p. 62627, 2013, https://doi.org/10.3389/fpsyg.2013.00863
[60] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," J. R. Stat. Soc. Ser. B Stat. Methodol., vol. 57, no. 1, pp. 289–300, 1995, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
[61] Syaiful, N. Huda, A. Mukminin, and Kamid, "Using a metacognitive learning approach to enhance students' critical thinking skills through mathematics education," SN Soc. Sci., vol. 2, no. 4, p. 31, 2022, https://doi.org/10.1007/s43545-022-00325-8
[62] C. E. Hmelo-Silver, "Problem-based learning: What and how do students learn?," Educ. Psychol. Rev., vol. 16, no. 3, pp. 235–266, 2004, https://doi.org/10.1023/B:EDPR.0000034022.16470.f3
[63] T. Tafakur, H. Retnawati, and A. A. M. Shukri, "Effectiveness of project-based learning for enhancing students critical thinking skills: A meta-analysis," JINoP (Jurnal Inov. Pembelajaran), vol. 9, no. 2, pp. 191–209, 2023, https://doi.org/10.22219/jinop.v9i2.22142
[64] S. M. M. Loyens, J. E. van Meerten, L. Schaap, and L. Wijnia, "Situating higher-order, critical, and critical-analytic thinking in problem- and project-based learning environments: A systematic review," Educ. Psychol. Rev., vol. 35, p. 39, 2023, https://doi.org/10.1007/s10648-023-09757-x
[65] P. Guo, N. Saab, L. S. Post, and W. Admiraal, "A review of project-based learning in higher education: Student outcomes and measures," Int. J. Educ. Res., vol. 102, p. 101586, 2020, https://doi.org/10.1016/j.ijer.2020.101586
[66] E. Rexigel, J. Kuhn, S. Becker, and S. Malone, "The more the better? A systematic review and meta-analysis of the benefits of more than two external representations in STEM education," Educ. Psychol. Rev., vol. 36, p. 124, 2024, https://doi.org/10.1007/s10648-024-09958-y
[67] Z. Ji, K. Guo, and S. Song, "Effects of dynamic mathematical software on students' performance: A three-level meta-analysis," J. Educ. Comput. Res., vol. 62, no. 4, pp. 815–840, 2024, https://doi.org/10.1177/07356331241226594
[68] Y. B. Zhang, P. Wang, W. Jia, A. Zhang, and G. Chen, "Dynamic visualization by GeoGebra for mathematics learning: A meta-analysis of 20 years of research," J. Res. Technol. Educ., vol. 57, no. 2, pp. 437–458, 2025, https://doi.org/10.1080/15391523.2023.2250886
[69] J. Shao, Y. Chen, X. Wei, X. Li, and Y. Li, "Effects of regulated learning scaffolding on regulation strategies and academic performance: A meta-analysis," Front. Psychol., vol. 14, p. 1110086, 2023, https://doi.org/10.3389/fpsyg.2023.1110086
[70] V. Clinton-Lisell and C. Litzinger, "Is it really a neuromyth? A meta-analysis of the learning styles matching hypothesis," Front. Psychol., vol. 15, p. 1428732, 2024, https://doi.org/10.3389/fpsyg.2024.1428732
[71] L. Melzner and C. Kappes, "Testing the meshing hypothesis in prospective teachers: Are there effects of

matching learning style and presentation mode on learning performance and on metacognitive aspects of learning?," Instr. Sci., vol. 53, pp. 365–389, 2025, https://doi.org/10.1007/s11251-024-09689-1

[72] M. Peters and S. Fàbregues, "Missed opportunities in mixed methods EdTech research? Visual joint display development as an analytical strategy for achieving integration in mixed methods studies," Educ. Technol. Res. Dev., vol. 72, pp. 2477–2497, 2024, https://doi.org/10.1007/s11423-023-10234-z

[73] D. Juandi, S. Suparman, A. D. Putri, J. A. Dahlan, and D. Dasari, "The utilization of computer-assisted mathematics education integrated into project-based learning to optimize Indonesian students' mathematics achievement: A systematic review and meta-analysis," Int. Electron. J. Math. Educ., vol. 20, no. 4, p. em0853, 2025, https://doi.org/10.29333/iejme/16899

SUPPLEMENTARY

Table S1. Operational Implementation of the Six E-Module-Assisted PjBL Meetings

Meet	Duration	PjBL Phase	Teacher Activities	Student Activities	Trigonometry Project	E-Module Features
1	90 min	Problem Orientation	Introduces the contextual trigonometry problem; explains project objectives, task requirements, and expected project outcomes; facilitates initial questioning and problem identification.	Observe the contextual problem, identify relevant mathematical information, formulate questions, and identify the trigonometric concepts required for the project.	Initial project problem formulation and identification of the mathematical problem to be investigated.	Interactive instructional video, project introduction, learning objectives, and contextual trigonometry materials.
2	90 min	Project Planning	Guides students in developing the project plan; facilitates group discussion; provides feedback on the feasibility of proposed strategies and task distribution.	Develop a project plan, formulate investigation steps, distribute group responsibilities, determine required data, and select appropriate mathematical strategies.	Project plan and digital worksheet containing investigation procedures, task allocation, required data, and proposed mathematical strategies.	Digital project worksheet, planning prompts, trigonometry learning materials, and guided activities.
3	90 min	Investigation	Facilitates data exploration and mathematical investigation; monitors group progress; provides guiding questions and feedback without directly providing solutions.	Collect and examine relevant information, explore mathematical relationships, conduct calculations, test representations, and investigate possible solutions.	Investigation results and mathematical evidence, including calculations, representations, and exploration results.	GeoGebra-based dynamic simulations, instructional explanations, interactive representations, and supporting trigonometry materials.
4	90 min	Product Development	Guides students in constructing and evaluating the project solution; checks the mathematical validity of students' work; provides formative feedback.	Construct the mathematical solution/model, apply the selected strategy, test the solution, identify errors, revise the work, and prepare the project result.	Completed mathematical project solution/model based on the investigation results.	Interactive exercises, formative tests with feedback, digital worksheets, and mathematical representations.
5	90 min	Presentation	Facilitates group presentations; asks questions concerning mathematical reasoning; provides feedback on the accuracy and justification of the project results.	Present the project results, explain the mathematical reasoning, interpret representations, answer questions, and justify the selected solution.	Presented project result and digital project portfolio containing the group's mathematical work and explanation.	Online portfolio submission, presentation-support materials, project documentation, and digital learning resources.
6	90 min	Reflection and Evaluation	Facilitates reflection and peer evaluation; guides students to evaluate the project process and outcomes; provides final feedback.	Evaluate the project process and outcome, conduct peer evaluation, identify strengths and weaknesses, revise conclusions where necessary, and reflect on their mathematical reasoning.	Final revised project portfolio and individual reflection/self-assessment.	Self-assessment quiz, reflection activities, formative evaluation, and project portfolio.

Table S2. Analytic Scoring Rubric for the Mathematical Critical Thinking Test

MCT Dimension	4 – Excellent	3 – Good	2 – Fair	1 – Poor	0 – No Evidence
Interpretation	Accurately identifies, selects, and interprets all relevant mathematical information, conditions, symbols, and representations.	Identifies and interprets most `relevant information correctly, with only minor omissions or inaccuracies.	Identifies some relevant information but interpretation is incomplete or contains several inaccuracies.	Identifies very limited information and shows substantial misunderstanding of the mathematical situation.	Does not identify or interpret relevant mathematical information.

MCT Dimension	4 – Excellent	3 – Good	2 – Fair	1 – Poor	0 – No Evidence
Analysis	Clearly identifies mathematical relationships, assumptions, and components and develops an appropriate and logically coherent strategy.	Identifies the main mathematical relationships and develops an appropriate strategy with minor weaknesses.	Identifies some relationships but the strategy is incomplete, partially appropriate, or insufficiently justified.	Shows limited recognition of mathematical relationships and uses an inappropriate or poorly developed strategy.	Provides no meaningful analysis or strategy.
Evaluation	Systematically examines the validity of procedures, mathematical arguments, representations, and results and provides accurate justification.	Evaluates the main procedure or result correctly and provides generally appropriate justification.	Performs partial evaluation but misses important aspects or provides limited justification.	Provides a superficial or largely unsupported evaluation and fails to identify important errors.	Does not evaluate the procedure, argument, representation, or result.
Inference	Draws a mathematically correct, logical, and well-supported conclusion from relevant evidence and relationships.	Draws a correct conclusion supported by relevant evidence, with minor gaps in reasoning.	Draws a partially correct conclusion but the supporting reasoning or evidence is incomplete.	Draws an unclear, unsupported, or substantially incorrect conclusion.	Provides no conclusion or inference.
Explanation	Clearly communicates the mathematical reasoning, procedures, evidence, and conclusion using accurate mathematical language and representations.	Communicates the reasoning and conclusion clearly, with minor omissions or imprecision.	Provides a partially understandable explanation but important reasoning steps are missing or unclear.	Provides a limited or difficult-to-follow explanation with substantial gaps in reasoning.	Provides no meaningful mathematical explanation.
Self-Regulation	Systematically monitors, checks, and evaluates the reasoning; identifies errors or limitations and appropriately revises the solution.	Checks the reasoning and result and identifies or corrects most errors with minor limitations.	Performs some checking but monitoring is incomplete and revision is limited.	Shows minimal checking or monitoring and rarely identifies or corrects errors.	Provides no evidence of monitoring, checking, evaluation, or revision.

Table S3. Inter-Rater Reliability of the Mathematical Critical Thinking Scores

MCT dimension	Pre-test ICC (2,1)	95% CI	Interpretation	Post-test ICC (2,1)	95% CI	Interpretation
Interpretation	0.840	[0.756, 0.897]	Good	0.922	[0.878, 0.950]	Excellent
Analysis	0.846	[0.765, 0.901]	Good	0.923	[0.880, 0.951]	Excellent
Evaluation	0.848	[0.768, 0.902]	Good	0.895	[0.837, 0.933]	Good
Inference	0.866	[0.794, 0.914]	Good	0.920	[0.875, 0.949]	Excellent
Explanation	0.818	[0.724, 0.882]	Good	0.892	[0.833, 0.931]	Good
Self-regulation	0.878	[0.812, 0.922]	Good	0.924	[0.881, 0.952]	Excellent
Total MCT	0.947	[0.917, 0.967]	Excellent	0.944	[0.912, 0.965]	Excellent

Table S4. ANCOVA of Posttest Mathematical Critical Thinking by VAK Preference, Controlling for Pretest Performance

Source	Sum of squares	df	Mean square	F	p	Partial η^2
Pretest MCT	0.731	1	0.731	0.021	0.885	< 0.001
VAK preference	110.476	2	55.238	1.600	0.209	0.045
Error	2,347.439	68	34.521	-	-	-
Corrected total	2,458.646	71	-	-	-	-

Table S5. Pretest-Posttest Change in MCT Dimensions Across VAK Preference Groups

MCT dimension	Maximum transformed	Visual (n = 24) Mean (SD)	Auditory (n = 24) Mean (SD)	Kinaesthetic (n = 24) Mean (SD)	F(2,69)	FDR-adjusted p
Interpretation	16.67	5.19 (2.28)	5.67 (2.44)	6.37 (2.08)	1.627	0.713
Analysis	16.67	5.87 (2.26)	6.10 (2.59)	5.54 (3.02)	0.266	0.972
Evaluation	16.67	5.79 (1.71)	5.62 (1.91)	5.88 (2.40)	0.100	0.972
Inference	16.67	5.25 (2.16)	5.97 (2.30)	5.47 (2.31)	0.642	0.972

MCT dimension	Maximum transformed	Visual (n = 24) Mean (SD)	Auditory (n = 24) Mean (SD)	Kinaesthetic (n = 24) Mean (SD)	F(2,69)	FDR-adjusted p
Explanation	16.67	6.06 (2.29)	6.02 (2.31)	5.92 (1.92)	0.029	0.972
Self-regulation	16.67	6.10 (2.07)	5.32 (2.38)	4.93 (2.72)	1.468	0.713

Table S6. Expanded Joint Display Integrating Quantitative and Qualitative Findings

MCT outcome / quantitative evidence	Qualitative process evidence	Relationship between strands	Integrated interpretation
Overall MCT: 43.95 (SD = 6.26) -> 78.18 (SD = 5.77); mean change = 34.23; dz = 4.08; p < 0.001.	Across interviews, observations, and artifacts, students increasingly identified relevant information, connected mathematical relationships, tested strategies, communicated reasoning, and monitored solutions.	Convergence	The observed score increase was accompanied by more explicit enactment of critical-thinking processes during mathematical problem solving.
Interpretation: 7.64 -> 13.52; change = 5.88; FDR-adjusted p < 0.001.	Dynamic graphs, diagrams, and the unit-circle representation were used to connect symbolic identities with spatial representations.	Convergence + complementarity	Improvement in interpretation was accompanied by stronger coordination of symbolic and visual information.
Analysis: 7.21 -> 13.08; change = 5.87; FDR-adjusted p < 0.001.	Students identified relationships, selected strategies, tested substitutions, and revised unsuccessful approaches using scratchpads and peer discussion.	Convergence + complementarity	Analysis was expressed as an iterative process of relationship identification, strategy construction, testing, and revision.
Evaluation: 6.85 -> 12.38; change = 5.53; FDR-adjusted p < 0.001.	Students checked transformations and compared procedures, although some required prompts to verify validity or domain restrictions.	Convergence + qualification	Evaluation improved, but verification was not uniformly spontaneous or independent.
Inference: 7.42 -> 13.19; change = 5.77; FDR-adjusted p < 0.001.	Students tested mathematical relationships and alternative transformations before finalizing conclusions.	Convergence + complementarity	Inference was expressed through relationship identification, testing, evidence checking, and conclusion formation.
Explanation: 7.50 -> 13.01; change = 5.51; FDR-adjusted p < 0.001.	Verbalization, peer dialogue, and translation between visual and symbolic representations were used to articulate reasoning.	Convergence + complementarity	Explanation involved making mathematical relationships explicit and defensible rather than merely stating final answers.
Self-regulation: 7.33 -> 13.00; change = 5.67; FDR-adjusted p < 0.001.	Reflection prompts supported error detection, procedural monitoring, domain checking, and revision of solution strategies.	Convergence + complementarity	Self-regulation operated across checking, correction, revision, and reflection.
VAK comparison: no significant posttest or gain differences across groups.	Visual, Auditory, and Kinaesthetic groups showed different recurring emphases in visualization, verbalization, manipulation, and trial-and-check, but these processes overlapped across categories.	Complementarity + qualification	Similar MCT outcomes coexisted with variation in how reasoning processes were expressed.
Cross-source qualitative pattern.	Interviews, observations, and artifacts differed in the visibility of particular reasoning processes.	Convergence + divergence	No single qualitative source fully represented students' reasoning; triangulation added substantive information.
Integrated outcome.	Students engaged with the common PjBL-e-module environment through different combinations of representation, verbalization, manipulation, collaboration, checking, and reflection.	Overall convergence with process differentiation	Outcome-level similarity across preference groups coexisted with heterogeneous process expression.

Table S7. Representative Qualitative Evidence for MCT Process Patterns

Process pattern	Representative participant evidence	Participant / source	Analytical interpretation
Visualization	"When I opened the dynamic unit circle in the E-Module and saw the right triangle transform visually ... the identity immediately made sense."	P02-V; INT-02, OBS-S3	Dynamic representation supported coordination between symbolic and spatial information.
Verbalization	"Verbalizing the logic helped me spot errors in my own reasoning."	P06-A; INT-06, OBS-S2	Spoken reasoning supported sequencing and error detection.
Digital manipulation	"Being able to drag the angle slider directly on the screen and watch the trigonometric function graph stretch in real-time let me test how period transformations actually behave."	P09-K; INT-09, ART-S4	Parameter manipulation supported exploration of functional relationships.

Process pattern	Representative participant evidence	Participant / source	Analytical interpretation
Trial-and-check	“If an approach collapsed into an unsolvable expression, I checked where it broke down and restarted.”	P10-K; INT-10, OBS-S3	Iterative testing supported strategy evaluation and revision.
Peer discussion	“Combining our methods made the final solution much tighter.”	P05-A; INT-05, OBS-S5	Peer exchange supported comparison and justification of alternative representations and strategies.
Reflection	“I realized I kept forgetting domain restrictions for tangent functions, so I added a checking step to my habit routine.”	P11-K; INT-11, ART-S5	Structured reflection supported monitoring and self-correction.

Note. Participant codes identify preference group: V = Visual, A = Auditory, K = Kinaesthetic. INT = interview transcript; OBS = classroom observation; ART = student artifact.

Table S8. Cross-Source Evidence Coverage for Qualitative MCT Process Patterns

Process pattern	Documented source tag(s)	Source coverage	Evidence interpretation
Visualization	INT-02; OBS-S3	Interview + observation	Dynamic representation supported coordination between symbolic and spatial information.
Verbalization	INT-06; OBS-S2	Interview + observation	Spoken reasoning supported sequencing and error detection.
Digital manipulation	INT-09; ART-S4	Interview + artifact	Parameter manipulation supported exploration of functional relationships.
Trial-and-check	INT-10; OBS-S3	Interview + observation	Iterative testing supported strategy evaluation and revision.
Peer discussion	INT-05; OBS-S5	Interview + observation	Peer exchange supported comparison and justification of alternative strategies.
Reflection	INT-11; ART-S5	Interview + artifact	Structured reflection supported monitoring and self-correction.
Cross-category overlap / discrepant cases	Cross-source synthesis	Interview + observation + artifact	Reasoning modes crossed nominal VAK categories, and different sources captured different aspects of the same reasoning processes.