

Supporting critical thinking in sequences and series through a problem-solving e-module with algebraic tasks

Nasrun^{1*}, Irwan Akib¹, Muhammad Ikram², Muhammad Muzaini¹, Rahmat Kamaruddin³

¹Department of Mathematics Education, Universitas Muhammadiyah Makassar, South Sulawesi, Indonesia

²Department of Mathematics, Universitas Negeri Makassar, South Sulawesi, Indonesia

³Department of Mathematics Education, STKIP Andi Matappa Pangkep, South Sulawesi, Indonesia

*Correspondence: nasrun.anthy@unismuh.ac.id

Received: May 11, 2026 | Revised: Jul 25, 2026 | Accepted: Aug 26, 2026 | Published Online: Sep 25, 2026

Abstract

Examining sequences and series often leads students to rely on formulas, especially when assignments require algebraic modeling and generalization. This condition may hinder the advancement of critical thinking, especially in evaluation, inference, and reasoning. This study examined the extent to which a problem-solving e-module with algebraic challenges enhanced students' critical thinking and identified the elements that best facilitated critical thinking during the learning process. The study used a sequential explanatory mixed-methods design. The quantitative phase encompassed 89 senior high school students from three classes. Students completed a pretest, engaged with an e-module featuring guiding questions, scaffolding, feedback, and reflection, and then completed a posttest. The quantitative data were examined utilizing descriptive statistics, normalized gain, pretest-posttest comparisons, and effect size analysis. Data were gathered from students' e-module assignments, classroom observations, written reflections, and task-oriented interviews. The data were examined using theme coding and synthesized with the quantitative results. The results indicated an improvement in overall critical thinking and across all metrics; however, the improvements were inconsistent. Evaluation and interpretation improved more significantly than explanation. Feedback, directed inquiries, scaffolding, and reflection facilitated critical thinking; however, argument quality and algebraic reasoning varied among students.

Keywords:

Algebraic tasks, Critical thinking, E-module, Problem-solving, Sequences and series

How to Cite:

Nasrun, N., Akib, I., Ikram, M., Muzaini, M., & Kamaruddin, R. (2026). Supporting critical thinking in sequences and series through a problem-solving e-module with algebraic tasks. *Infinity Journal*, 15(4), 1031-1060. <https://doi.org/10.22460/infinity.v15i4.p1031-1060>

This is an open access article under the CC BY-SA license.



1. INTRODUCTION

Students sometimes have difficulties with sequences and series when assignments require algebraic modeling (Kertil et al., 2019; Lee & Lee, 2021). A significant number of students can extend a pattern for the initial terms but find it challenging to accurately formulate the n th-term expression (Patterson & McGraw, 2018; Susanta et al., 2023; Wilkie, 2020).

Students frequently confuse recursive rules with explicit forms, especially when assignments require conversion between the two (Sandefur & Manaster, 2022). In contextual concerns, students often misconstrue the given information, resulting in the creation of algebraic relationships that do not align with the actual scenario (Scheiner & Pinto, 2019). Errors in algebraic manipulation occur when students simplify expressions, formulate equations, or manage parameters (Soneira et al., 2018). Mathematical nomenclature, encompassing sigma notation and indices, often leads to inaccuracies (Yoon & Thompson, 2020). These problems undermine the quality of students' solutions and hinder their ability to provide logical justification.

Previous studies suggest that difficulties in sequences and series often arise from problems associated with generalization (Yao, 2022) and representation (Elia, 2020). Research on secondary school students shown heterogeneity in their levels of representation while articulating sequences, ranging from unstructured to fully ordered forms, and indicated that generalization strategies changed when tasks required extensive generalization (Yao, 2022). In close generalization, students mostly employed counting and addition operations, whereas in far generalization, strategies diversified, with correspondence methods emerging as the most common (Melhuish et al., 2020). Further research has highlighted pedagogical obstacles and the gap between knowledge obtained in secondary education and that necessary for higher education, leading students to depend on rote memorization of formulas and struggle with the transition between empirical and theoretical frameworks (Kop et al., 2020). A quasi-experimental study on numerical patterns in Singapore revealed that computational activity-based interventions did not consistently produce significant overall improvement; however, some students demonstrated substantial progress, suggesting that learning outcomes are greatly affected by the design of activities and the quality of instructional support (Vicente et al., 2022). These findings indicate that education should emphasize how students develop overarching principles, select strategies, and verify outcomes, rather than only rehearsing the application of formulas.

These issues have necessitated the implementation of teaching practices designed to improve critical thinking. Problem-solving instruction enables students to understand obstacles, formulate strategies, execute those methods, and assess their results (Voica et al., 2020). Problem-based (Xin et al., 2020) and inquiry-based learning (Pedersen & Haavold, 2023) require students to explain rationales, evaluate facts, and justify their strategic choices. In educational contexts, leading questions are often utilized to stimulate interpretation, analysis, and reasoning rather than just obtaining definitive answers (Foster et al., 2022). Scaffolding and metacognitive queries are employed to aid students in tracking their progress, recognizing faults, and enhancing their strategies (Çakmak Gürel, 2025). Delivering this support consistently, however, is difficult in conventional media. Printed worksheets fix a single solution path and cannot adapt prompts to a student's response; whole-class discussion distributes scaffolding unevenly and leaves little trace of individual reasoning; and video or static presentations convey content without eliciting or recording students' own justifications. An e-module addresses these constraints directly: it can sequence graduated tasks, embed guiding questions and scaffolding at the point of need, deliver staged feedback tied to reasoning rather than to final answers alone, and capture the process traces through which

critical thinking becomes observable (Swidan et al., 2022). This last affordance is not merely convenient but methodologically necessary here, since analyzing which features elicit which critical thinking indicators during task performance (RQ2) requires a medium that records students' step-by-step interaction—data that conventional media do not provide. Such designs are expected to shift learning from mechanical formula application to analytical decision-making.

These strategies, however, share a limitation that constrains their capacity to promote critical thinking reliably: they attend to the products of student work rather than to the reasoning that generates them. Routine assessments reward procedural accuracy while leaving students' analysis and evaluation of strategies unexamined (Shaughnessy et al., 2019); where classroom discussion does occur, it is rarely anchored to explicit, measurable critical thinking criteria (Llinares, 2019); and e-modules are frequently reduced to digital worksheets that let students reproduce worked examples without constructing justifications (Amelia et al., 2026; Hidayat et al., 2025; Johar et al., 2025; Martin et al., 2021; Utami et al., 2018). The same product orientation recurs in how these approaches give feedback, which privileges the final answer over the quality of reasoning, the soundness of the method, and the plausibility of the result, and in how they are evaluated, typically stopping at pretest–posttest scores without documenting how students' cognitive processes change. This orientation also obscures an inconsistency in the evidence itself: even interventions that report positive effects do so unevenly, with substantial gains confined to only a subset of students (Vicente et al., 2022), yet without process-level data such divergent outcomes cannot be traced to the design features that drive them. Across otherwise different media and designs, then, the recurring gap is the absence of process-level evidence. This is why reported gains remain difficult to explain, and why the contribution of any specific design feature to those gains stays unidentified—the unresolved issue this study sets out to address.

This situation underscores the need for algebra-centric problem-solving activities that intentionally stimulate indicators of critical thinking. Assignments must require students to convert patterns and contexts into appropriate algebraic expressions (Huang et al., 2019). Tasks should facilitate the examination of connections and constraints before students choose the suitable formula (Klein & Leikin, 2020). They should also commence inference, such as by deriving the n th-term formula or linking recursive and explicit representations. Tasks must require evaluation, such as assessing the feasibility of an outcome or comparing two strategies to determine which is more effective and valid. Tasks must necessitate clarification by a methodical exposition of reasoning and verification of processes. This method facilitates the recognition of critical thinking during students' work processes, rather than exclusively through final grades.

A preliminary investigation in the specified context should discover patterns of errors and inadequacies in students' reasoning concerning sequences and series. Diagnostic evaluations and analyses of written responses may reveal common problems, such the failure to distinguish between arithmetic and geometric sequences, inadequacies in generalizing the n th term, and inconsistency in result verification (Yao, 2020). Task-based interviews may corroborate these findings by illustrating that certain students understand the topic but have difficulties in formalizing it into an algebraic model (Leiss et al., 2019). Contemporary

research on e-modules indicates improvements in higher-order thinking skills, such as critical thinking, following its implementation in classrooms. Nevertheless, many studies do not reliably identify specific e-module aspects that elicit indicators of critical thinking as students engage with tasks. Many studies emphasize quantitative results, giving less attention to process-oriented findings. This gap highlights the need for research that links outcome improvement with activation mechanisms at both the feature and task levels.

This study contributes to e-module research by shifting the unit of evidence from whether an e-module works to how it works. Existing studies generally demonstrate that an e-module can raise critical thinking as an aggregate outcome; this study instead specifies the mechanism behind that outcome by identifying which design features—guiding questions, scaffolding, feedback, and reflection (elicit which indicators of critical thinking)interpretation, analysis, inference, evaluation, and explanation—as students work through algebraic tasks on sequences and series. Theoretically, this feature-to-indicator model advances the account of how digital learning environments support critical thinking, reframing it as something evidenced in students' observable work processes rather than inferred solely from outcome scores. Practically, the resulting design principles help teachers and developers decide which features to embed in order to elicit particular reasoning processes, and support assessment that attends to the quality of reasoning rather than to final answers alone. These contributions motivate two research questions. The first concerns the extent of improvement in critical thinking after students use a problem-solving e-module with algebraic tasks on sequences and series; the second concerns which e-module features most consistently elicit those indicators during task performance, as determined through process analysis and student feedback: (1) RQ1: To what degree does the implementation of a problem-solving e-module with algebraic problem challenges enhance students' critical thinking skills in sequences and series? (2) RQ2: Which e-module attributes most reliably elicit critical thinking indications during task execution, as determined by process analysis and student feedback?

2. METHOD

2.1. Settings

This study utilized an explanatory sequential mixed-methods design as its primary objective necessitated two complimentary outputs. The quantitative phase was undertaken initially to acquire a validated measure of enhancement, ensuring that the preliminary findings were not exclusively dependent on process-based interpretation. The qualitative phase was conducted subsequent to the initial quantitative analysis to inform case selection and exploration focus based on developing outcome patterns, including high, moderate, and low improvement groups. The study was conducted in standard classroom environments to evaluate the e-module intervention under genuine conditions, hence enhancing the practical significance and ecological validity of the results. The intervention focused exclusively on a singular unit concerning sequences and series throughout multiple meetings to ensure content alignment, implementation consistency, and the quantifiability of increases in students' proficiency. The synthesis of findings occurred during the interpretation phase by merging quantitative and qualitative results, as this method facilitates a cohesive elucidation of both

"the degree of change" and "the elements that consistently elicited critical thinking indicators" within a unified scientific argument.

Because the study prioritizes explaining the mechanism through which particular e-module features elicit critical thinking indicators, rather than establishing comparative causal superiority over an alternative medium, it uses a single-group design without a control or comparison class. This design fits the second research question, which concerns how features activate indicators during task performance, something a between-group contrast would not reveal. The choice does, however, constrain causal inference: with no comparison group, the observed pretest-to-posttest gains cannot be attributed to the e-module alone, since maturation, repeated testing, and concurrent instruction may also contribute. The quantitative phase is therefore interpreted as evidence of change accompanying the intervention rather than proof of its exclusive effect, and the qualitative phase traces how that change arose at the feature and task levels.

2.2. Participants

This study involved 89 senior high school students drawn from three intact classes at the same grade level within a single school. The classes were selected through purposive sampling: [because they constituted the available classes at the target grade level in this school,] all sharing the same grade and institutional background, which held the students' educational context constant and allowed the e-module intervention to be examined under authentic classroom conditions rather than a reconstituted research group. Conducting the study across three classes of comparable background, rather than a single class, broadened the range of student work available for process analysis while keeping contextual variation controlled. The three classes comprised Class A ($n = 28$), Class B ($n = 29$), and Class C ($n = 32$); all participants were aged 15 to 17 years, consistent with the typical profile of students at this level. The sample was predominantly female (67 female, 22 male; 75.28% and 24.72% of the total), a distribution that was consistent across the three classes and is acknowledged as a limitation, since it constrains the extent to which the findings can be generalized across gender. [Table 1](#) reports the composition of each class.

Prior to the intervention, all participants completed a pretest assessing critical thinking on sequences and series. Because the qualitative phase required cases that reflected a range of entry points, the pretest scores were used to stratify students purposively into high ($n = 12$), moderate ($n = 53$), and low ($n = 24$) initial-ability groups, from which participants for the process analysis were selected. This two-stage procedure—intact classes for the quantitative phase and purposive, stratified case selection for the qualitative phase—ensured that the process data captured diversity in initial ability and helped explain the patterns observed in the quantitative results.

Table 1. Demographic profile of the research sample

Aspect	Category	n	Percentage
Class	Class A	28	31.46%
	Class B	29	32.58%
	Class C	32	35.96%
Age	15-17 years	89	100.00%

Aspect	Category	n	Percentage
Gender	M	9	32,14%
	F	19	67,86%
Class B	M	6	20,69%
	F	23	79,31%
Class C	M	7	21,88%
	F	25	78,12%
Initial critical thinking ability (pretest)	High	12	13.48%
	Moderate	53	59.55%
	Low	24	26.97%
	Total	89	100.00%

2.3. Instruments

This study utilized instruments designed to address the two research topics in accordance with the explanatory sequential mixed-methods framework. During the quantitative phase, the principal instrument utilized was a critical thinking assessment focused on sequences and series of algebraic problem-solving exercises. The assessment was conducted as both a pretest and a posttest utilizing the identical indication pattern to ensure that variations in students' abilities could be evaluated equitably and methodically. Assessment was performed utilizing an analytic rubric that delineated students' performance across critical thinking metrics, including interpretation, analysis, inference, assessment, and explanation. During the qualitative phase, the study supplemented the assessment with process-oriented tools that facilitated the identification of mechanisms that stimulated critical thinking while students engaged with tasks in the e-module. The process instruments were a worksheet for assessing students' work traces in the e-module, a task-based interview guide, and a feature engagement observation sheet to document evidence of leading questions, scaffolding, feedback, and reflection. Moreover, structured reflections or concise questionnaires, in conjunction with e-module activity logs, were employed as supplementary data to enhance triangulation and assess the consistency of feature engagement among students and between assignments. All instruments underwent expert assessment to guarantee content alignment with the measured constructs and were piloted on a limited scale to verify readability, implementation practicality, and scoring and coding consistency.

To measure students' critical thinking performance on sequences and series, the pretest and posttest were developed based on a blueprint that aligns each content focus with a corresponding algebraic task form, sample item, targeted critical thinking indicator, and evidence to be assessed. The blueprint spans six content areas (arithmetic sequences, geometric sequences, series, arithmetic series, geometric series, and mixed tasks) each constructed to elicit specific indicators such as interpretation, analysis, inference, evaluation, and explanation. For instance, the arithmetic sequence item requires students to generalize the formula for U_n and justify their reasoning, whereas the geometric series item requires students

to evaluate two competing solution strategies and identify the source of error in the incorrect one, thereby capturing evidence of evaluative and explanatory reasoning. The complete blueprint is presented in [Table 2](#).

Table 2. Blueprint of the critical thinking test (pretest and posttest) on sequences and series

Content Focus	Form of Algebraic Problem Task	Example	Targeted Critical Thinking Indicators	Evidence Assessed
Arithmetic sequence	Generalizing the n th term	“Given the sequence 4, 7, 10, 13, ... Write the formula for U_n . Explain your reasoning.”	Interpretation, Inference, Explanation	U_n formula, justification for parameter selection
Geometric sequence	Converting a recursive form to an explicit form	“Given $U_n = 3U_{n-1} - 1$, with $U_1 = 2$. Determine U_n . Verify your result for $n=1,2,3$ ”	Analysis, Inference, Evaluation	Transformation of form, value verification
Series	Sigma notation and reasoning	“Express the sum of the first five terms of an arithmetic sequence in sigma notation. Explain the meaning of the index.”	Interpretation, Analysis, Explanation	Sigma form, explanation of the index
Arithmetic series	Contextual modeling	“A savings amount increases by a constant value each week. Construct a model for the amount of savings in the n th week and the total savings up to the n th week.”	Interpretation, Analysis, Inference	Model of U_n , model of S_n
Geometric series	Strategy evaluation	“Two students provide different solutions for the sum of a geometric series. Which answer is correct? Explain the error in the other solution.”	Evaluation, Explanation	Error identification, argumentation
Mixed	Strategy selection	“A problem can be solved in two ways. Choose the most efficient method and explain your choice.”	Analysis, Evaluation, Explanation	Comparison of strategies, justification of choice

To address the second research question concerning the mechanisms that stimulate critical thinking as students engage with the e-module, the study employed a set of process-oriented instruments designed to capture qualitative evidence beyond test scores. These instruments (comprising a work-trace analysis worksheet, a task-based interview guide, a feature engagement observation sheet, a structured reflection or brief questionnaire, and an e-module activity log) differ in their unit of analysis and the type of data collected, ranging from solution-step segments and think-aloud transcripts to timestamped behavioral observations and activity logs. Each instrument was mapped to specific critical thinking indicators and

accompanied by illustrative evidence, such as strategy revision following feedback or verification behavior prior to submission, allowing triangulation between cognitive outcomes and the underlying reasoning processes. Table 3 summarizes the purpose, unit of analysis, and indicators associated with each instrument.

Table 3. Process instruments for RQ2

Instrument	Purpose	Unit of Analysis	Data Collected	Critical Thinking Indicators Captured	Examples of Evidence Sought
Worksheet for analyzing students' work traces in the e-module	To identify critical thinking indicators that emerge during task completion	Segments of solution steps for each task	Responses, scratch work, revisions, use of support, responses to feedback, reflections	Interpretation, Analysis, Inference, Evaluation, Explanation	Strategy revision after feedback, verification of results, justification for formula selection
Task-based interview guide	To explain the reasoning behind strategy use and responses to module features	Problem-solving episodes	Think-aloud protocols or post-task interviews, audio recordings, transcripts	Analysis, Evaluation, Explanation, Interpretation	Reasons for choosing a method, reasons for changing strategies, ways of checking errors
Feature engagement observation sheet	To record students' engagement when specific features appear	Per feature event	Observation notes, time stamps, visible behavior	Interpretation, Evaluation, Self-monitoring as part of evaluation	Reading a guiding question and then writing a reason, opening scaffolding and then correcting a step
Structured reflection or brief questionnaire	To capture perceptions of feature use and strategy	Per student per meeting	Short Likert-scale responses plus open-ended responses	Explanation through self-report, self-evaluation	"Feature X helped because...", "The most difficult part was... because..."
E-module activity log	To strengthen evidence of feature engagement	Per student per activity	Duration, number of attempts, scaffolding clicks, feedback access, answer revisions	Indicators are not assessed directly, but support explanation of consistency	Frequency of scaffolding access on a particular task, patterns of revision after feedback

To clarify how individual e-module features are theorized to elicit specific critical thinking indicators, the study operationally defined each feature and linked it to corresponding

observation items, work-trace items, and interview questions. Four features were examined (guiding questions, step-by-step scaffolding, feedback, and reflection) each associated with distinct indicator profiles, such as interpretation, analysis, explanation, and evaluation for guiding questions, and evaluation and explanation for feedback and reflection. Concrete behavioral markers were specified for each feature, for example, a student opening the scaffolding and subsequently revising a solution step, or checking a result through substitution before submission, providing operational criteria for coding feature engagement across data sources. This feature-to-indicator mapping is detailed in [Table 4](#).

Table 4. Mapping of e-module features to critical thinking indicators and sample observation items

E-module Feature	Operational Definition in the Study	Expected Critical Thinking Indicators	Items on the Observation Sheet	Items in the Work-Trace Analysis	Task-Based Interview Question
Guiding questions	Questions that require students to provide reasons, explain strategies, or verify results	Interpretation, Analysis, Explanation, Evaluation	"The student wrote a reason after the guiding question appeared."	"A written reason is present. A conceptual reference is included."	"Which part of the guiding question changed the way you were thinking?"
Step-by-step scaffolding	Gradual support provided when students reach an impasse	Analysis, Inference, Evaluation	"The student opened the scaffolding and changed a step."	"A revision occurred after the scaffolding was accessed."	"Why did you choose that support, and what changed afterward?"
Feedback	Information indicating correctness or providing cues for improvement	Evaluation, Explanation	"The student read the feedback and then revised the answer."	"An error correction is present, along with a new justification."	"What was the main error you realized after receiving the feedback?"
Reflection	End-of-task prompts about the process and the plausibility of the result	Evaluation, Explanation	"The student checked the result before submitting."	"Verification is present, such as substitution checking or pattern checking."	"How did you make sure that your answer was reasonable?"

The critical thinking test comprised six constructed-response items covering sequences and series. Its content validity was established through expert judgment by three experts in critical thinking scope research, yielding a Content Validity Index of 0.77. This value approaches, while falling marginally, the conventional threshold of 0.78 for a three-expert panel; the items were revised following the experts' feedback before administration. Internal-consistency reliability was acceptable, with Cronbach's alpha of 0.71, a value that is

satisfactory for a six-item instrument given the known sensitivity of alpha to test length. For the qualitative instruments, two raters independently coded students' work traces and interview data against the critical thinking indicators. Inter-rater reliability reached 0.73, indicating substantial agreement. Because the second research question relies on this coding rather than on test scores alone, the substantial inter-rater agreement provides the primary basis for the reported feature-to-indicator patterns.

2.4. Data Collection

The data collecting phase commenced with procedural preparation to guarantee uniform execution across the three classes. The collection of quantitative data commenced with the administration of a pretest on critical thinking regarding sequences and series before to the implementation of the intervention. The pretest was conducted for all three courses within a tightly coordinated timeframe to mitigate the impact of variations in the academic calendar. The researcher standardized the duration, testing protocols, and instructions across all classes, while underscoring the significance of documenting both rationale and solution procedures. The pretest results were gathered, classified by class and student number, and archived as baseline data. Subsequent to the pretest, the instructional intervention employing the problem-solving e-module was executed throughout the three classrooms following a consistent learning sequence. Each meeting adhered to a consistent sequence of activities, commencing with problem orientation, succeeded by strategy exploration, methodical problem solving, and concluding with verification and reflection. The e-module featured algebraic problems necessitating generalization, modeling, justification, and result evaluation to facilitate the emergence of critical thinking indicators in a systematic manner. In specific portions, the e-module presented leading questions prompting students to elucidate their thinking, select a technique, or assess the validity of their results. When students faced challenges, the e-module offered incremental scaffolding that could be utilized as required, allowing for the tracking of support usage and subsequent strategic modifications. The e-module included feedback that encouraged students to amend their responses or rectify their solution processes, concluding each exercise with a structured reflection and self-assessment. During the implementation, the researcher ensured that the teacher employed identical procedures across all three classrooms to prevent discrepancies in outcomes due to instructional variation.

During each implementation meeting, the researcher methodically gathered process data as the primary source for RQ2. Learning artifacts were gathered from the e-module, encompassing students' responses to each task, answer changes, written rationales, and feedback responses. Students' reflection notes were gathered at the conclusion of each exercise as evidence of self-assessment and retrospective analysis of strategy utilization. When accessible, the researcher also obtained e-module activity logs, including the frequency of scaffolding access, the duration of feedback reading, and the number of answer revision attempts. Concurrently, observers executed feature engagement observations through event-based recording, namely by documenting students' behaviors in response to leading questions, scaffolding, feedback, and reflection. The observations concentrated on observable activities, like reviewing the problem, pausing to verify a finding, altering strategy, or providing

justification. The researcher recorded field notes to describe the classroom atmosphere, including pertinent exchanges, technological challenges, and students' overall reactions. This sequence of data gathering meant that the study not only measured score changes but also elucidated the mechanisms that followed them.

Upon the completion of the instructional unit, the researcher conducted a critical thinking posttest on sequences and series, utilizing the identical indicator blueprint as the pretest. The posttest was conducted for all three classes using identical processes, encompassing the same duration, instructions, and testing regulations. The researcher reiterated the significance of documenting reasoning and solution processes to facilitate data analysis by indicator. The posttest results were gathered, encoded, and archived with the pretest data for comparative study. The researcher subsequently performed preliminary scoring to determine overall scores and indicator-level scores, along with the extent of progress for each student. The quantitative data provided the foundation for identifying patterns of enhancement at both the class and individual levels. This phase completed the quantitative data collecting and simultaneously initiated a more concentrated qualitative data collection phase.

The qualitative data collecting phase occurred subsequent to the completion of the preliminary quantitative analysis to ensure that participant selection was grounded in evidence of genuine progress. The researcher chose interview participants from each class by accounting for variations in improvement, categorizing them into high, moderate, and low-improvement groups, while also guaranteeing representation of original ability profiles. The chosen participants engaged in task-based interviews or think-aloud sessions utilizing various representative problems that mirrored the requirements of the critical thinking indicators and the primary e-module characteristics. The interviews addressed students' work artifacts and their experiences with certain aspects, allowing for the verification of their answers against their work traces. The researcher documented the interviews, made transcripts, and gathered supplementary materials, including revised solutions produced during the think-aloud sessions. To ensure consistency, the researcher employed a uniform interview guide for all participants, while permitting follow-up questions tailored to individual responses.

2.5. Analysis

The quantitative analysis addressed RQ1 by comparing students' critical thinking scores before and after the intervention, at both the total-score and indicator levels, using the same analytic rubric. Descriptive statistics (mean, standard deviation, and median) characterized the score distributions. Because the pretest–posttest comparison relies on a parametric paired-samples test, the normality of the difference (gain) scores was examined using the Shapiro–Wilk test. Since the difference scores met the normality assumption, a paired-samples t-test was conducted to evaluate the change from pretest to posttest for the total score and for each of the five indicators. Practical significance was quantified using Cohen's d_z for paired designs, complemented by the normalized gain computed at the individual level and aggregated to class and overall means. Results were reported at both class and aggregate levels and compared across indicators to assess the consistency of change. Given the single-

group design, these results are interpreted as change accompanying the intervention rather than as evidence of its exclusive causal effect.

The qualitative analysis addressed RQ2 by examining the processes through which students engaged the e-module features. Data were drawn from three sources (students' work traces in the e-module, task-based interviews, and classroom observation notes) for 27 students selected from the high, moderate, and low improvement groups to capture variation in outcome patterns. The data were segmented into meaning units, each corresponding to a single solution step, a response to feedback, or a reflective remark, yielding [total number] units. Coding followed a hybrid approach. The five critical thinking indicators served as a predefined deductive scheme derived from the analytic rubric, while inductive codes were added as new patterns emerged from the data, including specific algebraic errors and strategic adjustments. Each unit was coded for the relevant critical thinking indicator and the specific e-module feature in use. Codes were then grouped into themes to identify the features that most consistently elicited critical thinking indicators. Coding was carried out using N.Vivo software / manual. Credibility was further supported by triangulating the three data sources against one another. The complete coding scheme, comprising the critical thinking indicators, e-module features, and inductive codes with their operational definitions, coded data examples, and data sources, is presented in [Table 5](#).

Table 5. Coding scheme for the qualitative analysis of RQ2

Code Group	Code	Brief Operational Definition	Example of Coded Data Unit	Data Source
Critical thinking indicators	CT-INT	Accurately interpreting problem information or mathematical notation	"Given ..., asked ...", explaining the meaning of an index or sigma notation	Work traces, interviews
	CT-ANL	Analyzing relationships and selecting a strategy or formula with justification	Explaining why an arithmetic or geometric approach was chosen	Work traces, interviews
	CT-INF	Deriving a model or conclusion from a pattern or structure	Deriving U_n , converting a recursive form into an explicit form	Work traces, interviews
	CT-EVA	Checking results, detecting errors, and making corrections	Checking initial terms, correcting a step after feedback	Work traces, observations, interviews
	CT-EXP	Providing a coherent and verifiable justification	Giving reasons for a step, brief proof, written verification	Work traces, interviews
E-module features	F-GQ	Student response to guiding questions	Writing a reason after a prompt appears	Work traces, observations
	F-SC	Use of scaffolding and its effect on solution steps	Opening support and then changing strategy	Work traces, observations, logs if available
	F-FB	Response to feedback	Revising an answer after reading feedback	Work traces, observations, logs if available
	F-RF	Response to reflection	Writing a self-evaluation or strategy rationale at the end of a task	Work traces, reflections

Code Group	Code	Brief Operational Definition	Example of Coded Data Unit	Data Source
Inductive codes, field-based patterns	IND-ALG	Recurrent algebraic manipulation errors	Incorrect substitution, incorrect simplification, incorrect parameter use	Work traces, interviews
	IND-MOD	Errors in translating context into algebraic form	An algebraic relationship that does not match the problem context	Work traces, interviews
	IND-SW	Moments of strategy change	Shifting from a formula-based approach to a table, or the reverse, with justification	Work traces, interviews, observations

3. RESULTS AND DISCUSSION

3.1. Results

3.1.1. Quantitative Results (RQ1): Pretest–Posttest Changes in Critical Thinking

Table 6 indicates that the first critical thinking skills of the three classes were notably comparable. The pretest mean for Class A was 52.02 (SD = 11.79), while the median was 52.50. Class B had a median score of 51.00 and a pretest mean of 50.93 (SD = 12.17). Class C had a median score of 49.00 and a pretest mean of 48.58 (SD = 10.99). The pretest median was 50.00, and the mean was 50.43 (SD = 11.59). The initial score distribution of the pretest was very balanced and not skewed by outliers, as evidenced by the closeness of the mean and median. All courses exhibited improved posttest outcomes subsequent to the intervention. Class A demonstrated progress, evidenced by a median score of 68.00 and a posttest mean of 67.46 (SD = 16.89). Class B demonstrated growth, achieving a median score of 68.00 and a posttest mean of 67.91 (SD = 14.93). Class C demonstrated advancement, achieving a median score of 66.00 and a posttest mean of 66.04 (SD = 16.48). The posttest median was 67.00, and the total mean was 67.09 (SD = 15.96). $\Delta M = 15.44$ for Class A, 16.98 for Class B, 17.46 for Class C, and 16.66 overall is the difference in means between the pretest and posttest. Furthermore, the median increased by 15.50 for Class A, 17.00 for Class B, 17.00 for Class C, and 17.00 overall. The mean normalized gain, situated inside the moderate range, signified the degree of enhancement. The N-gain for Class A was 0.34, for Class B was 0.32, and for Class C was 0.36. The N-gain for the complete sample was 0.34. In each class, the pretest-posttest comparison demonstrated a statistically significant difference ($p < 0.05$). The effect size of the adjustment was significant, with d values of 1.28 for Class A, 1.03 for Class B, 1.49 for Class C, and an overall value of 1.24. The greater variability in achievement after the intervention is evidenced by the larger posttest standard deviation relative to the pretest standard deviation, signifying that students' improvement was inconsistent. The implementation of the algebraic problem-solving e-module led to an enhancement in critical thinking abilities related to sequences and series, demonstrated by an increase in the overall score from 50.43 to 67.09 ($\Delta M = 16.66$), an N-gain of 0.34 (moderate category), and a substantial effect size ($d = 1.24$). This enhancement was uniform throughout the three classes.

Table 6. Descriptive statistics of critical thinking scores (pretest and posttest)

Class	n	Mean Pre (SD)	Median Pre	Mean Post (SD)	Median Post	N-Gain	Comparison Test	Effect Size
A	28	52.02 (11.79)	52.50	67.46 (16.89)	68.00	0.34	t = 6.75 (df = 27)	d = 1.28
B	29	50.93 (12.17)	51.00	67.91 (14.93)	68.00	0.32	t = 5.53 (df = 28)	d = 1.03
C	32	48.58 (10.99)	49.00	66.04 (16.48)	66.00	0.36	t = 8.45 (df = 31)	d = 1.49
Total	89	50.43 (11.59)	50.00	67.09 (15.96)	67.00	0.34	t = 11.69 (df = 88)	d = 1.24

An indicator-level analysis was subsequently conducted to ascertain which facets of critical thinking exhibited the most significant improvement post-intervention. The analysis assessed the statistical significance and practical importance of the changes, with the differences between the pretest and posttest averages. The mean and standard deviation for the pretest and posttest, the mean difference (Δ), the results of the pretest-posttest comparison test, and the effect size were utilized to report each indication. [Table 7](#) and [Figure 1](#) present the findings.

Table 7. Modifications by indicator, comparative test outcomes, and effect sizes (N = 89)

Indicator	Mean Pre (SD)	Mean Post (SD)	Δ (Post-Pre)	Paired-samples t-test (df = 88)	p-value	Effect Size (Cohen's dz)
Interpretation	11.20 (2.80)	14.80 (3.10)	3.60	t = 9.70	< 0.05	1.03
Analysis	10.40 (2.70)	13.50 (2.90)	3.10	t = 8.10	< 0.05	0.86
Inference	9.60 (2.60)	12.70 (3.00)	3.10	t = 7.70	< 0.05	0.82
Evaluation	9.00 (2.40)	12.90 (3.20)	3.90	t = 9.95	< 0.05	1.05
Explanation	10.23 (2.65)	13.19 (3.05)	2.96	t = 7.15	< 0.05	0.76

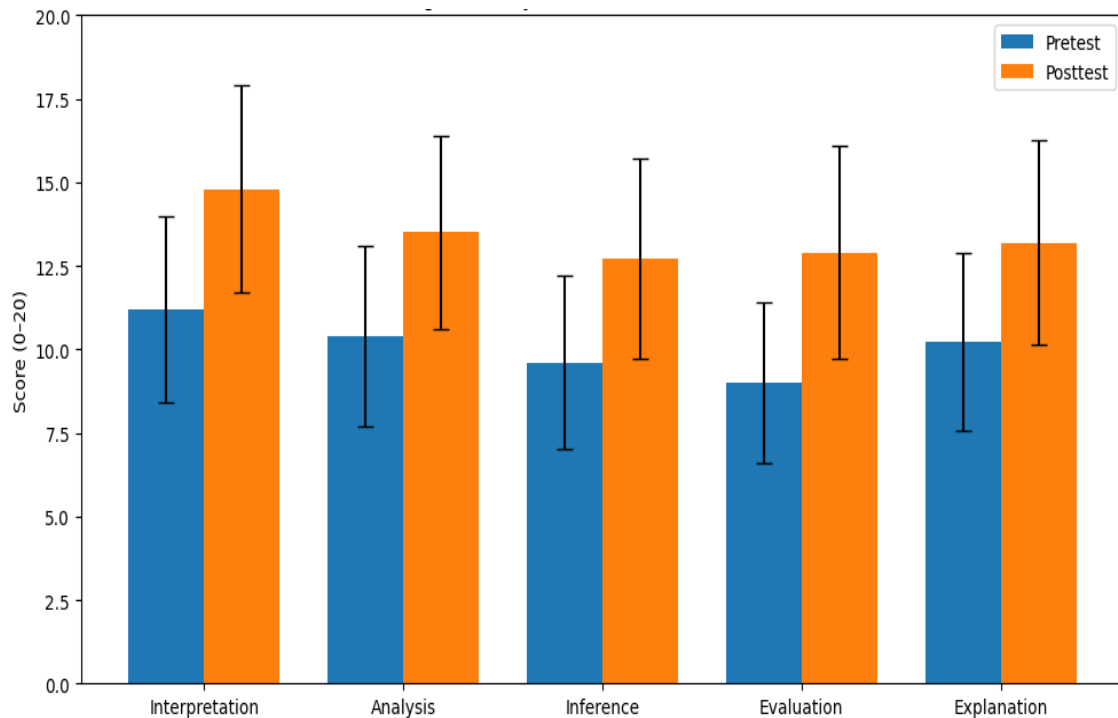


Figure 1. Average critical thinking scores by indicators

The mean score of the interpretation indicator increased from 11.20 (SD = 2.80) to 14.80 (SD = 3.10), reflecting a mean difference of $\Delta = 3.60$. The difference exhibited a substantial effect size ($d_z = 1.03$) and was statistically significant ($t(88) = 9.70$, $p < 0.05$). This pattern indicates enhanced understanding of problem specifics, terminology, and the inquiry. Some students may have exhibited more rapid improvement than others, as seen by the higher standard deviation in the posttest results. The analytical indicator's mean score increased from 10.40 (SD = 2.70) to 13.50 (SD = 2.90) with $\Delta = 3.10$. A statistically significant difference with a considerable effect size ($d_z = 0.86$) was established by the paired-samples t-test ($t(88) = 8.10$, $p < 0.05$). This outcome indicates that pupils' ability to understand relationships and discern more effective strategies or formulas through enhanced logical reasoning has advanced. Although the overall trajectory of growth remained steady, the slight rise in standard deviation indicates persistent variability in pupils' progress.

The mean score of the inference indicator increased by $\Delta = 3.10$, from 9.60 (SD = 2.60) to 12.70 (SD = 3.00). The difference exhibited a substantial effect size ($d_z = 0.82$) and was statistically significant ($t(88) = 7.70$, $p < 0.05$). This enhancement demonstrates an enhanced ability to formulate mathematical conclusions, like deriving Un, establishing a valid generalization, or converting a recursive expression into an explicit form. Individual variance must still be considered through process-based evidence, since the rise in standard deviation from 2.60 to 3.00 signifies that the level of achievement in inference remains unevenly distributed. The mean score of the assessment indicator rose significantly, from 9.00 (SD = 2.40) to 12.90 (SD = 3.20), with a change of $\Delta = 3.90$. The comparison test indicated a substantial effect size ($d_z = 1.05$) and a statistically significant difference ($t(88) = 9.95$, $p < 0.05$). This study indicates an improved ability to evaluate the credibility of findings, detect errors, and adjust procedural procedures in the solution. Some students shown substantial

growth in evaluation, whereas others exhibited more modest improvement, as indicated by the notable increase in the posttest standard deviation. With $\Delta = 2.96$, the mean score of the explanation indicator increased from 10.23 (SD = 2.65) to 13.19 (SD = 3.05), representing the least improvement among all indicators. The effect size of $d_z = 0.76$, categorized as moderate-to-large, indicates that the difference remains statistically significant ($t(88) = 7.15$, $p < 0.05$). This trend indicates that, although at a slower rate than other components, students' ability to construct explanations and arguments has improved. The increase in standard deviation indicates that the quality of students' arguments varied significantly among individuals.

All indicators demonstrated significant enhancements overall, however the extent of improvement varied by indication. Evaluation and interpretation demonstrated the most substantial benefits, while explanation shown the least enhancement, but with a significant rise. The elevated standard deviations of the posttest indicate that students' improvement was inconsistent. This pattern establishes the basis for examining the triggering mechanisms in RQ2, particularly identifying the e-module features that most often elicited the indicators that exhibited the greatest improvement and the factors that contributed to the slower progression of the explanation.

3.1.2. Qualitative Results (RQ2) Features of the E-Module That Most Consistently Triggered Critical Thinking Indicators

Two attributes were identified as the most dependable indicators of critical thinking: guiding questions (F-GQ) and feedback (F-FB). The students' inclination to assess results, recognize errors, and adjust their methods following feedback suggests that F-FB predominantly stimulated evaluation (CT-EVA). Rationale for strategy selection and justification of steps arose following the emergence of prompts, demonstrating that F-GQ predominantly elicited analysis (CT-ANL) and explanation (CT-EXP). Inference (CT-INF) was activated by scaffolding (F-SC), although its consistency was predominantly contingent upon students' capacity to implement the instructions and secure the support. CT-EVA and CT-EXP were elicited through reflection (F-RF); nonetheless, numerous students provided only succinct responses, resulting in their contribution appearing inferior to that of F-FB and F-GQ. The inductive codes revealed two primary obstacles that often hindered the emergence of inference and explanation: faults in algebraic manipulation (IND-ALG) and modeling inaccuracies (IND-MOD). Following F-FB and F-SC, instances of strategic alteration (IND-SW) were most prevalent, indicating that these two attributes acted as significant catalysts for modifications in students' cognitive processes throughout task execution. As summarized in [Table 8](#) and visually illustrated in [Figure 2](#), F-FB and F-GQ exhibited the highest student frequency and consistency (85.19% and 77.78%, respectively), whereas F-SC and F-RF showed moderate to low consistency, reinforcing the pattern of feature-indicator reliability.

Table 8. Summary of e-module features triggering critical thinking indicators (RQ2)
(N = 27 observed students)

Feature	Most Frequently Triggered Indicators	Observable Process Evidence Pattern	Student Frequency (%)	Consistency
F-GQ (guiding questions)	CT-ANL, CT-EXP, CT-INT	Students articulated their rationale for choosing a technique or formula and elucidated the provided information and the inquiries posed following the emergence of the guiding questions.	21 (77.78%)	High
F-FB (feedback)	CT-EVA, CT-EXP, IND-SW	Students reviewed their results, detected inaccuracies, and amended procedures or parameters subsequent to examining the input.	23 (85.19%)	High
F-SC (scaffolding)	CT-INF, CT-ANL, IND-SW	Students transformed recursive forms into explicit forms following assistance, developed U_n , and altered representations, such as from a table to a formula.	16 (59.26%)	Moderate
F-RF (reflection)	CT-EVA, CT-EXP	Students composed self-evaluations; nonetheless, these were frequently succinct and descriptive.	10 (37.04%)	Low

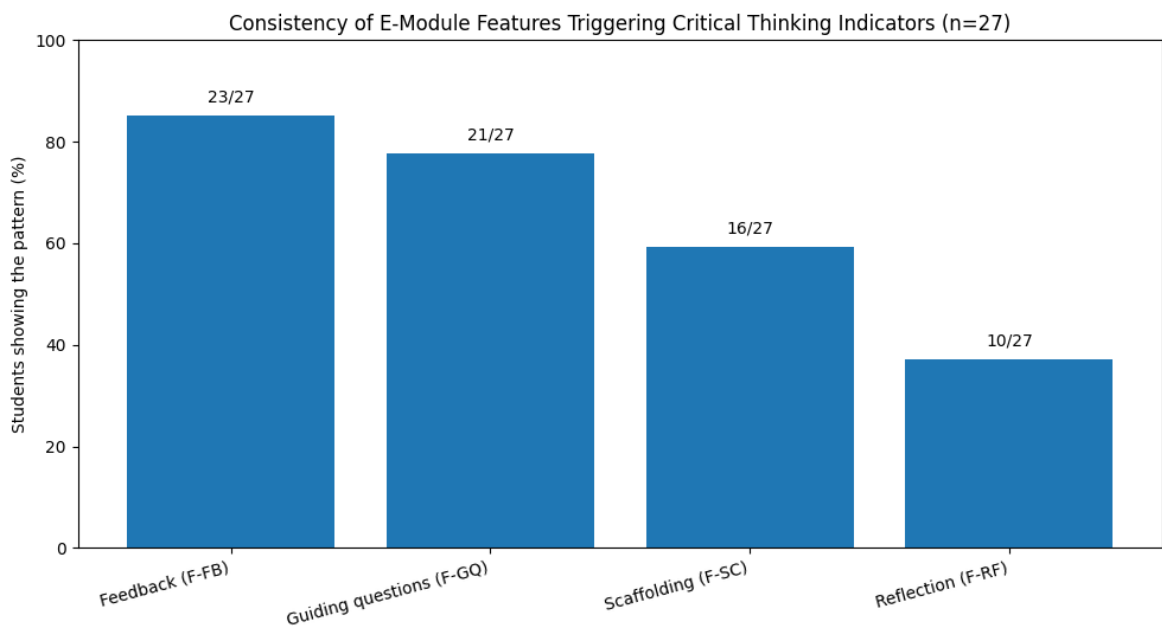


Figure 2. Consistency of e-module features triggering critical thinking indicators

The work traces of students, observation notes, and verbal responses from task-based interviews are subsequently synthesized to provide thematic evidence for each characteristic. The interview samples were selected to exemplify both prevalent patterns and variations in responses across students with differing improvement profiles. This presentation elucidates the relationship between features and indications based on the identified CT, F, and IND codes. This approach yields triggering mechanisms that may be monitored through process-based evidence, surpassing mere frequency counts.

Table 8 indicates that directing questions (F-GQ) stimulated critical thinking indicators, specifically analysis (CT-ANL) and explanation (CT-EXP), in 21 out of 27 students, equating to 77.78%. According to the work traces, students frequently provided justifications for their choice of an arithmetic or geometric sequence immediately following inquiries such as "why" or "how do you know." The observation data confirmed this tendency, revealing that students favored composing justification statements not present in their first solution steps after extending their pause at the prompt area. This pattern fulfilled the criterion of "consistency" at the feature level, as it was detected across several classes and initial ability groups. The quality of the justifications varied, however. Low-gain students sometimes provided cursory explanations such as "it increases" without examining the difference or ratio, but high-gain students typically connected their reasoning to the concepts of constant difference or constant ratio. Although the quality of explanation differed, these findings position F-GQ as an effective catalyst for enhancing analytical processes and encouraging students to express their ideas. This is a presentation of the interview excerpts.

- S17 : *"When the question 'why' appeared, I checked the difference. Since the difference was always 3, it meant the sequence was arithmetic. So, the formula was $U_n = 4 + (n - 1)3$ " [F-GQ + CT-ANL + CT-EXP]*
- S19 : *"I wrote arithmetic because the numbers kept increasing. But when I was asked 'why,' I was confused about how to explain the difference. I did not have time to check the differences one by one." [F-GQ + CT-EXP + IND-ALG]*
- S21 : *"I initially assumed it was geometric because it appeared to increase rapidly. I computed the ratio after the guiding question and discovered it was not constant. I then switched it to arithmetic." [F-GQ + CT-ANL + CT-EVA + IND-SW]*

The indicators of analysis (CT-ANL) and explanation (CT-EXP) appeared more frequently in their work traces due to guiding questions (F-GQ) consistently prompting students to articulate their arguments and select concepts with greater intentionality. Nevertheless, there remained variability in the intricacy of argument, especially among low-gain students who often provided cursory justifications. When students observed disparities in their findings, this pattern indicates that other factors exerted a more substantial influence on triggering reconsideration and modification of strategy.

The most consistent pattern emerged in feedback (F-FB), with 23 out of 27 students, or 85.19%, indicating that this characteristic stimulated critical thinking indicators, notably evaluation (CT-EVA), and was often succeeded by strategy modification (IND-SW). The analysis of the work traces indicated that students frequently reverted to their initial activities,

validated the coherence of the opening sentences, and contrasted their results with the established criteria upon receiving feedback. Students often corrected erroneous forms of *Un* and defective arithmetic procedures, modified parameters, or rectified substitutions. The conclusion was corroborated by observational data indicating that students often paused to review the feedback prior to revisiting the "given and asked" section or the computation of the initial terms before modifying their solutions. This pattern clearly indicated that CT-EVA had been activated, as students were now verifying, recognizing, and rectifying their results instead of accepting them uncritically. For certain students, IND-SW also included assessment, such as transitioning from a direct substitution method to constructing a term table or initially recognizing the pattern. The results unequivocally corroborated the CT-EVA indication by establishing that F-FB was the characteristic most consistently associated with the emergence of discernible evaluative practices in the process-based evidence. This is a presentation of the interview excerpts.

- S13 : *"It stated that my response did not match at the feedback stage. I discovered that the difference I had used was incorrect when I went back to compare the first and second words. I then made revisions to the formula."* [F-FB + CT-EVA + CT-INT]
- S22 : *"I double-checked the original terms after the feedback." I adjusted my Un and recalculated it after realizing that I had entered the value improperly.* [F-FB + CT-EVA + IND-SW]
- S23 : *"I assumed it was already right, but I reexamined the processes after seeing the criticism. I went through the procedure again using what was provided and what was requested, but the algebraic simplification step still baffled me."* [F-FB + CT-EVA + IND-ALG]

The evaluation indication (CT-EVA) was predominantly activated by feedback (F-FB), as students evidently reviewed their work, recognized errors, and modified their methods following the feedback received. The alteration of strategy (IND-SW) often adhered to this pattern, indicating that the assessment extended beyond merely "recognizing a mistake" to encompass a modification of the approach itself. These results, however, also indicate that several students persisted in facing difficulties when revision necessitated more complex algebraic manipulation. When students encounter a standstill, this situation raises the question of how incremental support mechanisms could facilitate their development of inference and generalization skills.

According to Table 8, scaffolding (F-SC) exhibited a moderate level of consistency, with 16 out of 27 students, or 59.26%, displaying a pattern where this feature elicited critical thinking indicators, notably inference (CT-INF), and occasionally succeeded by analysis (CT-ANL) and strategy change (IND-SW). Students predominantly employed scaffolding when encountering difficulties with tasks necessitating the generalization of *Un*, the derivation of a general form from a pattern, or the conversion of a recursive form into an explicit form. The work traces indicated a more structured inference among students who actively utilized the scaffolding. For example, before articulating the formula for *Un*, students initially illustrated the pattern of difference or ratio and subsequently correlated it to the general form. This result was supported by observational data indicating that students typically accessed the help section

after a prolonged period of stagnation, then restarting their work with more systematic approaches. Nonetheless, the limitations of algebraic manipulation compromised the coherence of this feature (IND-ALG). Scaffolding assisted certain students in identifying the structure; however, errors persisted during expression simplification, substitutions, or equation formulation, rendering the conclusion not totally robust. Consequently, scaffolding appeared beneficial in guiding cognition; yet, its efficacy depended on the essential algebraic competencies necessary for executing inferential processes accurately. This is a presentation of the interview excerpts.

- S07 : *"I utilized the sequential assistance to transform the recursive structure into a general format." Subsequently, I composed Un and verified it for various values of n to ensure accuracy."* [F-SC + CT-INF + CT-EVA]
- S09 : *"I accessed the step-by-step support, and only then did I comprehend how to construct Un ." However, I still erred in the simplification, resulting in an inaccurate outcome."* [F-SC + CT-INF + IND-ALG]
- S12 : *"The scaffolding prompted me to identify the pattern of the difference initially." Subsequently, I altered my methodology from direct computation to constructing a table of terms, and only thereafter did I formulate the equation."* [F-SC + CT-ANL + CT-INF + IND-SW]

Scaffolding (F-SC) assisted certain students in developing inference, particularly when they faced challenges in generalizing Un or transforming a recursive form into an explicit one; nonetheless, its effectiveness was limited by difficulties in algebraic manipulation (IND-ALG). When students' foundational mathematical skills were inadequate, scaffolding only outlined the procedural stages without ensuring precise execution, rendering the subsequent inference susceptible to inaccuracies. This pattern elucidates why F-SC failed to attain the same degree of consistency as F-FB and F-GQ in eliciting critical thinking markers. These findings indicate the necessity for mechanisms that promote students to deliberately reassess their methods and mistakes, rather than just adhere to the scaffolding phases.

Ultimately, Table 8 indicates that reflection (F-RF) had the lowest degree of consistency, with merely 10 out of 27 students, or 37.04%, demonstrating a pattern where reflection distinctly elicited critical thinking signs. The reflection traces indicated that this feature prompted students to compose concise summaries of their processes; however, numerous responses were vague, stating remarks like "the formula was difficult" or "I already understood," without specifying the erroneous step, the rationale for the mistake, or the foundation for revision. This trend indicated that the activation of evaluation (CT-EVA) and explanation (CT-EXP) did not consistently manifest among most students, as the reflective reactions failed to produce discernible evidence of thinking. Students who generated more comprehensive reflections exhibited specific self-evaluation, such as recognizing an error in replacement or in calculating the common difference or ratio, followed by an explanation of the corrective measures implemented. Nonetheless, this more complex pattern of reflection was not uniformly observed among individuals or assignments, so F-RF cannot currently be classified as the "most consistent" trigger. The interview excerpts are delineated as follows.

- S03 : *“In the reflection, I noted that the error occurred during the substitution step, which I subsequently rectified by verifying the initial terms for consistency.” [F-RF + CT-EVA + CT-EXP]*
- S05 : *“In the reflection, I noted that my method lacked sufficient efficiency.” Subsequently, I attempted to utilize a table initially before deriving the formula, nevertheless I remained uncertain about the most suitable way.” [F-RF + CT-EVA + IND-SW]*
- S07 : *“I completed the reflection succinctly.” I said, ‘I am perplexed regarding the formula.’ I was uncertain about which aspect I should clarify as incorrect.” [F-RF + CT-EXP]*

The findings indicate that reflection can stimulate CT-EVA and CT-EXP; however, its efficacy is significantly affected by the quality of reflection prompts and students' practices in articulating their reasons in writing. Thus, the function of reflection is more accurately perceived as a supplementary element that requires enhancement, rather than as a principal catalyst like F-FB and F-GQ.

The analysis demonstrated a distinct pattern of consistency in the manner in which the e-module elements elicited critical thinking signs. Feedback (F-FB) revealed as the most reliable catalyst, particularly for evaluation (CT-EVA) and strategy revision (IND-SW), as students distinctly reviewed and rectified their work following the provision of feedback. Guiding questions (F-GQ) regularly elicited analysis (CT-ANL) and explanation (CT-EXP) by prompting students to articulate their rationale and elucidate the foundations for their concept selections. Scaffolding (F-SC) facilitated inference (CT-INF), although its efficacy diminished when students faced challenges in algebraic manipulation (IND-ALG) that hindered precise execution. Reflection (F-RF) had the least consistency, as numerous replies were vague and failed to illustrate precise strategy appraisal. The findings indicate that an effective e-module design must prioritize feedback and leading questions, while also enhancing scaffolding and reflection prompts to enable students to generate more traceable evaluations and justifications.

3.2. Discussion

This section cohesively presents the quantitative and qualitative findings to address the two research topics. The implementation of the problem-solving e-module improved critical thinking abilities about sequences and series. The improvement in overall scores and all metrics corresponds with the logic of problem-solving, requiring a sequence of cognitive tasks, from understanding the issue to assessing the final result. Polya described this process as a cyclical framework comprising understanding, strategizing, execution, and reflection, which automatically promotes reflective thinking in problem-solving activities (DiNapoli & Miller, 2022; Voica et al., 2020). The e-module consistently upheld this cycle due to the tasks being organized in a set sequence that could not be easily circumvented. Algebraic issues increased cognitive demand as students needed to develop generalizations and select the appropriate sequence structure, rather than merely calculate the next term (Blanton et al., 2019; Çekmez, 2020; Yao, 2022). This combination of process structure and work needs clarifies why the improvement was not merely incremental.

The overall improvement can be understood through a critical thinking paradigm that emphasizes key skills such as interpretation, analysis, inference, evaluation, and explanation. Facione's paradigm views critical thinking as a conscious and self-regulatory judgment process, characterized by indications that emerge when students are engaged in processing information, forming connections, drawing conclusions, and evaluating the validity of their actions (Lehmann, 2023; Yasin et al., 2019). The problem-solving tasks in the e-module offer several opportunities to use these abilities specifically within the context of sequences and series (Scheibling-Sève et al., 2020; Xin et al., 2020). This topic requires students to transform numerical patterns, symbolic representations, and general formulas, often necessitating interpretative and analytical endeavors (El Mouhayar, 2018; Wilkie, 2021). Moreover, the imperative to conclude U_n or S_n prompted inference, as students were compelled to transition from specific instances to generalizations. The repeated procedure within a single teaching unit reinforced these cognitive habits, hence the changes in scores can be viewed as the culmination of systematic practice (Hadar & Ruby, 2019).

The degree of improvement across indicators was inconsistent. The assessment and analysis enhanced considerably, as the topic of sequences and series provides an immediate chance for verification through the examination of starting terms or the validation of consistent common differences and ratios. Students assessed a formula in relation to U_1 , U_2 , or U_3 , participating in monitoring and correction, which are essential components of self-regulation according to Zimmerman (2002). Formative feedback improved this cycle by clarifying the difference between the current response and the proper one, so motivating students to adjust their methods, as emphasized in the formative feedback theories (Kontorovich, 2021). The expected increase in interpretation arose from the need for students to differentiate known knowledge from questions, understand notation, and categorize the sort of sequence before progressing. The instructional design increasingly requires the development of interpretation and assessment in observable and trainable activities.

Conversely, explanation advanced at a diminished pace as this metric requires explicit mathematical justification rather than merely selecting the suitable technique. Argumentation theory, as illustrated by Toulmin's model, distinguishes between claims, data, and warrants; yet, many students stop at the claims stage, failing to furnish supporting evidence and conceptual underpinnings (Toulmin, 2003). In mathematics, justification refers to the criteria for proof and rationales that are subject to scrutiny, as detailed in the literature on mathematical justification and proof (Mata-Pereira & da Ponte, 2017). A predominantly procedural learning culture encourages the pursuit of definitive answers while undermining the expression of reasoning, making explanation the most difficult measure to reconcile with the others. This condition does not negate the observed progress; instead, it signifies that advancement in explanation requires more focused design support, such as prompts that urge students to illustrate the common difference or ratio and relate it to the fundamental notion. This result establishes a clear avenue for optimizing the design to attain a more equitable enhancement across indicators.

Mechanisms of Feature-Driven Critical Thinking

The components of the e-module did not stimulate critical thinking indicators to the same extent. Feedback (F-FB) emerged as the most dependable catalyst, particularly for evaluation (CT-EVA) and strategic adjustment, followed by guiding questions (F-GQ), which primarily stimulated analysis (CT-ANL) and elucidation (CT-EXP). Scaffolding (F-SC) enhanced inference (CT-INF), however its effectiveness was more vulnerable to difficulties in algebraic manipulation (IND-ALG). Reflection (F-RF) demonstrated the least consistency, as many responses were vague and lacked specific evaluative approaches. This trend suggests that the development of critical thinking is affected not just by the presence of traits but also by how those traits interact with monitoring, justification, and revision processes throughout task performance. In this study, these critical thinking indicators are understood as domain-specific manifestations of general cognitive processes: within sequences and series, evaluation, inference, and explanation take concrete form as validating an algebraic model, deriving an n th-term expression, and justifying a claim from the definitions of arithmetic and geometric sequences. A mathematical activity was treated as a critical thinking indicator only when it involved reasoning rather than procedural execution alone.

Feedback emerged as the most dependable trigger for assessment, as it activated a certain self-regulation mechanism inside the problem-solving process. Upon detecting a signal of inconsistency, students participated in a monitoring and regulation phase, which is essential to self-regulated learning (Kontorovich, 2021). This procedure entails evaluating results against predetermined standards, identifying the source of the error, and enacting modifications. In formative assessment frameworks, feedback provides insight into the gap between current performance and the desired goal, so encouraging corrective actions rather than simply classifying responses as correct or incorrect (Ayalon & Wilkie, 2020). The effect amplifies when feedback operates at both the task and process levels, allowing students to link the input to a specific step, then retracing their steps to initial words, analyzing frequent disparities or ratios, and modifying parameters. This pattern aligns that effective feedback addresses three questions: the learner's goal, the learner's current status, and the next steps to undertake. Thus, it is not surprising that CT-EVA became famous and was frequently followed by IND-SW, as correcting a mistake often required a strategic shift, such as moving from direct substitution to a term-table approach or doing iterative verification.

Guiding questions enabled analysis and clarification by requiring students to express their reasoning and link conclusions to evidence. From an argumentative perspective, questions such as “why” and “how do you know” require a reasoning structure that minimally includes a claim and supporting evidence, ideally supplemented by a warrant that links the evidence to the claim (Chorney et al., 2024; Lee et al., 2023). In sequences and series, this argumentative structure is realized mathematically: the supporting evidence is typically a constant difference or constant ratio, and the warrant derives from the definitions and properties of arithmetic or geometric sequences. Argument quality varied, as many students offered only claims without evidence or rationale, as in “it is arithmetic because it increases,” yielding a superficial form of CT-EXP. More proficient students presented data producing a fuller representation of CT-ANL and CT-EXP. This variation reflects the standards of justification specific to mathematics, which are more rigorous than everyday explanation

because reasoning must be scrutinized and grounded in formal concepts. Thus guiding questions prompted analysis and explanation, but the depth of argument depended on students' ability to marshal mathematical evidence and connect it to concepts.

Scaffolding demonstrated its value by providing temporary support during periods of impasse, promoting advancement within the zone of proximal development, as emphasized by Vygotsky and the scaffolding framework that requires gradual aid followed by ultimate removal (Kim et al., 2022; Rezat et al., 2022). In tasks requiring inference, such as deriving U_n or converting a recursive form into an explicit form, scaffolding aided students in deconstructing the problem into manageable steps, including identifying the pattern of difference or ratio, formulating the general expression, and subsequently validating it. The effectiveness of scaffolding depended on algebraic problems, as mathematical reasoning requires accurate symbolic execution (Li et al., 2024). The principles of symbolic literacy and symbol sense are essential, as students may comprehend the general concept but encounter difficulties with manipulation, simplification, or substitution, resulting in an erroneous final conclusion. Scaffolding that provides only conceptual direction is insufficient when the main problem lies in the accuracy of symbolic operations. Meanwhile, reflection was frequently superficial, as many students regarded it only as a recounting of feelings or understanding, rather than as an evaluation of methodology and a demonstration of revision (DiNapoli & Miller, 2022). This suggests that metacognitive skills, especially accurate monitoring, were not sufficiently developed, leading to pupils' inability to recognize the stage at which an error occurred, the reasons behind it, and the strategies for rectification. Reflection would be enhanced if prompts included a distinct metacognitive structure, such as an error-cause-revision-verification format, to promote the consistent development of CT-EVA and CT-EXP.

The design implications of these findings are apparent when linked to the profile of indication augmentation. The most significant improvement was noted in indicators supported by mechanisms requiring visible revision, especially feedback, as it directly initiated criterion-based evaluation and encouraged the correction of solution phases. Thus, the e-module should prioritize feedback that is specific and process-oriented rather than merely score-based, thereby encouraging students to participate in focused verification and correction. Guiding questions should go beyond simple calls for justification, requiring explicit evidence so that arguments include claim, data, and warrant. Scaffolding should deliver targeted algebraic support at critical moments, such as precise symbolic manipulation or verification that preserves the integrity of the initial terms, preventing execution errors from degrading inference. Reflection should be structured as a systematic evaluation process rather than open exploration, enabling focused, verifiable assessment. These modifications strengthen the e-module's capacity to support explanation and to promote more balanced progress across indicators.

4. CONCLUSION

This study's principal contribution lies not in showing that critical thinking improved, but in identifying a differentiated pattern of associations between specific e-module features and particular critical thinking processes. Feedback was most consistently associated with

evaluation, prompting students to check their work and revise their strategies; guiding questions were linked to analysis and explanation; scaffolding supported inference, though its effect depended on students' algebraic manipulation; and reflection, as implemented, elicited self-evaluation least consistently. This feature-to-indicator mapping advances the understanding of how digital learning environments support critical thinking, reframing it as a process observable in students' work rather than inferred from outcome scores alone. Alongside this, the findings provide evidence of significant improvement in students' critical thinking performance following participation in the e-module while the absence of a non-intervention comparison group precludes definitive causal attribution. Several limitations qualify these conclusions. The single-group design without an external comparison group means the results indicate improvement accompanying the intervention rather than its exclusive cause. The study was confined to one topic within a single school, limiting generalizability to other mathematical content and settings. The qualitative findings, though supported by source triangulation and inter-rater agreement, remain subject to interpretive variability, and the assessment of feature "consistency" rested on the subset of observed students, which may not fully capture variation in how features were engaged across all tasks. These limitations point to clear directions for future work. Stronger designs—incorporating a control group or a staggered replication—would sharpen causal inference. Replication across other mathematical topics would test whether the identified feature-to-indicator processes hold across content areas. At the design level, guiding questions could be structured to require a claim-evidence-reasoning format and evaluated for their effect on explanation; scaffolding could be augmented with symbolic support targeting algebraic manipulation to prevent execution errors from undermining inference; and reflection could be structured around an error-cause-revision-verification sequence to deepen self-evaluation.

Acknowledgments

The authors would like to express their sincere gratitude to *RisetMu (Riset Muhammadiyah)*, Universitas Muhammadiyah Makassar, for providing financial support through the Internal Research Grant Program. The authors also thank everyone who contributed to completing this research.

Declarations

Author Contribution	: N: Conceptualization, Formal analysis, Investigation, Methodology, Writing - original draft, and Writing - review & editing; IA: Data curation, Methodology, Validation, and Writing - review & editing; MI: Data curation, Methodology, Validation, and Writing - review & editing; MM: Data curation, Investigation, Visualization, and Writing - review & editing; RK: Supervision, Validation, and Writing - review & editing.
Funding Statement	: This research was funded by <i>RisetMu (Riset Muhammadiyah)</i> Universitas Muhammadiyah Makassar for supporting and funding this research.

Conflict of Interest : The authors declare no conflict of interest.
 Additional Information : Additional information is available for this paper.

REFERENCES

- Amelia, R., Herman, T., Dasari, D., & Jupri, A. (2026). E-Miracle: An inquiry-realistic digital module to support students' 6C competencies. *Infinity Journal*, 15(2), 483–508. <https://doi.org/10.22460/infinity.v15i2.p483-508>
- Ayalon, M., & Wilkie, K. J. (2020). Investigating peer-assessment strategies for mathematics pre-service teacher learning on formative assessment. *Journal of Mathematics Teacher Education*, 24(4), 399–426. <https://doi.org/10.1007/s10857-020-09465-1>
- Blanton, M., Isler-Baykal, I., Stroud, R., Stephens, A., Knuth, E., & Gardiner, A. M. (2019). Growth in children's understanding of generalizing and representing mathematical structure and relationships. *Educational Studies in Mathematics*, 102(2), 193–219. <https://doi.org/10.1007/s10649-019-09894-7>
- Çakmak Gürel, Z. (2025). Indication of scaffolding in mathematical modeling. *International Journal of Science and Mathematics Education*, 23(7), 2597–2628. <https://doi.org/10.1007/s10763-025-10576-5>
- Çekmez, E. (2020). What generalizations do students achieve with respect to trigonometric functions in the transition from angles in degrees to real numbers? *The Journal of Mathematical Behavior*, 58, 100778. <https://doi.org/10.1016/j.jmathb.2020.100778>
- Chorney, S., Evans, K. R., & Staples, M. (2024). Conceptualizing reasoning practices in the context of sociomathematical issues. *The Journal of Mathematical Behavior*, 73, 101124. <https://doi.org/10.1016/j.jmathb.2024.101124>
- DiNapoli, J., & Miller, E. K. (2022). Recognizing, supporting, and improving student perseverance in mathematical problem-solving: The role of conceptual thinking scaffolds. *The Journal of Mathematical Behavior*, 66, 100965. <https://doi.org/10.1016/j.jmathb.2022.100965>
- El Mouhayar, R. (2018). Trends of progression of student level of reasoning and generalization in numerical and figural reasoning approaches in pattern generalization. *Educational Studies in Mathematics*, 99(1), 89–107. <https://doi.org/10.1007/s10649-018-9821-8>
- Elia, I. (2020). Word problem solving and pictorial representations: Insights from an exploratory study in kindergarten. *ZDM – Mathematics Education*, 52(1), 17–31. <https://doi.org/10.1007/s11858-019-01113-0>
- Foster, C., Woodhead, S., Barton, C., & Clark-Wilson, A. (2022). School students' confidence when answering diagnostic questions online. *Educational Studies in Mathematics*, 109(3), 491–521. <https://doi.org/10.1007/s10649-021-10084-7>
- Hadar, L. L., & Ruby, T. L. (2019). Cognitive opportunities in textbooks: the cases of grade four and eight textbooks in Israel. *Mathematical thinking and learning*, 21(1), 54–77. <https://doi.org/10.1080/10986065.2019.1564968>
- Hidayat, W., Aripin, U., & Widodo, S. A. (2025). Integration of ethno-modelling and 3N: An innovative digital worksheet framework to enhance students' mathematical critical thinking skills. *Infinity Journal*, 14(4), 1019–1042. <https://doi.org/10.22460/infinity.v14i4.p1019-1042>

- Huang, R., Zhang, Q., Chang, Y.-p., & Kimmins, D. (2019). Developing students' ability to solve word problems through learning trajectory-based and variation task-informed instruction. *ZDM – Mathematics Education*, 51(1), 169–181. <https://doi.org/10.1007/s11858-018-0983-8>
- Johar, R., Annisa, D., Mailizar, M., Zubainur, C. M., Zubaidah, T., Usman, U., Indriyani, L., & Musa, M. (2025). Development of algebraic expressions e-modules through realistic mathematics education approach. *Infinity Journal*, 14(4), 973–994. <https://doi.org/10.22460/infinity.v14i4.p973-994>
- Kertil, M., Erbas, A. K., & Cetinkaya, B. (2019). Developing prospective teachers' covariational reasoning through a model development sequence. *Mathematical thinking and learning*, 21(3), 207–233. <https://doi.org/10.1080/10986065.2019.1576001>
- Kim, N. J., Vicentini, C. R., & Belland, B. R. (2022). Influence of scaffolding on information literacy and argumentation skills in virtual field trips and problem-based learning for scientific problem solving. *International Journal of Science and Mathematics Education*, 20(2), 215–236. <https://doi.org/10.1007/s10763-020-10145-y>
- Klein, S., & Leikin, R. (2020). Opening mathematical problems for posing open mathematical tasks: what do teachers do and feel? *Educational Studies in Mathematics*, 105(3), 349–365. <https://doi.org/10.1007/s10649-020-09983-y>
- Kontorovich, I. (2021). Minding mathematicians' discourses in investigations of their feedback on students' proofs: a case study. *Educational Studies in Mathematics*, 107(2), 213–234. <https://doi.org/10.1007/s10649-021-10035-2>
- Kop, P. M. G. M., Janssen, F. J. J. M., Drijvers, P. H. M., & van Driel, J. H. (2020). The relation between graphing formulas by hand and students' symbol sense. *Educational Studies in Mathematics*, 105(2), 137–161. <https://doi.org/10.1007/s10649-020-09970-3>
- Lee, H. S., Sanei, H., Famularo, L., Masters, J., Bradshaw, L., & Schellman, M. (2023). Validating a concept inventory for measuring students' probabilistic reasoning: The case of reasoning within the context of a raffle. *The Journal of Mathematical Behavior*, 71, 101081. <https://doi.org/10.1016/j.jmathb.2023.101081>
- Lee, M. Y., & Lee, J.-E. (2021). Pre-service teachers' selection, interpretation, and sequence of fraction examples. *International Journal of Science and Mathematics Education*, 19(3), 539–558. <https://doi.org/10.1007/s10763-020-10062-0>
- Lehmann, T. H. (2023). Using algorithmic thinking to design algorithms: The case of critical path analysis. *The Journal of Mathematical Behavior*, 71, 101079. <https://doi.org/10.1016/j.jmathb.2023.101079>
- Leiss, D., Plath, J., & Schwippert, K. (2019). Language and mathematics - Key factors influencing the comprehension process in reality-based tasks. *Mathematical thinking and learning*, 21(2), 131–153. <https://doi.org/10.1080/10986065.2019.1570835>
- Li, R., Cevikbas, M., & Kaiser, G. (2024). Mathematics teachers' beliefs about their roles in teaching mathematics: orchestrating scaffolding in cooperative learning. *Educational Studies in Mathematics*, 117(3), 357–377. <https://doi.org/10.1007/s10649-024-10359-9>

- Llinares, S. (2019). What mathematics teachers do and know: looks from inside and outside of the classroom to the mathematic-specificity of teacher work. *Journal of Mathematics Teacher Education*, 22(3), 227–230. <https://doi.org/10.1007/s10857-019-09435-2>
- Martin, M., Syamsuri, S., Pujiastuti, H., & Hendrayana, A. (2021). Pengembangan e-modul berbasis pendekatan contextual teaching and learning pada materi barisan dan deret untuk meningkatkan minat belajar siswa SMP [Development of an e-module on sequences and series, based on the contextual teaching and learning approach, to enhance junior high school students' interest in learning]. *Jurnal Derivat: Jurnal Matematika dan Pendidikan Matematika*, 8(2), 72–87. <https://doi.org/10.31316/j.derivat.v8i2.1927>
- Mata-Pereira, J., & da Ponte, J.-P. (2017). Enhancing students' mathematical reasoning in the classroom: teacher actions facilitating generalization and justification. *Educational Studies in Mathematics*, 96(2), 169–186. <https://doi.org/10.1007/s10649-017-9773-4>
- Melhuish, K., Thanheiser, E., & Guyot, L. (2020). Elementary school teachers' noticing of essential mathematical reasoning forms: justification and generalization. *Journal of Mathematics Teacher Education*, 23(1), 35–67. <https://doi.org/10.1007/s10857-018-9408-4>
- Patterson, C. L., & McGraw, R. (2018). When time is an implicit variable: An investigation of students' ways of understanding graphing tasks. *Mathematical thinking and learning*, 20(4), 295–323. <https://doi.org/10.1080/10986065.2018.1509421>
- Pedersen, I. F., & Haavold, P. Ø. (2023). Students' mathematical beliefs and motivation in the context of inquiry-based mathematics teaching. *International Journal of Mathematical Education in Science and Technology*, 54(8), 1649–1663. <https://doi.org/10.1080/0020739x.2023.2189171>
- Rezat, S., Malik, S., & Leifeld, M. (2022). Scaffolding close reading of mathematical text in pre-service primary teacher education at the tertiary level: Design and evaluation. *International Journal of Science and Mathematics Education*, 20(S1), 215–236. <https://doi.org/10.1007/s10763-022-10309-y>
- Sandefur, J., & Manaster, A. B. (2022). Encouraging research on recursive thinking through the lens of a model of the spread of contagious diseases. *ZDM – Mathematics Education*, 54(4), 895–907. <https://doi.org/10.1007/s11858-022-01354-6>
- Scheibling-Sève, C., Pasquinelli, E., & Sander, E. (2020). Assessing conceptual knowledge through solving arithmetic word problems. *Educational Studies in Mathematics*, 103(3), 293–311. <https://doi.org/10.1007/s10649-020-09938-3>
- Scheiner, T., & Pinto, M. M. F. (2019). Emerging perspectives in mathematical cognition: Contextualizing, complementizing, and complexifying. *Educational Studies in Mathematics*, 101(3), 357–372. <https://doi.org/10.1007/s10649-019-9879-y>
- Shaughnessy, M., Boerst, T. A., & Farmer, S. O. (2019). Complementary assessments of prospective teachers' skill with eliciting student thinking. *Journal of Mathematics Teacher Education*, 22(6), 607–638. <https://doi.org/10.1007/s10857-018-9402-x>
- Soneira, C., González-Calero, J. A., & Arnau, D. (2018). An assessment of the sources of the reversal error through classic and new variables. *Educational Studies in Mathematics*, 99(1), 43–56. <https://doi.org/10.1007/s10649-018-9828-1>

- Susanta, A., Sumardi, H., Susanto, E., & Retnawati, H. (2023). Mathematics literacy task on number pattern using Bengkulu context for junior high school students. *Journal on Mathematics Education*, 14(1), 85–102. <https://doi.org/10.22342/jme.v14i1.pp85-102>
- Swidan, O., Bagossi, S., Beltramino, S., & Arzarello, F. (2022). Adaptive instruction strategies to foster covariational reasoning in a digitally rich environment. *The Journal of Mathematical Behavior*, 66, 100961. <https://doi.org/10.1016/j.jmathb.2022.100961>
- Toulmin, S. E. (2003). *The uses of argument* (2 ed.). Cambridge University Press. <https://doi.org/DOI: 10.1017/CBO9780511840005>
- Utami, T. N., Jatmiko, A., & Suherman, S. (2018). Pengembangan modul matematika dengan pendekatan science, technology, engineering, and mathematics (STEM) pada materi segiempat [Development of a mathematics module using the science, technology, engineering, and mathematics (STEM) approach for the topic of quadrilaterals]. *Desimal: Jurnal Matematika*, 1(2), 165–172. <https://doi.org/10.24042/djm.v1i2.2388>
- Vicente, S., Verschaffel, L., Sánchez, R., & Múñez, D. (2022). Arithmetic word problem solving: Analysis of Singaporean and Spanish textbooks. *Educational Studies in Mathematics*, 111(3), 375–397. <https://doi.org/10.1007/s10649-022-10169-x>
- Voica, C., Singer, F. M., & Stan, E. (2020). How are motivation and self-efficacy interacting in problem-solving and problem-posing? *Educational Studies in Mathematics*, 105(3), 487–517. <https://doi.org/10.1007/s10649-020-10005-0>
- Wilkie, K. J. (2020). Investigating students' attention to covariation features of their constructed graphs in a figural pattern generalisation context. *International Journal of Science and Mathematics Education*, 18(2), 315–336. <https://doi.org/10.1007/s10763-019-09955-6>
- Wilkie, K. J. (2021). Seeing quadratics in a new light: secondary mathematics pre-service teachers' creation of figural growing patterns. *Educational Studies in Mathematics*, 106(1), 91–116. <https://doi.org/10.1007/s10649-020-09997-6>
- Xin, Y. P., Park, J. Y., Tzur, R., & Si, L. (2020). The impact of a conceptual model-based mathematics computer tutor on multiplicative reasoning and problem-solving of students with learning disabilities. *The Journal of Mathematical Behavior*, 58, 100762. <https://doi.org/10.1016/j.jmathb.2020.100762>
- Yao, X. (2020). Unpacking learner's growth in geometric understanding when solving problems in a dynamic geometry environment: Coordinating two frames. *The Journal of Mathematical Behavior*, 60, 100803. <https://doi.org/10.1016/j.jmathb.2020.100803>
- Yao, X. (2022). Representation transformations in transition from empirical to structural generalization. *The Journal of Mathematical Behavior*, 66, 100964. <https://doi.org/10.1016/j.jmathb.2022.100964>
- Yasin, M., Jauhariyah, D., Madiyo, M., Rahmawati, R., Farid, F., Irwandani, I., & Mardana, F. F. (2019). The guided inquiry to improve students mathematical critical thinking skills using student's worksheet. *Journal for the Education of Gifted Young Scientists*, 7(4), 1345–1360. <https://doi.org/10.17478/jegys.598422>
- Yoon, H., & Thompson, P. W. (2020). Secondary teachers' meanings for function notation in the United States and South Korea. *The Journal of Mathematical Behavior*, 60, 100804. <https://doi.org/10.1016/j.jmathb.2020.100804>

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2