

Addressing symbolic obstacles in undergraduate calculus through artificial intelligence scaffolding

Isbadar Nursit^{1,2} , Toto Nusantara^{3*} , Erry Hidayanto³ 

¹ Doctoral Student of Mathematics Education Department, State University of Malang, INDONESIA

² Mathematics Education Study Program, Universitas Islam Malang, East Java, INDONESIA

³ Department of Mathematics Education, State University of Malang, East Java, INDONESIA

*Corresponding Author: toto.nusantara.fmipa@um.ac.id

Citation: Nursit, I., Nusantara, T., & Hidayanto, E. (2026). Addressing symbolic obstacles in undergraduate calculus through artificial intelligence scaffolding. *Pedagogical Research*, 11(4), em0295. <https://doi.org/10.29333/pr/19631>

ARTICLE INFO

Received: 14 Oct. 2025

Accepted: 16 Aug. 2026

ABSTRACT

Symbolic obstacles in calculus prevent students from moving away from intuition and towards formal thinking, but are rarely studied in the context of artificial intelligence (AI) interaction logs. We investigated symbolic obstacles and AI-supported representational transitions among 12 Indonesian undergraduate mathematics education students in a non-intrusive intrinsic case study. Semi-structured interviews, eight classroom observations, diagnostic tests, academic artifacts, and AI logs were analyzed through inductive-deductive thematic analysis, based on Tall's Three Worlds of Mathematics. Four obstacles are highlighted: graph-to-symbol translation, integral and sigma notation, limit notation, and symbolic algebraic manipulation. Participant-level comparisons showed 27.4%–53.8% fewer coded symbolic errors in post-activity assessments than at baseline. AI scaffolding through dynamic visualization, adaptive feedback, step-by-step guidance, and solution validation was associated with representational cycling, metacognitive reflection, reported self-efficacy gains, and lower mathematics anxiety. The study integrates AI-log evidence with the Three Worlds framework to explain how scaffolding can support students in addressing symbolic obstacles.

Keywords: artificial intelligence, calculus, digital scaffolding, representational fluency, symbolic obstacles, three worlds of mathematics

INTRODUCTION

University calculus is a demanding subject for higher education because it is expected to help students develop conceptual reasoning but it is also a gateway to study science, technology, engineering and mathematics. Reviews of undergraduate calculus education suggest that calculus courses can serve as a platform to push students to the next level of learning but also filter out participation when nothing really meaningful is happening and there is no fundamental knowledge (Biza et al., 2022). The obstacles in moving from school to university mathematics are not just the implementation of rules but also in their language and form of justification and expectations to coordinate visual, numerical, verbal and symbolic representations (Di Martino et al., 2023). Mathematical notation can also be a source of structure or an impediment when students use symbols but do not relate them to the quantities and relationships that they represent.

These difficulties can be seen in many calculus tasks. Students may recognize the geometric idea of an area under a curve but struggle to interpret the role of dx , determine the meaning and orientation of integral bounds, or relate a Riemann sum to integral notation. Research with university students has documented unstable interpretations of differentials and definite integrals, including difficulties coordinating symbolic expressions with graphical and contextual meanings (Bajracharya et al., 2023; Nilsen & Knutsen, 2023). Similar instability occurs in derivative learning when students can apply differentiation rules but cannot adequately explain the derivative in terms of covariation, local rate of change, slope, or graphical behavior (Aniswita et al., 2024; Thompson & Harel, 2021). These patterns mean that a wrong symbol is not a single transcription or procedural error. It may reveal much more about the fact that there is an underlying disruption in the process of creating and recording mathematical meaning across representations.

We think of these types of disruptions as symbolic obstacles. Mathematical understanding from a semiotic perspective, in other words, is about coordinating and changing representations rather than operating within a single representational register (Duval, 2017). In other words, representational fluency is not just about finding a mathematical object of the same form but about forming various forms. It is about students building, interpreting, selecting, and connecting representations while keeping the mathematical relationships that are universal to them (Fonger, 2019). In Indonesian mathematics education studies, students

have also found compartmentalized understanding when working with a variety of representations and not preserving semantic congruence in translating between symbolic, verbal, and graphical representations (Afriyani et al., 2018; Nurrahmawati et al., 2021). The external visualization literature also emphasizes that graphs, diagrams, and dynamic displays are pedagogically useful when students actively connect them with mathematical objects and inferences expressed in symbols and language (Schoenherr & Schukajlow, 2024). This is why students will know the spatial and visual tendency of a function, an area image, or a rapidly changing slope but can't express that in the terms of limit, derivative, integral, or sigma notation.

Tall's Three Worlds of Mathematics provides an epistemological framework to locate these representational challenges. Mathematical activity as a result of perception and action is embodied, symbolic activity through processes and manipulable signs are defined, and formal activity involves definitions, properties, generalization, and deductive reasoning (Tall, 2008). Embodied cognition research also shows that mathematical meaning can be determined by the coordination of perception and action, technological tools, and symbolic systems rather than purely symbolic manipulation by the imagination (Halák, 2022; Shvarts et al., 2021). Mathematical worlds are not a fixed set of things to learn, and students can move in and out of them or draw from more than one universe at once. But symbolic expression can be just as important in calculus. If notation does not preserve the meaning of graphs and actions as well as the language we use for verbal expression, then mathematical reasoning into a formalistic explanation can be at a standstill.

Digital and artificial intelligence (AI) tools can help students maintain those transitions when they help students connect representations without taking away the time to compute with math. Connected visualizations can reflect change simultaneously in visual and symbolic form, and technology-supported environments can allow students to explore calculus relationships and to see gaps in understanding (Chytas et al., 2024; Li et al., 2022). Recent classroom research on AI in calculus has found a positive impact on explanation, engagement, feedback, and individualized assistance, but this same research also points out that good instructional design and careful guidance are needed (Serhan & Welcome, 2024; Torres-Peña et al., 2024). Interactive feedback is also of particular importance if it looks at a problematic symbolic step and points out the inconsistency in the representation, and students are encouraged to correct their own representation and not just accept the final solution (Barana et al., 2021). AI can provide a temporary digital scaffolding that students can check for relationships and the differences that they might have missed.

Despite these advances, there are still three gaps. First, there are few studies on problems in calculus where we can find symbolic errors in students' transitions from embodied, symbolic, and formal thinking. It is not clear whether a given error is due to an unstable intuitive meaning, a translation to notation, an invalid symbolic shift, or a lack of formalization. Second, research on AI in calculus has mainly focused on students' perceptions, general experience in the classroom, or task performance. AI interaction logs are not widely used as process data to show the students' requests for help or feedback, how students change their symbolic work, and how they use different representations as they tackle a problem. Third, the instructors need to understand how dynamic visualization, adaptive feedback, step-by-step guidance, and solution validation can be organized to facilitate representational transitions without entailing reliance on automated answers. These gaps are theoretical as well as practical because they concern both the explanation of symbolic obstacles and the design of instruction for addressing them.

Accordingly, this exploratory intrinsic case study investigates symbolic obstacles among undergraduate mathematics education students at a university in Malang, East Java, Indonesia, and examines how AI-based scaffolding is associated with students' attempts to address those obstacles. The study integrates Tall's Three Worlds of Mathematics with representational fluency and embodied cognition to identify where representational breakdowns occur and to interpret how students respond to dynamic visualization, adaptive feedback, step-by-step guidance, and validation. Three research questions guide the study: What types of symbolic obstacles are evident in students' work on limits, derivatives and integrals? How do these symbolic obstacles shape students' transitions between embodied, symbolic and formal mathematical thinking? How do AI-supported visualization, feedback, guidance and validation support students' representational and strategic responses to symbolic obstacles?

The study provides a holistic view of symbolic obstacles by answering diagnostic questions, conducting interviews, observing in the classroom and providing academic artifacts and AI interaction logs. Its contribution is not to argue that AI causes conceptual development, but to show how specific types of scaffolding can make representational relationships, symbolic inconsistencies and revision processes visible to learners and researchers. The conceptual and theoretical connections that are the basis of this analysis are described in the next section and in **Figure 1**.

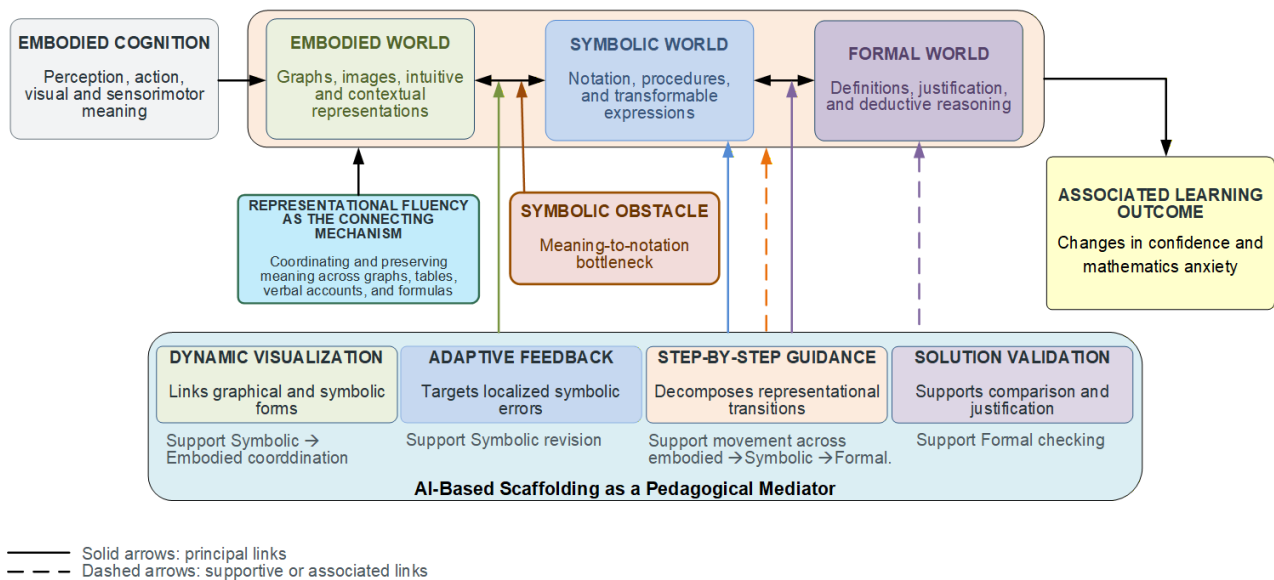


Figure 1. Conceptual framework for addressing symbolic obstacles through AI scaffolding (Source: Authors' own elaboration, based on Tall (2008) and the study's conceptual framework)

CONCEPTUAL AND THEORETICAL FRAMEWORK

In this work, we distinguish between its theoretical lens and conceptual model. Tall's Three Worlds of Mathematics is the primary theoretical framework, and representational fluency and embodied cognition further solidify how mathematical meaning is constructed and organized. Artificial intelligence (AI)-based scaffolding is the teaching mechanism in the conceptual model, not the other mathematical world.

Tall's Three Worlds of Mathematics

Tall (2008) distinguishes between mathematical thinking in the embodied, symbolic, and formal worlds. These three approaches together encapsulate how undergraduate students experience symbolic hurdles when they manage to juggle intuitively, symbolically, and formally based calculus reasoning. The embodied world is formed by perception, imagery, spatial connections, action, and experience with mathematical objects. The symbolic world is the notation and operations that describe processes and the objects produced by those processes. The formal world is formed by definitions, properties, generalization, and deductive reasoning. These worlds are not rigid developmental stages: students may move recursively in one world and use more than one world at once. They can face difficulties when meaning that is set in one world is not preserved in another.

In undergraduate calculus, there are graphs such as changing quantities, areas, slopes, motion, and visual approximation to support embodied reasoning. Limit, derivative, integral, sigma, and differential notation, and algebraic transformations are all symbolic activities. Formal reasoning is apparent when students justify a limit, explain the relation between differentiation and integration, or connect symbolic procedures to definitions and theorems. In a study of the basic calculus concepts, it is found that what makes learning difficult for students is the meanings and ways of thinking they bring to these concepts in general and not only procedural errors (Thompson & Harel, 2021). Differentials and definite integrals are also studied to show that students can perform some procedures with their different symbols dx , limits, and integral notation being different (Nilsen & Knutsen, 2023).

Analytically, this study treats the symbolic world as a critical hinge. The student might picture a graphical tendency or accumulate but cannot express that in a reasonable calculus notation. Tall's framework shows where a transition difficulty is and represents the way in which a connection is being made.

Representational Fluency as the Connecting Mechanism

Representational fluency is the ability to produce, interpret, select, coordinate and translate between representations with mathematical meaning. It is not just about displaying the same object in many ways. Fluency is clear when graphs, tables, verbal accounts, numerical patterns, and symbolic expressions indicate structurally consistent relationships (Fonger, 2019). External representations are the ones that help students learn and coordinate them better in terms of reasoning as opposed to just being examples (Schoenherr & Schukajlow, 2024). Technology-based environments can make this coordination even stronger when multiple representations are connected, and the students need to explain their correspondence (Li et al., 2022).

Representational fluency is the key to linking the three mathematical worlds. Reading the behavior of a graph near a point requires some embodied reasoning; to translate this behavior into a limit requires some symbolic encoding; to explain why the notation gives a definition is a formal one. Symbolic obstacles arise when the semantic coherence is lost along this path. They may be graph-to-symbol errors, a misunderstanding of dx , incorrect integral bounds or sigma indices, unstable limit notation, or transformations that alter the meaning of an expression.

Embodied Cognition as a Basis for Meaning Construction

Embodied cognition describes how mathematical meaning can be understood in relation to perception, action, gesture, spatial organization, and interaction with artifacts. Mathematical thinking is not limited to internal manipulation of symbols. Embodied instrumentation is the study of learning in which we study the learner's actions, perceptual attention, and technological artifacts (Shvarts et al., 2021). In calculus, tracing a curve, manipulating a dynamic point, comparing local slopes, or observing increasingly fine partitions might make variation, rate, and accumulation perceptually possible.

These experiences don't automatically lead to symbolic or formal understanding. Students might notice a visual pattern without being able to encode or justify it mathematically. Embodied activity is thus viewed as an underlying source of meaning that needs to be synchronized with language and notation, not as a preliminary stage that is simply replaced. This approach is consistent with embodied learning approaches that emphasize purposeful relationships between action, visual attention, representation, language, and reflection (Way & Ginns, 2024).

AI-Based Scaffolding as a Pedagogical Mediator

AI-based scaffolding is conceptualized as contingent and gradually fading support for movement between embodied, symbolic and formal reasoning. Contingency and fading of responsibility are the key characteristics of scaffolding, as the support should be responsive to the current needs of a learner and less so as independent performance develops (van de Pol et al., 2010). In this study, we implemented AI support through dynamic visualization, adaptive feedback, step-by-step guidance and solution validation.

These mechanisms have different representational properties. Dynamic visualization may make variation and invariants perceptually available. Adaptive feedback can highlight a particular symbolic inconsistency rather than simply mark a final answer as wrong. Step-by-step guidance enables you to break down a representation change into actionable steps, and validation can help compare the learner's work and digitally generated output. Interactive feedback is most effective when it is specific to the learner's current step and supports revision rather than replacing the learner's reasoning (Barana et al., 2021). Reviews of AI in mathematics education found the involvement of individualized support, feedback and diagnosis in the pedagogical practice of AI (Hwang & Tu, 2021). Calculus-based classroom studies also show that digital and AI-enabled activities can increase student engagement with changing quantities and multiple representations in structured educational designs (Chytas et al., 2024; Torres-Peña et al., 2024).

AI scaffolding is therefore not assumed to lead to conceptual development independently. It can support representational coordination and metacognitive regulation, but too much guidance can stifle productive struggle or induce dependency. The framework thus views fading and independent symbolic production as necessary conditions to regard AI support as scaffolding, not answering.

Integrated Analytical Proposition

The integrated framework proposes that calculus understanding develops through recursive coordination among embodied, symbolic, and formal worlds. Embodied cognition explains how meanings are grounded; representational fluency explains how those meanings are translated and preserved; symbolic obstacles identify disruptions in this coordination; and AI-based scaffolding provides temporary support at the point of disruption.

The framework guides analysis of: (a) the mathematical world in which a student's reasoning is situated; (b) the representational transition being attempted; (c) the symbolic obstacle that disrupts the transition; (d) the form of AI scaffolding provided; and (e) subsequent evidence of greater representational coherence, strategic independence, or formal justification. It does not presume linear transitions or a causal treatment effect. Changes in confidence and mathematics anxiety are interpreted as associated learning responses rather than constitutive elements of the theoretical framework.

MATERIALS AND METHODS

Research Design and Case Selection

This study employed an exploratory qualitative intrinsic case-study design to investigate how symbolic obstacles emerged and changed during an artificial intelligence-supported undergraduate calculus learning sequence (Creswell & Poth, 2018; Sneed et al., 2020). The case was bounded by one mathematics education program at a university in Malang, East Java, Indonesia; twelve participating students; the topics of limits, derivatives, and integrals; and eight observed learning sessions. A qualitative case study was appropriate because the study sought an in-depth, context-sensitive explanation of students' representational activity rather than statistical generalization. Recent methodological guidance emphasizes that case-study quality depends on clearly bounding the case, defining the unit of analysis, and explaining why the selected setting is information-rich for the research purpose (Miller et al., 2023; Mtisi, 2022).

The university was selected because it provided access to early-stage mathematics education students who were taking or had recently completed Calculus I or II and were therefore actively negotiating transitions among visual, symbolic, and formal representations. The setting also enabled the same students to be followed across diagnostic assessment, AI-supported activities, interviews, classroom observations, and post-assessment. The case is not presented as statistically representative of Indonesian universities; rather, its instructional and sociocultural boundaries are reported to support judgments of transferability.

Table 1. Data sources and their analytical contributions

Data source	Procedure	Primary analytical contribution
Written diagnostic assessments	Pre- and post-assessments consisting of limits, derivatives, and integrals tasks	Identification of symbolic errors, representational coherence, and descriptive changes in error rates
Semi-structured interviews	Individual sessions lasting approximately 45–60 minutes	Students' interpretations of symbols, representational transitions, AI support, and affective responses
Non-participant observations	Eight calculus learning sessions	Students' use of notation, responses to prompts, representation switching, and interaction with AI
Academic artifacts	Assignments, notes, assessment responses, and relevant course documents	Corroboration of recurring symbolic-error patterns
AI interaction logs	Recorded prompts, hint levels, responses, revisions, and outputs	Identification of scaffolding mechanisms and changes in students' representational activity

Participants

Participants were twelve mathematics education students in semesters 1–3, selected through purposive sampling. Inclusion criteria were being enrolled in or having recently completed Calculus I or II, demonstrating varied forms of symbolic difficulty in preliminary work, being able to describe their learning experiences, and providing informed consent. Students who had completed advanced calculus courses or had substantial absences during the observed learning sequence were excluded. Participant identities were replaced with codes P1–P12 in the analytical dataset and with corresponding pseudonyms in qualitative excerpts.

The sample size was considered appropriate for the intrinsic purpose of examining variation and commonality within a bounded case. No claim of population-level generalizability was made. Instead, detailed descriptions of the participants, tasks, learning sequence, and institutional context were retained to enable readers to determine the applicability of the findings to comparable settings.

Data Sources and Research Procedure

Data were collected from five complementary sources: written diagnostic assessments, semi-structured interviews, non-participant classroom observations, academic artifacts, and AI interaction logs. The multiple evidence sources enabled symbolic steps, verbal explanations, observed behavior, and AI-supported revisions to be examined together.

The research process took six stages: case selection, pre-assessment, initial diagnosis, an eight-session AI-supported learning sequence, post-assessment, and integrated analysis. We selected 12 undergraduate mathematics education students from a university in Malang, East Java, with a bounded calculus learning context. We asked the students first to write a pre-assessment in which they obtained written answers that were used for semi-structured interviews and the review of relevant academic artifacts. See **Table 1** for data sources and their analytical contributions.

During the eight-session learning sequence, the AI acted as a scaffold through three levels of support: intuitive or visual prompts, symbolic or procedural guidance, and formal justification or verification. Higher-level suggestions were provided only when earlier support was not sufficient. Non-participant observations and AI interaction logs summarized students' notation use, representational changes, responses to feedback, prompts, outputs, and revisions.

The participants then completed a post-assessment in the same format, time, and procedures as the pre-assessment. AI or mathematical software was only allowed for certain kinds of graphing and solution-validation tasks. Responses (pre- and post-assessment) were compared with the same symbolic error rubric, and all assessments, interviews, observations, artifacts, AI logs, and reflections were combined for NVivo-supported thematic analysis and mapping to Tall's embodied, symbolic, and formal worlds.

Figure 2 shows the stages in order. The data and the initial interpretation were very close to each other in eight AI-supported sessions, with observations, student artifacts, and AI interaction logs being reviewed to fine-tune subsequent interview questions and to identify areas of need for further analysis. After the post-assessment, all of the materials were collected into a qualitative corpus which included diagnostic results, interview transcripts, observation records, academic artifacts, AI interaction logs, and written reflections. This corpus was then subjected to the iterative inductive-deductive analysis described in qualitative data analysis.

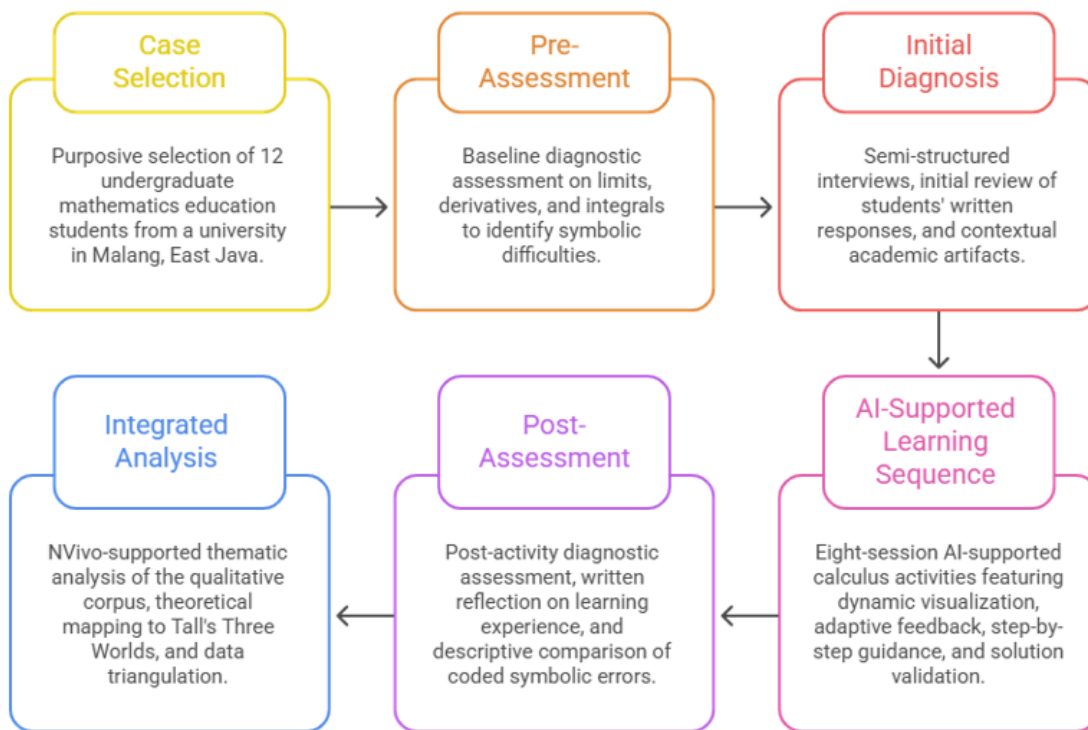


Figure 2. Research procedure from case selection and diagnostic assessment to AI-supported learning and integrated analysis (Source: Authors' own elaboration)

Diagnostic Instrument and Symbolic-Error Metrics

The written diagnostic instrument comprised three problems with sub-items a–c and required approximately 30–40 minutes. The limit task assessed algebraic manipulation, graphical estimation, and the relationship between graphical behavior and the formal meaning of a limit. The derivative task assessed symbolic differentiation, graphical interpretation, and the relationship between f and f' . The integral task assessed manual integration, AI-supported validation, and interpretation of dx in relation to accumulation and area.

Two complementary scoring layers were applied. First, each diagnostic sub-item received a level-of-evidence score: 2 for a coherent response, 1 for a partial response, and 0 for an incorrect or unsupported response. These scores supported the qualitative interpretation of transitions among embodied, symbolic, and formal reasoning.

Second, individual symbolic representation or transformation steps were assessed using the symbolic-error rubric. Steps at rubric Levels 0–1 were classified as errors, whereas Levels 2–3 were classified as non-errors. The unit of analysis was therefore one scorable symbolic step per problem per participant. Because students produced different numbers of scorable transformations, the denominators varied slightly across participants and assessment occasions.

$$ER_{it} = \frac{(\text{number of incorrect symbolic steps}_{it})}{(\text{number of evaluated symbolic steps}_{it})}$$

Where ER denotes error rate, i denotes the participant and t denotes the pre- or post-assessment. Percentage Reduction (%) was calculated as:

$$PR_i = \frac{(ER_{i,pre} - ER_{i,post})}{ER_{i,pre}} \times 100\%$$

These values were used only as descriptive indicators embedded within the qualitative case study. No inferential significance test or causal-effect estimate was calculated because the study did not include random assignment or a comparison group.

Qualitative Data Analysis

The qualitative corpus comprised interview transcripts, observation records, open-ended diagnostic explanations, academic artifacts, and AI interaction logs. Audio-recorded interviews were transcribed verbatim and checked against the recordings. The researchers then repeatedly read the complete corpus to identify units of meaning related to symbolic obstacles, representational transitions, AI scaffolding, adaptive strategies, and affective responses.

We used a hybrid inductive-deductive and codebook-oriented thematic analysis. This analysis was codebook-oriented, not purely reflexive due to the combination of codes based on theoretical derivation, an operational rubric, coders, and frequency summaries at the level of participants. The thematic analysis literature often stresses the importance of identifying a particular thematic-analytical approach and keeping its epistemological assumptions and codebook procedures consistent as well as the role of coder and theme development (Braun & Clarke, 2021; Byrne, 2022; Kiger & Varpio, 2020).

The analysis involved five phases. Open coding first took into account data-near meanings, including confusion with notation, symbolic transformations that did not work, graph-to-symbol disconnections, AI suggestions, and reflective checking. In axial coding, related open codes were compared and grouped into broader categories, such as difficulty of graph-to-symbol translation, notation interpretation, tiered guidance, representational cycling, and metacognitive reflection. Selective coding then brought these categories into higher-level themes and related them to embodied, symbolic, and formal mathematical work.

In **Figure 3**, we show how representative excerpts are translated from data-near open codes into axial categories, selective themes, and theoretical interpretations. We chose the examples to show how the analytical logic was applied in the wider corpus

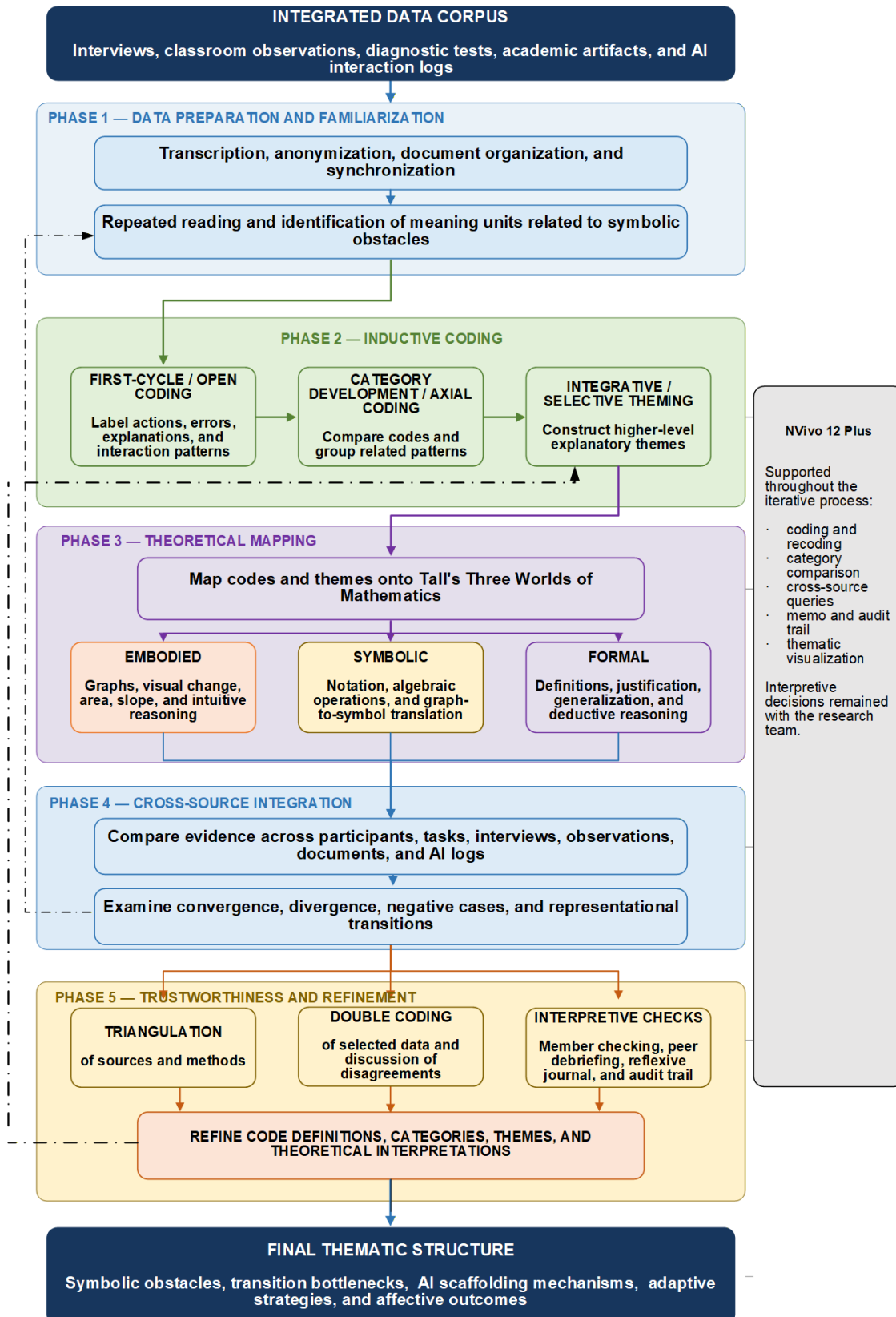


Figure 3. Illustrative NVivo coding excerpts and data-to-theme mapping (Source: Authors' own elaboration)

Table 2. Illustrative NVivo coding excerpts and data to theme mapping

Raw data excerpt	Open code	Axial category	Selective theme	Three Worlds interpretation
"When I look at the graph, I understand what it means. But when I use the notation Σ or integral, I am confused about what it means." (Student-03)	Visual meaning understood but not encoded symbolically	Graph-to-symbol translation difficulty	Symbolic obstacle as a representational bottleneck	Breakdown in the Embodied $\rightarrow$ Symbolic transition
"This AI gives me the steps, so I can see from the graph to the formula, then to the answer." Student-05)	Staged connection among representations	Tiered AI guidance	AI scaffolding supporting representational transition	AI-mediated Embodied $\rightarrow$ Symbolic movement
"If I'm wrong, the AI tells me where I went wrong. So I try again in a different way." (Student-02)	Error identification followed by revision	Recursive consultation and reflective checking	Adaptive and metacognitive response	Symbolic consolidation supported by feedback

and not the complete codebook. The first pathway illustrates how visual meaning is lost in the process of moving from embodied to symbolic reasoning. The second shows how tiered AI guidance supported movement between representations, while the third illustrates how feedback made recursive consultation and reflective revision possible.

Open coding first captured data-near meanings like notation confusion, invalid symbolic transformations, disconnections from graph to symbol, questions about additional AI clues, and reflective checking. In axial coding, other open codes were grouped into broader categories like graph-to-symbol translation, notation interpretation, tiered guidance, representation cycling, and metacognitive reflection. Selective coding then connected these categories with the main themes of the code and mapped them into embodied, symbolic, and formal work.

NVivo 12 Plus was used to organize data, maintain the codebook, retrieve coded excerpts, construct participant-by-code matrices, and generate thematic visualizations. NVivo reference counts were interpreted as the prevalence of coded instances within the corpus, not as effect sizes, statistical correlations, or independent measures of learning outcomes (See [Table 2](#)).

Trustworthiness and Research Ethics

Credibility was supported through triangulation across the five data sources, comparison of pre- and post-assessment evidence, and member checking of concise interpretive summaries (Birt et al., 2016). At least 25% of the qualitative corpus was independently coded by a second researcher. Coding differences were discussed, operational definitions were refined, and final decisions were documented (O'Connor & Joffe, 2020). Peer debriefing with mathematics education researchers, reflexive journaling, and an audit trail of instrument versions, coding decisions, merged or separated categories, and theme development were used to strengthen dependability and confirmability (Korstjens & Moser, 2018; Nowell et al., 2017). A thick description of the institutional context, participants, tasks, and AI-supported sequence supported transferability. These procedures correspond with recent guidance emphasizing triangulation, member engagement, transparent contextual description, reflexivity, and audit trails as complementary components of qualitative trustworthiness (Ahmed, 2024).

RESULTS

Our interviews, classroom observations, diagnostic responses and academic artifacts along with AI interaction logs led to four interrelated results: the distribution of symbolic obstacles, the locations of transition bottlenecks in Tall's three worlds of mathematics, the changes in coded symbolic errors in the participants' spaces, and the mechanisms and learning responses for AI-based scaffolding. Numerical summaries are given as descriptive data within the qualitative case study.

Symbolic Obstacle Categories

The types of symbolic obstacles were identified in our data set: graph-to-symbol translation, integral and sigma notation, limit notation, and symbolic algebraic manipulation. Graph-to-symbol translation was the most commonly used and most popular of the four types of symbolic obstacles in our data set: 38 coded references for all 12 participants. The remaining categories were present in most cases, too: integral and sigma notation in 11 participants, limit notation in 10, and symbolic algebraic manipulation in nine.

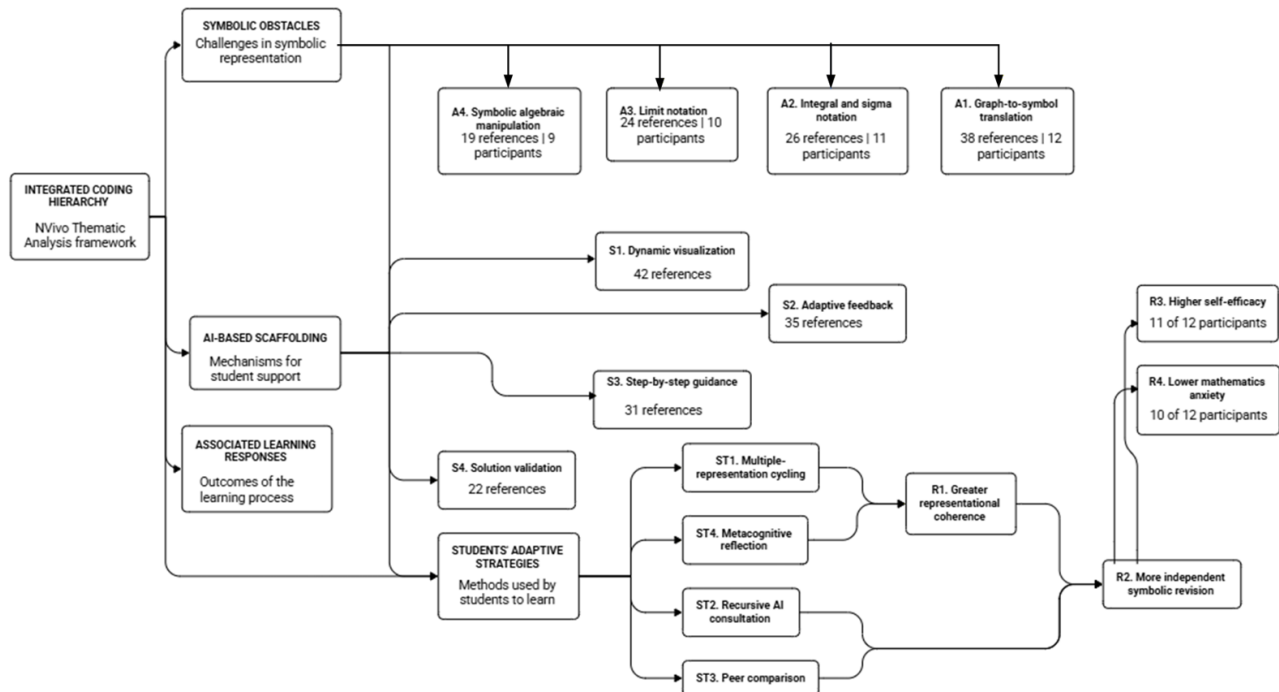
The dominant pattern was not a complete lack of intuition. Participants would often describe the visual tendency of a graph, an area under a curve, a change in slope, but their written expressions did not reflect such elements. This could be missing or misplaced bounds, confused dx and sigma indices, confusion between a function and a limit, to algebraic transformations that do not satisfy. Student-03 summarized this separation between visual recognition and symbolic encoding as:

"When I look at the graph, I understand what it means. But when I use the sigma or integral notation, I am confused about what it means." (Student-03)

Table 3 shows integrated coding summary of symbolic obstacles. Note that NVivo references indicate data that is coded and should not be taken as independent observations. Participant coverage was available for the obstacle categories.

Table 3. Integrated coding summary of symbolic obstacles

Domain	Category / mechanism	Operational evidence in the data	NVivo references	Participant coverage
Symbolic obstacle	Graph-to-symbol translation	Difficulty converting graphical or intuitive meaning into valid calculus notation.	38	12/12
Symbolic obstacle	Integral and sigma notation	Confusion about dx, integral bounds, sigma indices, or the relation between accumulation and notation.	26	11/12
Symbolic obstacle	Limit notation	Confusion about approaching versus equaling, two-sided behavior, or formal control.	24	10/12
Symbolic obstacle	Symbolic algebraic manipulation	Invalid transformation, simplification, substitution, or use of an equality sign.	19	9/12

**Figure 4.** Integrated coding hierarchy derived from NVivo coding and mapped to Tall's Three Worlds of Mathematics (Source: Authors' own elaboration, based on NVivo coding and Tall (2008))

The Embodied-to-Symbolic Transition as the Principal Bottleneck

The obstacles converged from embodied or intuitive reasoning to symbolic representation at the same time. Participants were generally more successful when they could view a graph, describe a changing quantity, or refer to an area or slope. The problem became harder when the meaning was encoded in limit, derivative, integral, or summation notation. Student-07 addressed this transition problem directly:

"I can imagine the area under the curve, but when asked to write it down as an integral, I do not know where to start." (Student-07).

This bottleneck also prevented us from moving towards formal explanations. Formal explanations, such as connecting graphical behavior to an appropriate definition, justifying a derivative relationship, or explaining accumulation based on a theorem, appeared in only three of the 12 participants in the relevant diagnostic and interview segments. The other participants typically remained at an embodied explanation or produced a symbolic procedure without a stable conceptual justification.

Figure 4 integrates the coding hierarchy with Tall's Three Worlds framework. The obstacle categories are concentrated in the symbolic world, while AI-supported visualization, feedback, guidance, and validation are positioned at the points where students attempted to coordinate embodied, symbolic, and formal reasoning.

Descriptive Changes in Coded Symbolic Errors

Within-participant comparisons showed fewer coded symbolic-error occurrences in the post-activity materials for all 12 participants. Percentage reductions ranged from 27.4%–53.8%, with a mean reduction of 42.3% and a median of 40.4%. Six participants were classified in the high-reduction group (>40%), and six were classified in the moderate-reduction group (25%–40%); no participant was classified in the low-reduction group (<= 25%). The largest reductions were observed for Student-06 (53.8%), whereas Student-08 showed the smallest reduction (27.4%) (**Table 4**).

Table 4. Participant level comparison of coded symbolic error occurrences

ID	Pre-Step (K_pre)	Pre Error	ER Pre	Post Step (K_post)	Post Error	ER Post	PR %	Group
Student-01	14	9	0.643	15	6	0.400	37.8%	Moderate
Student-02	13	8	0.615	14	5	0.357	42.0%	High
Student-03	12	7	0.583	14	5	0.357	38.8%	Moderate
Student-04	12	6	0.500	13	4	0.308	38.5%	Moderate
Student-05	15	9	0.600	16	6	0.375	37.5%	Moderate
Student-06	12	6	0.500	13	3	0.231	53.8%	High
Student-07	13	7	0.538	14	4	0.286	46.9%	High
Student-08	14	9	0.643	15	7	0.467	27.4%	Moderate
Student-09	12	6	0.500	12	3	0.250	50.0%	High
Student-10	13	8	0.615	13	4	0.308	50.0%	High
Student-11	12	6	0.500	13	4	0.308	38.5%	Moderate
Student-12	13	7	0.538	14	4	0.286	46.9%	High

Table 5. NVivo coding references and representative evidence for AI scaffolding mechanisms

Code and NVivo node	Coding references, <i>n</i>	Principal data sources	Representative evidence or observed pattern	Analytical interpretation
S1. Dynamic visualization	42	AI logs, classroom observations, diagnostic tasks	Students inspected graphical and symbolic representations within the same problem-solving sequence.	Made relationships across embodied and symbolic representations available for comparison.
S2. Adaptive feedback	35	AI logs, interviews	Students revised localized symbolic errors after feedback identified the problematic step.	Supported error monitoring and revision within the symbolic world.
S3. Tiered step-by-step guidance	31	AI logs, interviews	Guidance progressed from intuitive or visual prompts to symbolic procedures and requests for formal explanation.	Mediated movement across embodied, symbolic, and formal reasoning.
S4. Solution validation	22	AI logs, written reflections	Students compared manual and AI-supported solutions and examined whether discrepancies represented errors or equivalent forms.	Supported verification and metacognitive evaluation of symbolic work.

Note. PR = percentage reduction in coded symbolic errors, calculated as = (pre-activity coded errors - post-activity coded errors) / pre-activity errors x 100%. Groups: High >40%; Moderate (25%-40%]; Low <=25%. ER (Error Rate), defined as the number of coded errors divided by the number of evaluated symbolic steps, is a separate metric and is not reported in this column.

AI Scaffolding as Representational and Procedural Support

The AI interaction evidence pointed to four common forms of support. Dynamic visualization was the most frequently coded mechanism with 42 references followed by adaptive feedback with 35 references, tiered step-by-step guidance with 31 references and solution validation with 22 references. **Table 5** summarizes the main data sources, representative evidence and analytical interpretation related to the mechanism.

Note. Coding references refer to data segments that are coded and not to independent observations. The number of references could be from different participants and interactions. The counts thus indicate the relative importance of each mechanism in the qualitative data set and are not effect sizes or causal estimates.

As shown in **Table 5**, the mechanisms were not quite related in their immediate analysis. Dynamic visualization made relationships across representations perceptually available, but adaptive feedback and tiered guidance focused students on the symbolic steps specific to each given representation. Successful solution validation allowed for comparison and evaluation. Student-05 described the motivation behind this sequence:

“This AI gives me the steps, so I can see from the graph to the formula and then to the answer. The flow is clearer.” (Student-05)

The coded evidence, therefore, described AI support as a bridge that unlocked intermediate relationships and discrepancies. It did not show that every interaction became formal. Some of the sequences ended with a slightly corrected symbolic response and stronger sequences had a student-generated explanation of why the revised representation was consistent with the graph, procedure or definition.

Adaptive Strategies and Reported Affective Responses

Four adaptive strategies appeared across the AI-supported activities: cycling among graphical, numerical, verbal, and symbolic representations; recursive consultation through progressively more specific questions; comparison with peers or alternative solutions; and metacognitive reflection on the source of an error. Recursive consultation was evident when students reformulated a question after receiving feedback rather than accepting the first response. Student-02 explained this iterative use of feedback:

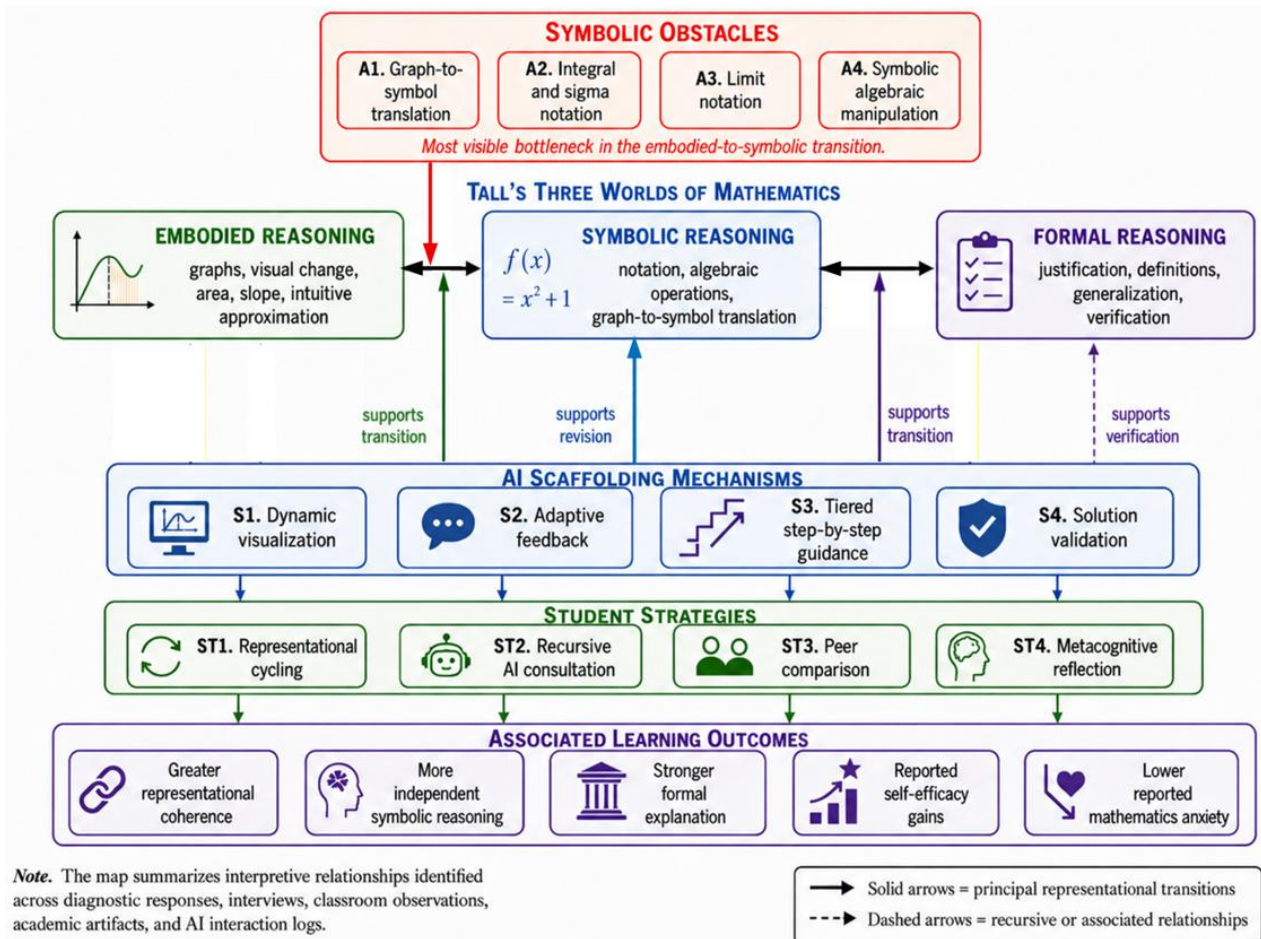


Figure 5. Thematic map of representational transitions, symbolic obstacles, and AI scaffolding pathways (Source: Authors' own elaboration)

"If I am wrong, the AI tells me where I went wrong. I try again in a different way until I get it right." (Student-02)

Metacognitive reflections and representation cycling were two strategies that were most commonly linked in participant accounts with clearer symbolic changes. Students used graphs or numerical checks to reconsider notation and explained what had changed between an initial solution and an updated solution. Peer comparison was less central to the AI logs, but was in classroom observations and interviews as a way of checking if a symbolic form was reasonable (Double et al., 2020; Yan et al., 2022). These strategic changes were accompanied by affective responses. Ten of the 12 participants reported having less mathematics anxiety during or after the supported activities and 11 reported greater confidence in attempting calculus tasks. Student-04 said, "I used to be afraid of integrals, but now I am more willing to try because there is a tool to help me." These statements were seen as self-reported case evidence, not scores from a psychological scale. No inferential correlation was calculated between strategy and affective codes.

These findings also suggest that the system is in a common pattern within every case: symbolic hurdles were the ones where intuitive or visual meaning had to be encoded and manipulated symbolically; AI support made some of these transitions more visible and revisable; and students responded through representation cycling, feedback seeking, comparison, and reflection. The degree of progress varied, particularly at the transition from symbolic process to formal justification.

Figure 5 summarizes the thematic relationships between students' representational transitions, symbolic obstacles, AI scaffolding, adaptation strategies, and learning outcomes. The map does not imply a linear development trajectory. Rather, students moved from one stage to another in embodied, symbolic, and formal reasoning, with the most visible disruption when visual or intuitive meanings must be encoded in symbolic notation.

DISCUSSION

This study examined symbolic obstacles in undergraduate calculus and the ways in which AI-supported scaffolding was associated with students' representational activity. There were four types of obstacles: graph-to-symbol translation, integral and sigma notation, limit notation, and symbolic algebraic manipulation. The biggest problem was when students tried to translate their visual or intuitive understanding into a valid symbolic expression. Across the AI-supported learning sequence, participants showed fewer coded symbolic-error occurrences in the post-activity assessment, with participant-level percentage reductions

ranging from 27.4% to 53.8% and a mean reduction of 42.3%. Error Rate (ER), defined separately as the proportion of incorrect symbolic steps among all evaluated symbolic steps, was used as a normalized descriptive indicator and should not be confused with the percentage reduction in coded error occurrences. The qualitative evidence also showed that dynamic visualization, adaptive feedback, tiered guidance, and solution validation helped students revise symbolic work, coordinate representations, and reflect on their problem-solving process. These findings should be seen as patterns in a bounded intrinsic case and not as indicators of some causal effect of AI.

Symbolic Obstacles as a Representational-Procedural Bottleneck

The high proportion of graph-to-symbol errors indicates that the problem was not simply that students could not read graphical information. Students could describe the behavior of a curve, the area, or intuitive tendency in integral notation, sigma notation, limits, or algebraically equivalent form. Symbolic obstacles were therefore the point where meaning had to be maintained while also changing the representation.

This is consistent with Duval's perspective on semiotic representation, which states that there are transformations in a representation and transformations to different representational registers (Duval, 2017). In calculus, the latter requires students to understand that a graph, a numerical pattern, and a symbolic formula can give the same mathematical result but in different representational forms. We have shown that students' difficulties arose when the correspondence was not reliable. As a result, dx , integral bounds, sigma indices, and limit markers were treated as procedural additions rather than mathematical elements.

The representational-procedural disconnection we found in this paper is consistent with Indonesian research that showed that students can perform integration without losing mathematical sense-making and that representational obstacles can involve epistemological, didactical, and language-related aspects (Agustin et al., 2024; Alifiani et al., 2023; Murniasih et al., 2020).

Our results also relate to recent studies on students' interpretations of differentials and definite integrals. Nilsen and Knutsen (2023) found that first-year engineers used graphical, derivative-related, or change-based interpretations when they explained differentials, but did not always relate those interpretations to integral notation. Such fragmentation helps us to understand why a student may know the area or change shown in a diagram while not knowing that the symbolic structure needed to express it.

The approach of the present study is to see these errors together as a representational-procedural bottleneck. The representational dimension is the continuation of meaning across graphs, verbal explanations, tables, and symbolic expressions, and the procedural dimension is the production of mathematically correct transformations after a symbolic expression has been set up. These are related but not the same: a student can apply a procedure correctly to a wrong symbolic model and understand the model or perform the symbolic transformation correctly, and the other way around. That distinction suggests that instruction should diagnose both the meaning represented and the validity of the symbolic steps rather than treating every error as a procedural weakness.

The Symbolic World as a Hinge Between Embodied and Formal Reasoning

When Tall's three worlds of mathematics are considered, the symbolic world is perceived as the bridge between embodied and formal thinking (Tall, 2008). Participants often reached a meaning of embodied meaning through curves, areas, slopes, movements or visual changes. But their ability to put formal or formal-lite explanations into those meanings was also affected by the fact that they had already encoded those meanings in a stable symbolic form.

This doesn't mean the students went from embodied to symbolic and then to formal reasoning. Rather, the AI interaction logs and diagnostic responses reflect continued movement among representations. In the end, the students returned to graphs if they found symbolic mistakes, compared visual and algebraic results and altered notation after getting feedback. The Three Worlds should thus be understood in the context of interacting modes of mathematical activity and not as a static stage of development.

In theory, these results support and refine rather than confirm the Three Worlds model in an AI-backed setting. The symbolic world was not only an intermediate level between intuition and proof. It was a kind of coordination space for students to connect visual meaning to what is familiar notation, executable procedures and formal justification. When symbolic expressions were unstable, the lack of formal reasoning inhibited students from making or communicating an argument because they did not have a mathematical language to do so. But methodological factors are also involved.

Diagnostic responses and AI interaction logs made some of the transition process visible. The analysis could identify when students needed visual support, changed a symbolic expression, compared outputs, or shifted towards explanation rather than getting a summary of representational movement from the end of the answer. So the logs were process evidence of representational activity, although they should not be considered as an indication of cognitive function.

AI Scaffolding as Support for Meaning-Making and Representational Coordination

The results indicate that the pedagogical role of AI had a lot to do with how AI was included in the learning activity. Dynamic visualization, adaptive feedback, tiered guidance, and validation supported different aspects of the representational process. Dynamic visualization made relationships between curves, numerical values, slopes, areas, and symbolic expressions more evident. Adaptive feedback focused on particular inconsistencies in symbolic expressions. Tiered guidance split representational transitions into smaller tasks, and validation required students to compare AI-supported outputs with their own reasoning.

This interpretation is in line with previous research that places external visualization as a fundamental ingredient in mathematical meaning-making, not a mere instructional device. Schoenherr and Schukajlow (2024) highlight that external visualizations are also important in mathematics learning because they are diverse and educational in nature. But they are valuable to students when we give them the tools to understand relationships and relate them to mathematical representations.

The mechanisms also fit interactive feedback models in digital mathematics. Such feedback is useful when we know where or what an error happens, and when we can try again and see if the answer is correct (Barana et al., 2021). Worked example research also shows that structured examples are useful for mathematical learning, but the timing and style of prompts might affect their use (Barbieri et al., 2023).

In the current study, AI scaffolding appeared to reorganize students' activity into the sequence of observing, predicting, encoding, checking, and explaining. This is a result of the triangulated data rather than a cognitive pathway. However, it does suggest that AI can be involved in conceptual meaning-making when prompts cause students to connect representations and justify their symbolic choices.

The change in error rates described in the text does not imply that AI was involved in this change in behavior independently. The absence of a comparison group means practice, repeated practice with the tasks, classroom instruction, peer interaction, or knowledge of the assessment may also have been involved. The clear conclusion is that post-assessment error rates were lower after the AI-based sequence and that the qualitative evidence identified plausible scaffolding processes associated with those changes.

Adaptive Strategies, Metacognitive Activity, and Self-Regulated Learning

Participants' recursive consultations, comparisons of manual and AI-driven solutions, representation cycling, and reflection on the feedback process suggest that AI interactions were sometimes part of a self-regulatory process as well. Students did not always accept the first output. Some reformulated prompts, went back to earlier steps, identified inconsistencies and considered alternatives. This is all part of monitoring, strategy adjustment, and evaluation.

Recursive checking in this study is indicative of ongoing metacognitive activity in which students are challenged to reevaluate information, understand relationships and adjust their strategies. This is consistent with Kholid et al. (2022) who showed that reflective thinking can be present throughout problem solving and Wulan et al. (2021) who demonstrated that scaffolding is most useful when students are able to identify mistakes and re-frame their reasoning. In this study, adaptive AI feedback supported this process by pointing out symbolic errors and making students feel that they should justify their revisions rather than merely accept them. The findings also relate to self-regulated learning because productive AI use required students to set goals, monitor progress, seek appropriate support, and evaluate both their own and AI-supported solutions (Chiu, 2024). Nuryanis et al. (2026) similarly identified self-regulated learning, perceived ease of use, and AI self-efficacy as strong predictors of students' perceptions of AI-based scaffolding. Although these findings do not establish causality, they highlight the importance of learner readiness, critical AI use, and instructional designs that promote active engagement rather than passive reliance.

Research on AI-supported self-regulated learning similarly suggests that AI applications may support cognitive, metacognitive, and behavioral regulation, but do not automatically regulate motivation or guarantee active learner engagement (Guan et al., 2025; Jin et al., 2023). Jin et al. (2023) found that university students perceived AI applications as potentially useful for several self-regulated learning processes while also emphasizing learner identity, activeness, and position in relation to the system. Effective AI scaffolding therefore requires students to remain decision-makers rather than passive recipients of generated responses (Darvishi et al., 2024).

In the literature on university mathematics, the same advice is held. Yoon et al. (2024) found that students' conceptualization of generative AI during mathematical proving was influenced by their definitions of proof, AI concepts, and ethical concerns. They concluded that students need guidance to engage critically with generated arguments, rather than simply be passively taking them up. Zhuang (2025) investigated two different cases in calculus which showed that students' perceptions of ChatGPT also influenced their willingness to accept mathematical errors and the persistence of using the system.

These studies reinforce the need for productive recursive consultation and not dependency. Repeatedly asking AI for support might be self-regulation when students compare, evaluate, and revise. It may also be an example of over-scaffolding if students ask more and more explicit answers but don't construct or validate their own reasoning. This is far from the difference in the number of interactions but in the epistemic work of the student.

Validation is especially important, since AI-generated mathematical answers are not always reliable. Gandolfi (2025) said there was significant variability and repeated errors in solving college calculus problems with GPT models. Thus, AI output should be considered a conjecture, feedback source or comparison object rather than a mathematical authority that is simply unquestioned.

Affective Responses and Willingness to Engage

Some people said they were more willing to tackle symbolic calculus issues that had been structured with help. These comments could suggest that immediate support and improvement were available for them and that fear of errors may have been lessened. But no standardized measure of math anxiety and self-efficacy was developed. So we should interpret the findings as participants' experiences of affective experience and not as changes to stable psychological constructs.

This cautious interpretation is consistent with the literature. Sammallahti et al. (2023) found that interventions focused on cognitive support or emotional regulation might decrease mathematics anxiety and enhance mathematics performance, but not all intervention methods and time scales were used. Zakariya (2022) also found in a systematic review that mathematics self-efficacy can be promoted with teaching methods, learning strategies, feedback, modeling and successful completion of maths tasks.

In this study, affective responses were intertwined with representational and metacognitive experiences. Students' willingness to try appeared to be more likely when they separated complex symbolic tasks into manageable steps, errors could be modified without immediate judgment, and visual or procedural support was available. The affective value of AI may be partly in

creating opportunities for attainable progress. But such support should be gradually withdrawn so that students' own mathematical competence is tied to the continued availability of the tool.

Theoretical and Pedagogical Contributions

The study makes three theoretical contributions which are interrelated. It is the first to consider symbolic obstacles as a representational-procedural bottleneck and not just an isolated notation error. Second, it extends the Three Worlds model and explains how AI-based interactions can be used to support recurrent movement between embodied, symbolic, and formal activities. Third, it demonstrates how it is possible to combine written diagnostic evidence with AI logs to analyze not just the final answers but also how students' answers are submitted, revised, checked, and explained.

The findings also show AI-based scaffolding should not be defined simply by the presence of an AI tool. Its quality of teaching depends on whether the interaction leaves student agency intact, links representations, acknowledges a specific error, and requires mathematical justification. Recent reviews of ChatGPT in math education also highlight opportunities for personalized feedback and conceptual support, as well as risks of inaccurate outputs, passive acceptance, low critical involvement, and over-reliance (Pepin et al., 2025).

The findings support the following six-stage approach for classroom use:

1. Find the location of the obstacle first before giving assistance.
2. Start with an embodied, intuitive, or visual prompt to activate meaning.
3. Require the student to create the symbolic representation
4. Give specific feedback for each step of the process, without replacing the student's solution process.
5. Give verification, comparison, and explanation of the final symbolic work.
6. Fade AI support and end with an unaided transfer task.

Dynamic visualization is important in this sequence because it allows a transformation to be assessed in multiple representations at once. Students should be required to state what remains invariant, explain how the graph constrains the symbolic expression, and create the notation themselves.

Without these requirements, visualization may be observational rather than becoming a bridge to symbolic reasoning. Assessment should also consider more than final answer quality. Symbolic-step rubrics can show where transformation fails; cross-representation items can tell if the graphical, symbolic, and verbal answers are coherent. AI interaction logs may supplement these assessments with revisions and checking behavior, but log frequency should not be interpreted as learning quality without examining the content of the interaction.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

There are several limitations to the interpretation of the findings. First, 12 students in the study were studied as part of a single university. But the value of the study lies in context-specific analysis and not statistical generalization. Therefore, replication between institutions, programs, and regions is needed to understand which patterns are stable and which depend on local curriculum, language, technology access, or teaching culture.

Second, the diagnostic tool was solely oriented towards symbolic representations and transformation. While interviews, observations, and explanation items brought together the evidence, the study did not evaluate proof competence, understanding of concepts for the whole calculus curriculum, or long-term retention. Third, the pre-post descriptive changes cannot separate AI from practice, instruction, repeated assessment, or peer participation. Future studies may also be applicable to comparison conditions, equivalent alternate-form assessments, delayed post-tests, and unaided transfer. Fourth, dependency was not investigated in depth in the study after scaffolding was removed. Further study should compare fading schedules to determine whether students can maintain representational coherence without AI assistance. Process analyses could also differentiate productive help-seeking from answer-seeking behavior.

Finally, affective responses were based on qualitative reports rather than standardized instruments. Future research should combine interview evidence with task-specific self-efficacy measures, mathematics-anxiety scales, and longitudinal interaction data (Yang et al., 2024). Such designs would allow us to study how symbolic competence, AI-supported regulation, confidence, and anxiety build up over time without reducing these complex relationships to simple frequency associations.

CONCLUSION

In this study, we identified four interrelated symbolic obstacles for undergraduate calculus students, from translation from graph to symbol, interpretation of integral and sigma notation, interpretation of limit notation and symbolic algebraic manipulation. Students struggled most to transform intuitive or visual meaning into mathematical notation; the main bottleneck was from being embodied to symbolic reasoning. In this case, artificial intelligence scaffolding (dynamic visualization, adaptive feedback, step-by-step guidance, and solution validation) helped students to compare representations, identify symbolic inconsistencies, revise solutions, and reflect on their reasoning. Participant-level comparisons revealed 27.4%–53.8% fewer coded

symbolic errors post-activity than at baseline, but these changes should be taken as descriptive within-case evidence and not as a result of artificial intelligence.

The study provides a coherent approach to calculus education to integrate Tall's Three Worlds of Mathematics, representational fluency, embodied thinking and artificial intelligence interaction logs. The findings indicate that symbolic obstacles are not simply procedural errors, but representational errors in the way in which mathematics is understood and artificial intelligence is most useful as contingent scaffolding rather than as an answer-generating tool. Instructors should combine visual exploration, student-generated symbolic representations, targeted feedback, conceptual justification, and progressively faded assistance. Because this study is based on only a small sample from one institution and only in symbolic performance, future research should extend this framework to larger schools and use longitudinal or comparative designs and look at how independent representational and formal reasoning is maintained even when there is no artificial intelligence support.

Author contributions: **IN:** conceptualization, study design, data collection, writing – original draft; **TN:** supervision, methodology, writing – review & editing, formal analysis, interpretation. **EH:** supervision, validation, resources, writing – review & editing. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding: No funding source is reported for this study.

Acknowledgements: The authors thank the participating students and colleagues in the Department of Mathematics Education for their cooperation during data collection and discussion sessions. The authors also thank the research assistants who supported classroom observations and transcription, and the anonymous colleagues who provided valuable feedback on earlier drafts.

Ethical statement: The authors stated that formal ethics committee approval was not required for this study under the applicable institutional procedures. Participation was voluntary, informed consent was obtained from all participants, and participant identities were protected through pseudonymization. The authors further stated that the study involved educational activities and qualitative data collection within a university calculus-learning context and did not involve clinical intervention or the collection of sensitive biomedical data.

AI statement: The authors stated that during the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-5.6) to assist with language editing, paraphrasing, improving clarity, and generating ideas for organizing the manuscript. Napkin AI was used to assist in developing ideas and visual concepts for the figures. All AI-assisted outputs were critically reviewed, verified, and revised by the authors. AI tools were not used to generate, fabricate, or manipulate research data, analyses, results, or conclusions. The authors retain full responsibility for the accuracy, originality, integrity, and final content of the manuscript and its figures.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

- Afriyani, D., Sa'dijah, C., Subanji, & Muksar, M. (2018). Characteristics of students' mathematical understanding in solving multiple representation task based on SOLO taxonomy. *International Electronic Journal of Mathematics Education*, 13(3), 281-287. <https://doi.org/10.12973/iejme/3920>
- Agustin, R. D., Sa'dijah, C., Susiswo, & Sukoriyanto. (2024). Mathematical semantics representation obstacles of preservice mathematics teachers to solve geometry problems. *Pegem Journal of Education and Instruction*, 14(3), 376-392. <https://doi.org/10.47750/pegegog.14.03.35>
- Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health*, 2, Article 100051. <https://doi.org/10.1016/j.glmedi.2024.100051>
- Alifiani, Subanji, Permadi, H., & Irawati, S. (2023). Metacognitive intervention: Can it solve suspension of sense-making in integration problem-solving? *Mathematics Teaching Research Journal*, 15(1), 44-66. <https://files.eric.ed.gov/fulltext/EJ1391470.pdf>
- Aniswita, A., Sepriyanti, N., & Medika, G. H. (2024). Unravelling undergraduate mathematics students' understanding of derivatives. *JRAMathEdu: Journal of Research and Advances in Mathematics Education*, 9(4), 262-272. <https://doi.org/10.23917/jramathedu.v9i4.11267>
- Bajracharya, R. R., Sealey, V. L., & Thompson, J. R. (2023). Student understanding of the sign of negative definite integrals in mathematics and physics. *International Journal of Research in Undergraduate Mathematics Education*, 9, 62-91. <https://doi.org/10.1007/s40753-022-00202-y>
- Barana, A., Marchisio, M., & Sacchet, M. (2021). Interactive feedback for learning mathematics in a digital learning environment. *Education Sciences*, 11(6), Article 279. <https://doi.org/10.3390/educsci11060279>
- Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K. (2023). A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review*, 35, Article 11. <https://doi.org/10.1007/s10648-023-09745-1>
- Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member checking: A tool to enhance trustworthiness or merely a nod to validation? *Qualitative Health Research*, 26(13), 1802-1811. <https://doi.org/10.1177/1049732316654870>
- Biza, I., González-Martín, A. S., & Pinto, A. (2022). 'Scaffolding' or 'filtering': A review of studies on the diverse roles of calculus courses for students, professionals and teachers. *International Journal of Research in Undergraduate Mathematics Education*, 8(2), 389-418. <https://doi.org/10.1007/s40753-022-00180-1>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328-352. <https://doi.org/10.1080/14780887.2020.1769238>

- Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, 56(3), 1391-1412. <https://doi.org/10.1007/s11135-021-01182-y>
- Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development*, 72, 2401-2416. <https://doi.org/10.1007/s11423-024-10366-w>
- Chytas, C., van Borkulo, S. P., Drijvers, P., Barendsen, E., & Tolboom, J. L. J. (2024). Computational thinking in secondary mathematics education with GeoGebra: Insights from an intervention in calculus lessons. *Digital Experiences in Mathematics Education*, 10(2), 228-259. <https://doi.org/10.1007/s40751-024-00141-0>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, 210, Article 104967. <https://doi.org/10.1016/j.compedu.2023.104967>
- Di Martino, P., Gregorio, F., & Iannone, P. (2023). The transition from school to university in mathematics education research: New trends and ideas from a systematic literature review. *Educational Studies in Mathematics*, 113(1), 7-34. <https://doi.org/10.1007/s10649-022-10194-w>
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis. *Educational Psychology Review*, 32, 481-509. <https://doi.org/10.1007/s10648-019-09510-3>
- Duval, R. (2017). *Understanding the mathematical way of thinking: The registers of semiotic representations*. Springer. <https://doi.org/10.1007/978-3-319-56910-9>
- Fonger, N. L. (2019). Meaningfulness in representational fluency: An analytic lens for students' creations, interpretations, and connections. *The Journal of Mathematical Behavior*, 54, Article 100678. <https://doi.org/10.1016/j.jmathb.2018.10.003>
- Gandolfi, A. (2025). GPT-4 in education: Evaluating aptness, reliability, and loss of coherence in solving calculus problems and grading submissions. *International Journal of Artificial Intelligence in Education*, 35(1), 367-397. <https://doi.org/10.1007/s40593-024-00403-3>
- Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies*, 30, 4493-4518. <https://doi.org/10.1007/s10639-024-12881-y>
- Halák, J. (2022). Mathematics embodied: Merleau-Ponty on geometry and algebra as fields of motor enaction. *Synthese*, 200, Article 34. <https://doi.org/10.1007/s11229-022-03526-z>
- Hwang, G.-J., & Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, 9(6), Article 584. <https://doi.org/10.3390/math9060584>
- Jin, S.-H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, 20, Article 37. <https://doi.org/10.1186/s41239-023-00406-5>
- Kholid, M. N., Sa'dijah, C., Hidayanto, E., & Permadi, H. (2022). Students' reflective thinking pattern changes and characteristics of problem solving. *Reflective Practice*, 23(3), 319-341. <https://doi.org/10.1080/14623943.2021.2025353>
- Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher*, 42(8), 846-854. <https://doi.org/10.1080/0142159X.2020.1755030>
- Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, 24(1), 120-124. <https://doi.org/10.1080/13814788.2017.1375092>
- Li, S., Shen, Y., Jiao, X., & Cai, S. (2022). Using augmented reality to enhance students' representational fluency of linear functions in middle schools. *Mathematics*, 10(10), Article 1718. <https://doi.org/10.3390/math10101718>
- Miller, E. M., Porter, J. E., & Barbagallo, M. S. (2023). Simplifying qualitative case study research methodology: A step-by-step guide using a palliative care example. *The Qualitative Report*, 28(8), 2363-2379. <https://doi.org/10.46743/2160-3715/2023.6478>
- Mtisi, S. (2022). The qualitative case study research strategy as applied on a rural enterprise development doctoral research project. *International Journal of Qualitative Methods*, 21, Article 16094069221145849. <https://doi.org/10.1177/16094069221145849>
- Murniasih, T. R., Sa'dijah, C., Muksar, M., & Susiswo. (2020). Fraction sense: An analysis of preservice mathematics teachers' cognitive obstacles. *Center for Educational Policy Studies Journal*, 10(2), 27-47. <https://doi.org/10.26529/cepsj.742>
- Nilsen, H. K., & Knutsen, K. H. (2023). First-year engineering students' interpretations of differentials and definite integrals. *International Journal of Research in Undergraduate Mathematics Education*, 9(1), 173-200. <https://doi.org/10.1007/s40753-022-00208-6>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16, Article 1609406917733847. <https://doi.org/10.1177/1609406917733847>
- Nurrahmawati, Sa'dijah, C., Sudirman, & Muksar, M. (2021). Assessing students' errors in mathematical translation: From symbolic to verbal and graphic representations. *International Journal of Evaluation and Research in Education*, 10(1), 115-125. <https://doi.org/10.11591/ijere.v10i1.20819>

- Nuryanis, Nusantara, T., Murtadho, N., & Faizah, S. (2026). AI-based scaffolding and conceptual understanding: Evidence from Indonesian students using PLS-SEM. *International Journal of Evaluation and Research in Education*, 15(2), 1060-1078. <https://doi.org/10.11591/ijere.v15i2.38450>
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, Article 1609406919899220. <https://doi.org/10.1177/1609406919899220>
- Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 11(1), 9-41. <https://doi.org/10.1007/s40751-025-00172-1>
- Sammallahti, E., Finell, J., Jonsson, B., & Korhonen, J. (2023). A meta-analysis of math anxiety interventions. *Journal of Numerical Cognition*, 9(2), 346-362. <https://doi.org/10.5964/jnc.8401>
- Schoenherr, J., & Schukajlow, S. (2024). Characterizing external visualization in mathematics education research: A scoping review. *ZDM – Mathematics Education*, 56, 73-85. <https://doi.org/10.1007/s11858-023-01494-3>
- Serhan, D., & Welcome, N. (2024). Integrating ChatGPT in the calculus classroom: Student perceptions. *International Journal of Technology in Education and Science*, 8(2), 325-335. <https://doi.org/10.46328/ijtes.559>
- Shvarts, A., Alberto, R., Bakker, A., Doorman, M., & Drijvers, P. (2021). Embodied instrumentation in learning mathematics as the genesis of a body-artifact functional system. *Educational Studies in Mathematics*, 107(3), 447-469. <https://doi.org/10.1007/s10649-021-10053-0>
- Sneed, S., Nguyen, C. H. P., & Eubank, C. L. (2020). An introduction to case study methodology: Single case and multiple case approaches. *International Journal of Adult Education and Technology*, 11(4), 1-11. <https://doi.org/10.4018/IJAET.2020100101>
- Tall, D. O. (2008). The transition to formal thinking in mathematics. *Mathematics Education Research Journal*, 20(2), 5-24. <https://doi.org/10.1007/BF03217474>
- Thompson, P. W., & Harel, G. (2021). Ideas foundational to calculus learning and their links to students' difficulties. *ZDM – Mathematics Education*, 53(3), 507-519. <https://doi.org/10.1007/s11858-021-01270-1>
- Torres-Peña, R. C., Peña-González, D., Chacuto-López, E., Ariza, E. A., & Vergara, D. (2024). Updating calculus teaching with AI: A classroom experience. *Education Sciences*, 14(9), Article 1019. <https://doi.org/10.3390/educsci14091019>
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271-296. <https://doi.org/10.1007/s10648-010-9127-6>
- Way, J., & Ginns, P. (2024). Embodied learning in early mathematics education: Translating research into principles to inform teaching. *Education Sciences*, 14(7), Article 696. <https://doi.org/10.3390/educsci14070696>
- Wulan, E. R., Subanji, & Muksar, M. (2021). Metacognitive failure in constructing proof and how to scaffold it. *Al-Jabar: Jurnal Pendidikan Matematika*, 12(2), 295-314. <https://doi.org/10.24042/ajpm.v12i2.9590>
- Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, 37, Article 100484. <https://doi.org/10.1016/j.edurev.2022.100484>
- Yang, Y., Maeda, Y., & Gentry, M. (2024). The relationship between mathematics self-efficacy and mathematics achievement: Multilevel analysis with NAEP 2019. *Large-scale Assessments in Education*, 12, Article 16. <https://doi.org/10.1186/s40536-024-00204-z>
- Yoon, H., Hwang, J., Lee, K., Roh, K. H., & Kwon, O. N. (2024). Students' use of generative artificial intelligence for proving mathematical statements. *ZDM – Mathematics Education*, 56(7), 1531-1551. <https://doi.org/10.1007/s11858-024-01629-0>
- Zakariya, Y. F. (2022). Improving students' mathematics self-efficacy: A systematic review of intervention studies. *Frontiers in Psychology*, 13, Article 986622. <https://doi.org/10.3389/fpsyg.2022.986622>
- Zhuang, Y. (2025). Lessons from using ChatGPT in calculus: Insights from two contrasting cases. *Journal of Formative Design in Learning*, 9(1), 25-35. <https://doi.org/10.1007/s41686-025-00098-2>