

Critical Thinking Activation During Limits Problem Solving Mediated by Symbolab: An Instrumental Genesis Perspective

Enrique Ceballos-Alexis

eceballos1@ucol.mx

Instituto Politécnico Nacional

Research Article

Keywords: critical thinking, instrumental genesis, Symbolab, calculus education, computer algebra systems, Facione, CT-neutral zone

Posted Date: June 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9988433/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This study describes how Facione's (1990) six cognitive critical thinking (CT) skills (interpretation, analysis, inference, evaluation, explanation, and self-regulation) are activated during limits problem solving mediated by Symbolab, a widely used computer algebra system (CAS), from the perspective of instrumental genesis theory (Rabardel, 1995; Trouche, 2004). Using a qualitative, descriptive-interpretive design with reflexive thematic analysis (Braun & Clarke, 2006), written productions from 29 first-year students majoring in International Business at the Universidad de Colima (Mexico) were analyzed using a theoretically grounded nested instrument designed to make CT activation observable across three stages of instrumental genesis. A central construct is the CT-neutral zone: a structurally determined interval during which CT skills are transitionally suspended while students mechanically enter mathematical expressions into Symbolab's interface. This zone is interpreted not as a cognitive deficiency but as a predictable consequence of early-phase instrumental genesis, consistent with Rabardel's (1995) framework. Additional patterns include three qualitatively distinct evaluation subtypes (procedural, conceptual, and contextual) that emerged inductively, as well as two cross-stage inductive patterns that describe how gaps in Stage 1 analysis can propagate into uncritical Stage 2 tool acceptance. The study contributes an empirical-descriptive framework that articulates Facione's CT components with instrumental genesis processes in undergraduate calculus, an intersection that, to our knowledge, prior literature has not addressed systematically.

1. Introduction

Calculus remains among the courses with the highest failure rates in higher education worldwide. Research consistently documents that failure in calculus is associated not only with procedural gaps but with shallow understanding of mathematical objects and with problem-solving strategies that prioritize mechanical algorithm application over reflective reasoning (Schoenfeld, 1985; Tall, 2013). In this context, step-by-step solvers such as Symbolab, Wolfram Alpha, and Photomath, which provide fully worked solutions for limits, derivatives, and integrals, have become part of the everyday problem-solving environment of university students (Aguilar & Puga, 2020; Webel & Otten, 2015). The pedagogically relevant question is no longer whether students use these tools, but what happens cognitively when they do.

Despite a growing body of research on critical thinking in mathematical problem solving (Mastuti et al., 2022; Permadi & Zakiah, 2021; Wang & Abdullah, 2024) and a consolidated tradition of research on instrumental genesis with computer algebra systems (Artigue, 2002; Drijvers, 2002, 2019; Trouche, 2004), these two traditions have operated largely independently. We located no study that systematically describes how Facione's (1990) six cognitive CT components activate and transform during the instrumental genesis of a CAS tool in undergraduate calculus. This gap is the specific problem the present study addresses.

The study adopts a descriptive-interpretive approach. Its purpose is not to evaluate whether Symbolab improves or damages CT, but to describe the activation patterns of CT skills within the mediated problem-solving process: what activates them, what observable evidence exists of their presence, and whether they transform over the course of a single instrument application. The central research question is: How do Facione's cognitive critical thinking skills activate during limits problem solving mediated by Symbolab, from the perspective of instrumental genesis?

2. Theoretical Framework

2.1 Facione's Critical Thinking Framework

Critical thinking has been conceptualized in several traditions, from the informal-logic emphasis of Ennis (1962) and the discipline-embedded view of McPeck (1981) to the dispositions-and-skills synthesis of Paul (1990). The conceptual foundation of this study is Facione's (1990) Delphi Report, which represents the consensus of 46 experts from diverse disciplines on the nature of critical thinking and its component skills. The report characterizes critical thinking as purposeful, self-regulatory judgment that gives rise to interpretation, analysis, evaluation, inference, explanation, and self-regulation. This definition is adopted here for three reasons: its multi-disciplinary expert-consensus origin, its operationalization into six identifiable cognitive components, and its disciplinary independence, which permits systematic mathematization for the calculus domain.

A key distinction introduced in this study is the difference between CT-complete and CT-truncated tasks. A pre-formulated equation is a CT-truncated task: the work of interpretation, analysis, and inference has already been performed externally. A contextualized word problem is a CT-complete task: it requires the student to traverse the full process from natural-language comprehension to contextual result verification, activating all six CT components at identifiable points (cf. Polya, 1945; Verschaffel et al., 2020). This distinction justifies the exclusive use of word problems in the study's instrument design.

2.2 Instrumental Genesis Theory

Instrumental genesis theory was formulated by Rabardel (1995) in the context of ergonomic psychology and operationalized for mathematics education by Trouche (2004). Its central distinction is between artifact and instrument. An artifact is an external object designed to mediate an activity; an instrument is a mixed entity constructed by the subject through interaction with the artifact. The instrument does not pre-exist: it emerges from the subject's activity as the subject develops schemes of use through appropriation of the artifact.

Instrumental genesis involves two interdependent sub-processes. Instrumentalization is directed from the subject toward the artifact: the subject discovers, personalizes, and adapts it to their purposes. Instrumentation is directed from the artifact toward the subject: the artifact's possibilities and constraints condition and modify the subject's action schemes, reorganizing their reasoning strategies. Both processes coexist and feed each other throughout genesis.

2.3 Integrated Framework: CT Activation Across Three Stages

The integration of Facione's and Rabardel's frameworks generates a descriptive model that organizes CT activation across three stages corresponding to the moments of instrumental genesis. In Stage 1 (pre-Symbolab), the student works on the word problem without technological mediation. The predominant CT components are interpretation, analysis, and inference. The quality of inference is determinant: the expression formulated at this stage is what the student inputs into Symbolab, making any inferential error a source of downstream disruption.

In Stage 2 (during Symbolab), the interaction with the tool introduces a dual-level dynamic. At the external level, the mechanical entry of the mathematical expression into Symbolab's interface defines an interval in which CT skills tend to be transitionally suspended: the CT-neutral zone. At the internal level, once Symbolab generates the step-by-step solution, all six CT components can be reactivated within the evaluation of that solution, now directed specifically at the tool-generated output.

In Stage 3 (post-Symbolab), the student returns to the original problem with the result in hand. The predominant CT components are explanation and self-regulation. This stage is where the effects of instrumentation (understood as the modification of the subject by the artifact) become most visible.

2.4 The CT-Neutral Zone: Theoretical Characterization

The CT-neutral zone is defined in this study as the temporal interval during Symbolab-mediated problem solving in which Facione's cognitive CT skills are transitionally suspended. This interval coincides, in instrumental genesis terms, with the mechanical entry phase. It is essential to emphasize that the CT-neutral zone is not a failure of the student's critical thinking, nor an indicator of incompetence or cognitive disengagement. It is a structural characteristic of instrumentalization: when the subject is focused on the technical translation of their inference to the artifact's interface, cognitive resources orient toward the immediate technical task. The construct has independent theoretical support in Rabardel's (1995) account, which predicts that during the artifact-entry phase the subject's cognitive resources orient toward immediate technical coordination, a demand that by definition competes with the higher-order processes characteristic of critical thinking (see also Barzel et al., 2019, on the tension between instrumental automation and metacognitive reflection).

3. Literature Review and Research Gap

3.1 Critical Thinking in Mathematics Education

Research on critical thinking in mathematical problem solving has grown considerably. A recent systematic review by Wang and Abdullah (2024) maps the field in higher education and documents both its rapid expansion and its methodological heterogeneity. Within this body of work, several studies have applied Facione's framework directly to student productions in mathematical contexts: Mastuti et al.

(2022) operationalized Facione's components to characterize undergraduate CT, and Nur Juhainah Ulfa and Permatasari (2023) examined CT activation during word-problem solving across learning styles. In the calculus domain specifically, Permadi and Zakiyah (2021) reported that open-ended, peer-assessed task designs strengthened CT skills, and Evendi et al. (2022) and Orhan (2022) linked CT activation to problem-based designs and to problem-solving ability more broadly. These works consistently document substantial variation in component activation across students and tasks, and they repeatedly identify inference and self-regulation as the components least frequently activated in their full dimension under conventional assessment conditions, echoing longstanding evidence that metacognitive self-regulation is both central to and underdeveloped in mathematical problem solving (Schoenfeld, 1992). They establish that Facione's components are operationalizable through textual evidence, a methodological precedent the present study builds on, but they do not examine activation under tool mediation.

3.2 Technology and Computer Algebra Systems in Calculus

The integration of digital tools into mathematics learning has a long trajectory, from graphing calculators (Drijvers & Weigand, 2010) to desktop CAS such as Maple, Mathematica, and Derive, whose automation of symbolic calculation raised early questions about the role of by-hand work in learning (Buchberger, 1990, 2002; Heid, 1988). Heid (1988) showed that students who delegated procedural calculation to a CAS could achieve superior conceptual understanding when course time was redirected toward conceptual exploration. Artigue (2002) demonstrated that instrumental genesis with CAS is a markedly slower and more complex process than software designers tend to anticipate, requiring deliberate pedagogical orientation, and Drijvers (2002, 2015, 2019) documented both the possibilities and the obstacles of CAS use, consistently showing that cognitive benefits depend on task design and teacher mediation rather than tool presence alone (see also Geiger et al., 2015; Hoyles et al., 2004).

A more recent strand examines the freely available step-by-step solvers that now mediate much of students' independent work. Aguilar and Puga (2020) observed how undergraduates recruit internet tools to cope with a calculus task, and Webel and Otten (2015) and Zain et al. (2023) discuss how Photomath-type applications reshape what counts as doing a problem. On cognitive effects, Thomas and Hong (2004) described a dual thinking pattern that alternates between reasoning with the tool and reasoning about the problem; Jankvist and Misfeldt (2015) identified difficulties that emerge when CAS procedures are objectified without conceptual understanding; and Barzel et al. (2019) documented students' difficulty in articulating their own thinking while using such tools. Symbolab belongs squarely to this category of step-by-step solver, yet it has received little dedicated analytic attention, and none that tracks the full arc of critical reasoning around a single tool use.

3.3 Instrumental Genesis as an Analytic Lens

Trouche (2004) operationalized Rabardel's theory for mathematics education, introducing the notion of instrumental orchestration to describe the deliberate pedagogical management of the artifacts available in a learning environment. Subsequent work applied the instrumental approach to CAS and handheld technologies, documenting the coexistence of instrumentalization and instrumentation in everyday

student practice (Aldon, 2010; Artigue, 2002; Drijvers, 2002). This tradition supplies a vocabulary for describing how a tool becomes an instrument, but it has not, to date, been articulated with a fine-grained model of critical thinking components.

3.4 Synthesis and Research Gap

The specific gap this study addresses lies at the intersection of these three traditions. Database searches in ERIC, Scopus, and Google Scholar combining the terms “critical thinking,” “Facione,” “calculus,” “CAS,” “instrumental genesis,” and “Symbolab” returned studies on CT in mathematics, on CT with technology, and on CAS in calculus, but none that systematically articulates how all six of Facione's cognitive CT components activate and transform during the instrumental genesis of a CAS tool in undergraduate calculus. The present study constitutes a first attempt to fill that gap.

4. Methodology

4.1 Research Design

The study adopts a qualitative, descriptive-interpretive approach. This choice responds to the nature of the research question: describing how CT skills activate during a problem-solving process cannot be addressed through variable quantification, but through in-depth analysis of written productions that account for that process (Braun & Clarke, 2006). The design is observational and non-experimental: no variables are manipulated, no comparison groups are formed, and no pedagogical interventions are implemented. The study observes the CT-activation phenomenon under the ordinary teaching conditions of the investigated context.

The researcher served simultaneously as course instructor and as researcher. This dual role introduces reflexivity considerations that are addressed explicitly: coding criteria were derived from the theoretical framework before data analysis; interpretation consistently aimed at describing the CT-activation phenomenon rather than evaluating individual student performance; and the codebook provides transparent, auditable decision documentation available for external review.

4.2 Participants and Context

Participants were 29 first-year students enrolled in the Differential and Integral Calculus course at the Facultad de Economía, Universidad de Colima, Mexico, during one regular semester. The instrument analyzed here corresponds to the limits topic and was completed by the full group, constituting a census within that application; it is one of three instruments administered across the semester within a larger study (two on limits, one on derivatives), the remaining analyses of which are ongoing. The instrument was distributed through Google Classroom and completed individually within the designated activity time.

4.3 Instrument Design: The Nested Observation Structure

The data-collection instrument is a structured work guide organized around a theoretically grounded dual-level nesting design. At the external level, the instrument contains six sections corresponding to Facione's six CT components. The first three (Interpretation, Analysis, Inference) are completed without Symbolab, corresponding to Stage 1. The final two external sections (Explanation and Self-Regulation) are completed after the Symbolab interaction, corresponding to Stage 3.

At the internal level, Section 4 (Evaluation) constitutes the instrument's methodological core and corresponds to Stage 2. Within this section, the instrument embeds all six Facione CT components again, now applied specifically to Symbolab's generated solution. The student is prompted to interpret what Symbolab is doing at each step, analyze the relationships among those steps, infer whether the process leads to the correct solution, evaluate the quality of the mathematical argument the tool produces, explain each generated step in their own words, and self-regulate their understanding during the interaction. This nested design is termed a transitional observation structure: it makes observable the exact moment the student transitions from autonomous pre-tool reasoning (Stage 1) to tool-mediated reasoning (Stage 2) and back to autonomous post-tool synthesis (Stage 3).

The problem used in the instrument is a contextualized word problem involving the function $C(d) = (d^2 - 6d + 8) / (d - 2)$, which models the average cost per patient as a function of distance to an urban center. Students are asked to evaluate the limit as d approaches 2 and as d approaches 6 km, contextualized within a clinical investment decision. The case $d = 2$ yields a 0/0 indeterminate form requiring factoring (the limit equals -2); the case $d = 6$ admits direct substitution (the limit equals 2). This double-case structure (one requiring a special technique and one admitting direct substitution) is the primary source of analytical variability documented in the corpus.

The instrument was designed by the researcher from the study's theoretical framework, operationalizing each CT sub-skill into prompts accessible to first-year students without prior formal training in critical thinking as a discipline. It was reviewed by academics with experience in mathematics education and refined following a pilot application, with adjustments to instruction wording, question order, and prompt clarity.

4.3.1 The Instrument as an Observation Structure, Not a Thinking Inducer

A methodological objection that must be addressed explicitly is whether the instrument generates the thinking it purports to observe (the demand-characteristics problem). Labeling sections with the names of CT components could, in principle, prompt students to produce responses that resemble critical thinking without it genuinely occurring. The instrument's internal evidence answers this objection in three complementary ways. First, the coding system distinguishes between mentioning a component's vocabulary and operating its sub-skills: a student who writes "now I will evaluate" and then mechanically transcribes Symbolab's steps without any appraisal receives the CT-neutral code, not an activation code, because the criterion requires textual evidence of the constitutive sub-skills, not section completion.

Second, if the instrument merely induced critical-looking responses, one would expect high uniformity across sections and students; the data show the opposite, with substantial variability including blank sections, mechanical transcriptions, and analytically rich responses. The CT-neutral zone itself is the clearest evidence that the instrument is sensitive to the absence of CT, not only to its presence. Third, in ordinary classroom use the entire Symbolab process is opaque: the student inputs the problem, obtains the answer, and moves on. The instrument does not create the underlying cognitive processes; it creates a structured space in which they are externalized in writing, and the wide variation in those externalizations demonstrates that it captures genuine differences in cognitive engagement rather than producing artificial uniformity.

4.4 Data Analysis

Data was analyzed through reflexive thematic analysis following Braun and Clarke's (2006) approach, using a hybrid deductive-inductive coding system. Deductive codes were derived from the theoretical framework: the six Facione CT components with their sub-skills, organized across the three instrumental genesis stages, including the internal level of CT reactivation within the Evaluation section. Inductive codes emerged from the data, capturing patterns not anticipated by the framework. Three activation levels were used as qualitative descriptors: Not Evidenced (NE), no observable evidence of the component's sub-skills; Partially Evidenced (PE), at least one sub-skill active but incomplete or superficial; and Evidenced (E), the component's nuclear sub-skills clearly active in the written production.

A complementary qualitative rubric, grounded in Facione's (1990) sub-skill definitions, was used to systematize activation patterns across the 29 productions. Its epistemological status must be made explicit: the rubric is not a result of the study, nor a psychometric measurement instrument. Its three levels are not numerical and do not constitute an ordinal continuum with metric properties; they are qualitative descriptors that orient coding and comparison. The rubric is an observation device that complements, without replacing, the in-depth description produced by the thematic analysis.

Trustworthiness was ensured using Lincoln and Guba's (1985) criteria. Credibility was supported through triangulation between the thematic analysis and the rubric descriptors and through systematic grounding of interpretations in direct student-production quotes. Dependability was supported through explicit documentation of the coding system in a full codebook with operational definitions, textual indicators, and examples drawn directly from the data. Confirmability was supported through consistent linkage of interpretations to the integrated theoretical framework, and an audit trail documenting coding decisions, inductive-code emergence, and sampling rationale was maintained throughout.

Consistent with reflexive thematic analysis, the study did not employ inter-rater reliability. Braun and Clarke (2006) hold that inter-coder reliability is a positivist criterion inappropriate to reflexive analysis, whose logic is interpretive and does not presuppose a single correct coding that independent researchers should replicate identically. The dependability mechanism instead resides in the transparency of the codebook: any researcher applying the same codebook to the same corpus can follow the analytic reasoning and assess its internal consistency and plausibility. For the same reason,

the counts reported below (for example, 8 of 29) describe the analyzed corpus and are offered as evidence of pattern recurrence; they are not presented as prevalence estimates for a wider population, in line with the study's analytic unit: the CT-activation phenomenon rather than the individual student.

4.5 Ethical Considerations

The instrument header included an informed-consent notice explaining the study's purpose, guaranteeing the anonymity of student productions, and clarifying that participation was voluntary and would not affect course grades. Data were anonymized before analysis, with participants identified solely through alphanumeric codes (e.g., L1-001). The study adhered to the institutional norms for research with human participants of the Universidad de Colima.

5. Findings

5.1 Stage 1: CT Activation Prior to Symbolab Use

5.1.1 Interpretation: From Natural Language to Mathematical Representation

Analysis of the Interpretation section reveals a consistent pattern: most students successfully identified the main magnitudes and paraphrased the problem statement (situational interpretation) but showed greater difficulty completing the transition to mathematical interpretation, specifically assigning variables, selecting appropriate limit notation, and distinguishing the two cases ($d = 2$ and $d = 6$) as mathematically distinct.

Representative of complete situational interpretation with incomplete mathematical interpretation is the following production: "The problem talks about how much it costs per patient depending on the distance. We need to find what happens when d is 2 and 6 km" (L1-008). The student identifies the key magnitudes but does not formulate the limit notation nor specify the indeterminate condition at $d = 2$. In contrast, students with fuller mathematical interpretation produced responses such as: "We are looking for $\lim C(d)$ as d approaches 2 and $\lim C(d)$ as d approaches 6. At $d = 2$ there may be an indetermination, so we need to check" (L1-022). This evidences the simultaneous activation of situational and mathematical interpretation, including the categorization sub-skill that flags the potential special condition.

5.1.2 Analysis: Identification of Relationships and Conditions

The Analysis component showed the most variable activation in the corpus; nine of 29 participants received NE in this section. A significant inductive finding was the pattern coded ANL-OMITIDO: a subset of students passed directly from Interpretation to Inference without articulating the relationships between the function's structure, the point of evaluation, and the implied mathematical condition. This

analytical skip co-occurred with a higher frequency of errors in subsequent inference, particularly the failure to distinguish that $d = 2$ requires factoring while $d = 6$ admits direct substitution. This association was not anticipated in the framework and has implications for understanding the role of Analysis in the CT inferential chain.

Students with evidenced Analysis articulated structural conditions explicitly: “The function has a numerator that factors as $(d - 2)(d - 4)$ and a denominator $(d - 2)$. At $d = 2$ both become zero, so there is a $0/0$ indetermination. At $d = 6$ the denominator is not zero so I can substitute directly” (L1-017). This articulation made the downstream inference more precise and reduced the probability of entering an incorrect expression into Symbolab.

5.1.3 Inference: Constructing the Mathematical Model

Inference showed a bimodal distribution: one group formulated the complete mathematical expression correctly and with justification before using Symbolab; another introduced incorrect or partially formulated expressions. From an instrumental genesis perspective, the quality of inference is determinant, since the expression formulated here is precisely what the student inputs into Symbolab. Any inferential error generates the pattern coded INF-ERROR-NO-DETECTADO: an incorrect expression entered into Symbolab without the student questioning the pertinence of the generated result. This cross-stage pattern (a Stage 1 error compounded by Stage 2 uncritical acceptance) is one of the study’s key inductive findings and illustrates how the CT-neutral zone can amplify, rather than merely suspend, the consequences of weak reasoning.

The most extreme case in the corpus is L1-025, who entered an altogether different expression into Symbolab, obtained an infinite result, and accepted it without interrogation, including without questioning whether an infinite cost rate is coherent within a clinical investment decision.

5.2 Stage 2: CT Activation During Symbolab Use

5.2.1 The CT-Neutral Zone: Empirical Evidence

The CT-neutral zone’s theoretical prediction found consistent empirical support. Eight of 29 participants received the NE-PCN code in the Evaluation section, producing responses that consisted of literal transcription of Symbolab’s generated steps without evidence of critical processing: reproduction of the tool’s algebraic notation without paraphrase, commentary, or appraisal of each step’s pertinence.

A particularly striking sub-pattern was the appearance of identical verbal formulations across distinct participants. Three students (L1-002, L1-007, L1-016) produced the following response verbatim in the Evaluation section: “Yes, because Symbolab gives us the correct results and simplifies them.” An identical formulation appearing in three independent participants is compelling evidence, consistent with the construct, that the CT-neutral zone describes a structural phenomenon rather than an individual characteristic.

Analytically, the most compelling evidence for the CT-neutral zone as a structural cognitive phenomenon, distinct from simple disengagement, is its coexistence with high mathematical competence in the same participant. L1-007 produces the most detailed factoring description in the entire corpus in the Inference section, correctly identifying the relevant numbers, describing the cancellation, obtaining $(d - 4)$, and substituting. Immediately afterward, in the Evaluation section, the same participant writes the verbatim CT-neutral response cited above. A student who did not understand the factoring could not describe it with that precision; a student who does not evaluate, here, does so not from lack of competence but from a transitory cognitive state linked to the instrumentation process, as Rabardel's (1995) framework predicts.

5.2.2 Evaluation: Three Qualitatively Distinct Subtypes

Active evaluation of Symbolab's generated steps constitutes the opposite pole of the CT-neutral zone. Thematic analysis identified three qualitatively distinct evaluation patterns that emerged inductively from the data.

The first, coded EVL-PROC (procedural evaluation), involves verifying that steps are mathematically correct without connecting them to conceptual understanding: "Symbolab applied the factoring correctly and then canceled the $(d - 2)$ term. The result matches what I had done before" (L1-004). This evaluation is active in that it implies verification, but it remains procedural.

The second, coded EVL-CONC (conceptual evaluation), connects steps to underlying mathematical meaning: "The result $- 2$ is consistent with the algebra and with numerical checking. d approaches 2 from either side" (L1-013). This participant verifies against two independent checks (their own prior algebra and the bilateral numerical behavior of the function at $d = 2$), evidencing the sub-skill of questioning one's own evidence. This pattern was documented in three participants.

The third, coded EVL-CONT (contextual evaluation), connects the result to the problem context, evaluating its reasonableness. It was documented in a single participant (L1-029), who evaluated the negative limit value in terms of its economic meaning, reasoning that a negative cost rate signals a non-profitability indicator for the investment decision. It should be noted that this economic interpretation is imprecise: a negative limit value in this model would imply the clinic receives payment per patient, rendering the function economically incoherent at that point rather than merely unprofitable. Nonetheless, this appraisal of contextual reasonableness was not anticipated in the framework and represents the most sophisticated form of Evaluation activation observed.

That this contextual appraisal occurred within Section 4.4 rather than in Stage 3 is analytically significant: L1-029 did not defer contextual reasoning to the post-Symbolab stage but activated it during the evaluation of the tool's output itself, suggesting that for this participant the boundary between Stage 2 and Stage 3 reasoning was permeable. This anticipatory contextual move is precisely what distinguishes EVL-CONT from EXP-FUND: the former evaluates the result's real-world coherence as part of appraising the tool, while the latter explains it as part of synthesizing the full problem-solving process.

5.3 Stage 3: CT Activation Post-Symbolab

5.3.1 Explanation: Justification and Communication of the Result

The Explanation component showed high variability. Two qualitatively distinct types were identified. Superficial explanation (EXP-SUP) limits itself to stating the numerical result, often restating the problem with the calculated value but without justification or contextual connection. Grounded explanation (EXP-FUND) connects the result to the mathematical object's meaning in context and justifies the procedure followed. A recurring error in this section merits attention: four participants (L1-002, L1-007, L1-012, L1-016) stated that both limits equal 2, ignoring that the limit at $d = 2$ is -2 . The error is analytically notable because all four showed high mathematical competence in earlier sections. It is not computational but reflects uncritical acceptance of Symbolab's output without verification against the full set of results, illustrating how the CT-neutral zone can have observable downstream effects in Stage 3 even when not directly coded at Stage 2.

5.3.2 Self-Regulation: Reflection on the Process

Self-regulation was the least frequently activated component in its full dimension. Three levels were identified. Absent self-regulation (AUT-AUS) manifests as omission of any process reflection. Procedural self-regulation (AUT-PROC) involves step verification or error recognition. Reflective self-regulation (AUT-REF), the least frequent but most theoretically significant, involves reflection on the student's own reasoning process and Symbolab's role within it. Four participants showed AUT-REF (L1-001, L1-012, L1-022, and L1-019); three of them (L1-001, L1-012, and L1-022) located their greatest comprehension change during or after Symbolab use, identifying the specific mechanism of that change.

One participant (L1-019) constitutes a singular case: the only one declaring “ample” prior Symbolab experience, they located their greatest comprehension change before using the tool: “Before using Symbolab, because I already knew how to solve the problem almost just by seeing it.” This inversion of the typical pattern, in which the artifact produces the comprehension change, is coherent with instrumental genesis predictions for users with advanced instrumentation: already-established use schemes allow the artifact to confirm rather than construct understanding. This is the clearest evidence in the corpus that instrumental genesis can produce qualitatively distinct use logics depending on the level of instrumentation reached.

Table 1
CT Activation Patterns by Component and Stage (Limits Instrument, n = 29)

CT Component	Stage 1	Stage 2 (CT-neutral / Active)	Stage 3	Key Analytical Finding
Interpretation	Medium	(*)	Medium	Situational INT > Mathematical INT
Analysis	Low (9/29 NE)	(*)	(*)	ANL-OMITIDO co-occurs with INF errors
Inference	Medium	(*)	(*)	INF-ERROR-NO-DETECTADO: undetected error propagation into Stage 2
Evaluation	(*)	CT-neutral: 8/29. Active: EVL-PROC dominant; EVL-CONC 3 cases; EVL-CONT 1 case	(*)	3 inductive subtypes; verbatim identical response in 3 participants
Explanation	(*)	(*)	Medium (EXP-SUP dominant)	Downstream CT-neutral effect: both limits given as 2 by 4 high-competence participants
Self-Regulation	(*)	(*)	Low (AUT-REF: 4 cases)	Inverse pattern in advanced-experience participant (L1-019)
<i>Note. (*) = component not predominant in that stage per the integrated theoretical model. Intensity: Medium = predominance of PE; Low = predominance of NE. Counts describe the analyzed corpus and are not offered as prevalence estimates.</i>				

6. Discussion

6.1 The CT-Neutral Zone: Toward a More Precise Conceptualization

The CT-neutral zone received consistent empirical support. Its coexistence with high mathematical competence in the same participants (L1-007, L1-016) makes implausible any alternative explanation grounded in lack of knowledge or disengagement: a student who can articulate the logic of factoring with precision in Stage 1 does not fail to evaluate in Stage 2 from ignorance, but from a transitory cognitive state linked to the instrumentation process. This is consistent with Rabardel's (1995) account of cognitive-resource distribution during early instrumentalization.

A refinement to the initial conceptualization is that the CT-neutral zone can be interrupted by cognitive-dissonance events: when Symbolab's result does not match the student's expectations, a corrective evaluation activates and breaks the mechanical transcription. The extreme case of L1-025, where an

infinite cost rate was accepted without interrogation even within a clinical context, represents the zone at its most intense (suspension extending even to domain-contextual coherence) and shows that, without active evaluation, the tool's output can displace rather than support mathematical reasoning. These observations connect with Jankvist and Misfeldt's (2015) account of objectified CAS procedures and with Thomas and Hong's (2004) dual-thinking concept; students with extended CT-neutral zones exhibit precisely the pattern these authors describe as cognitively risky.

6.2 Evaluation Subtypes and Their Theoretical Significance

The three inductively emergent evaluation subtypes enrich the theorization of Facione's Evaluation component in the CAS-mediated context. EVL-PROC connects with the procedural thinking that Thomas and Hong (2004) identify as dominant in early CAS users; EVL-CONC reflects the conceptual understanding that Jankvist and Misfeldt (2015) position as absent in the risks of procedure objectification; and EVL-CONT, documented in a single case, anticipates the contextual reasoning that Verschaffel et al. (2020) identify as a higher-order component in word-problem solving. The rarity of EVL-CONT is itself informative: contextual evaluation requires holding the mathematical result, the problem context, and the evaluative connection between them simultaneously, a demand that competes with the instrumentation load of early Symbolab use, in line with Drijvers's (2002, 2015, 2019) argument that the cognitive benefits of CAS depend on task design and the degree of instrumental development reached.

6.3 The Instrument as a Transitional Observation Structure

The nested instrument design constitutes an original methodological contribution. It makes observable what conventional assessment formats leave invisible: the cognitive transition from autonomous pre-tool reasoning to tool-mediated reasoning and back to autonomous post-tool synthesis. As argued in Section 4.3.1, the coding system distinguishes between mentioning a component's vocabulary and operating its sub-skills, so the presence of a section does not guarantee activation of the corresponding component. The substantial variability across the 29 productions (blank sections, mechanical transcriptions, and analytically rich responses alike) demonstrates that the instrument captures genuine differences in cognitive engagement rather than inducing the thinking it observes.

7. Theoretical Contribution

The central theoretical contribution of this study is an empirical-descriptive framework of CT activation during Symbolab-mediated limits problem solving in undergraduate calculus. To our knowledge, the following elements are original to the framework. First, the systematic mathematization of Facione's six CT components for the calculus domain, including the distinction between situational and mathematical interpretation, the CT-complete versus CT-truncated task distinction, and the stage-differentiated activation model. Second, the CT-neutral zone as a structural characteristic of instrumental genesis, with both theoretical characterization and empirical evidence, including the verbatim triple-occurrence pattern and the coexistence with high mathematical competence. Third, the three inductive evaluation subtypes (EVL-PROC, EVL-CONC, EVL-CONT) and two inductive cross-stage codes (ANL-OMITIDO, INF-ERROR-NO-

DETECTADO) that extend the framework beyond its deductive foundations. Fourth, the CT-activation phenomenon as the unit of analysis, permitting description of CT-skill behavior at a level of granularity that student-centered approaches do not offer.

This articulation had not been available previously because the CT-in-mathematics and instrumental-genesis literatures had operated independently. The value of the contribution is that it gives future research a conceptually grounded framework for more precise questions about the relationship between technological mediation and critical reasoning in undergraduate mathematics.

8. Limitations

Several limitations should be borne in mind. In terms of data scope, the empirical results rest on a single instrument applied with 29 participants in one institutional context, covering the limits topic only; generalization beyond this context is not claimed. In terms of participants, the students are students majoring in International Business at one Mexican university, whose relationship with mathematics and with digital tools may differ from other settings. In terms of data type, the study relies exclusively on written productions; the absence of real-time observation such as video analysis or think-aloud protocols limits access to cognitive processes that do not externalize in writing, particularly within the CT-neutral zone, whose very nature implies reduced explicit discourse. In terms of analytic depth, the purposive sampling strategy means systematic in-depth coding concentrated on selected cases that best illustrated each pattern; consistent with reflexive thematic analysis and with the study's analytic unit, the findings should be read as a theoretically grounded description of how CT activation can manifest and transform, not as a quantitative count of pattern frequency. Applying the developed coding system to the full corpus is a line of future work that would support more precise prevalence estimates. Finally, the dual researcher-instructor role introduces potential biases that, although addressed through explicit reflexivity measures and a documented audit trail, cannot be entirely eliminated.

9. Implications

9.1 Implications for Future Research

The most immediate research line is the completion of thematic analysis for the derivatives and integrals instruments, enabling cross-topic comparison of CT activation as instrumental genesis progresses across the semester. Beyond this study, replication in different contexts (other CAS tools such as Wolfram Alpha, GeoGebra, Desmos, and Photomath; other educational levels; and other institutional and cultural settings) would allow evaluation of the framework's generality. A complementary methodological line involves incorporating think-aloud protocols and video analysis to access the cognitive processes occurring within the CT-neutral zone that written productions cannot capture.

9.2 Implications for Teaching Practice

Knowledge of the CT-neutral zone has direct implications for task design: work guides that explicitly require students to articulate their evaluation of tool-generated steps create conditions for CT activation to interrupt the mechanical-transcription state, and the nested instrument offers an adaptable design model for different contexts and tools. The ANL-OMITIDO pattern and its association with subsequent inferential errors suggests that deliberately scaffolding the Analysis stage (prompting students to articulate the structural conditions of the problem before formulating

10. Conclusion

This study described how Facione's cognitive CT skills activate during limits problem solving mediated by Symbolab among 29 first-year students majoring in International Business, from the perspective of instrumental genesis. The six CT components activate differentially across the three stages of instrumental genesis. The CT-neutral zone, defined as a structurally determined interval of transitional CT suspension during mechanical artifact entry, received consistent empirical support and constitutes the study's central theoretical construct. Three inductively emergent evaluation subtypes and two inductive cross-stage patterns enrich the framework. The coexistence of CT-neutral responses with high mathematical competence in the same participants, and the appearance of verbatim identical evaluations across distinct participants, provide the clearest evidence that the zone describes a structural cognitive phenomenon linked to instrumentation, not a product of disengagement or incompetence.

The study contributes an empirical-descriptive framework that, to our knowledge, is the first to articulate systematically Facione's CT components with Rabardel's instrumental genesis processes in undergraduate calculus mediated by a widely used CAS tool. Analysis of the derivatives and integrals instruments is ongoing; the present study reports the first instrument's findings as the empirical baseline from which cross-topic CT transformation will be evaluated in subsequent work.

Declarations

Competing interests.

The author has no relevant financial or non-financial interests to disclose.

Ethical standards.

This study was conducted in accordance with the ethical principles governing educational research at the Instituto Politécnico Nacional. Student participation was voluntary and participants were informed of the study's purpose and procedures prior to data collection. No personally identifying information was retained; all participants are identified in the data exclusively by anonymized codes. The study did not involve any interventions beyond routine classroom activity.

Supplementary materials.

The full data-collection instrument and the complete codebook (operational definitions, textual indicators, and data examples) are available from the author on request and are intended to accompany the article as supplementary material.

Funding.

No funding was received for conducting this study.

Author Contribution

Ceballos-Alexis, E. conceived and designed the study, developed the theoretical framework, designed the data-collection instrument, collected and analyzed the data, interpreted the results, and wrote and revised the manuscript.

Data Availability

The data that support the findings of this study are not publicly available due to ethical restrictions protecting participant confidentiality. The data consist of student-produced problem-solving protocols collected under informed consent agreements that restrict open dissemination. The full data-collection instrument and the complete codebook (operational definitions, textual indicators, and illustrative data examples) are available from the corresponding author upon reasonable request, subject to applicable ethical and institutional requirements.

References

1. Aguilar, M., & Puga, D. (2020). Mathematical help-seeking: Observing how undergraduates use the Internet to cope with a calculus task. *ZDM Mathematics Education*, 52(5), 1003–1016. <https://doi.org/10.1007/s11858-019-01120-1>
2. Aldon, G. (2010). Handheld calculators between instrument and document. *ZDM Mathematics Education*, 42(7), 733–745. <https://doi.org/10.1007/s11858-010-0275-4>
3. Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning*, 7(3), 245–274. <https://doi.org/10.1023/A:1022103903080>
4. Barzel, B., Ball, L., & Klinger, M. (2019). Students' self-awareness of their mathematical thinking: Can self-assessment be supported through CAS-integrated learning apps on smartphones? In G. Aldon & J. Trgalová (Eds.), *Technology in mathematics teaching* (pp. 173–199). Springer. https://doi.org/10.1007/978-3-030-19741-4_4

5. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
6. Buchberger, B. (1990). Should students learn integration rules? *SIGSAM Bulletin*, 24(1), 10–17. <https://doi.org/10.1145/382276.1095228>
7. Buchberger, B. (2002). Computer algebra: The end of mathematics? *ACM SIGSAM Bulletin*, 36(1), 3–9. <https://doi.org/10.1145/565145.565147>
8. Drijvers, P. (2002). Learning mathematics in a computer algebra environment: Obstacles are opportunities. *ZDM Mathematics Education*, 34(5), 221–228. <https://doi.org/10.1007/BF02655825>
9. Drijvers, P. (2015). Digital technology in mathematics education: Why it works (or doesn't). In S. J. Cho (Ed.), *Selected regular lectures from the 12th International Congress on Mathematical Education* (pp. 135–151). Springer. https://doi.org/10.1007/978-3-319-17187-6_8
10. Drijvers, P. (2019). Head in the clouds, feet on the ground: A realistic view on using digital tools in mathematics education. In A. Büchter, M. Glade, R. Herold-Blasius, M. Klinger, F. Schacht, & P. Scherer (Eds.), *Vielfältige Zugänge zum Mathematikunterricht* (pp. 163–176). Springer. https://doi.org/10.1007/978-3-658-24292-3_12
11. Drijvers, P., & Weigand, H.-G. (2010). The role of handheld technology in the mathematics classroom. *ZDM Mathematics Education*, 42(7), 665–666. <https://doi.org/10.1007/s11858-010-0285-2>
12. Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review*, 32(1), 81–111.
13. Evendi, E., Al Kusaeri, A., Pardi, M. H. H., Sucipto, L., Bahalwan, F., & Prayogi, S. (2022). Assessing students' critical thinking skills viewed from cognitive style: Study on implementation of problem-based blended learning model. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(7), em2129. <https://doi.org/10.29333/ejmste/12161>
14. Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. American Philosophical Association.
15. Geiger, V., Goos, M., & Dole, S. (2015). The role of digital technologies in numeracy teaching and learning. *International Journal of Science and Mathematics Education*, 13(5), 1115–1137. <https://doi.org/10.1007/s10763-014-9530-4>
16. Heid, M. K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. *Journal for Research in Mathematics Education*, 19(1), 3–25. <https://doi.org/10.5951/jresmetheduc.19.1.0003>
17. Hoyles, C., Noss, R., & Kent, P. (2004). On the integration of digital technologies into mathematics classrooms. *International Journal of Computers for Mathematical Learning*, 9(3), 309–326. <https://doi.org/10.1007/s10758-004-3469-4>
18. Jankvist, U. T., & Misfeldt, M. (2015). CAS-induced difficulties in learning mathematics? *For the Learning of Mathematics*, 35(1), 15–20.
19. Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

20. Mastuti, A. G., Abdillah, A., Sehuwaky, N., & Risahondua, R. (2022). Revealing students' critical thinking ability according to Facione's theory. *Jurnal Pendidikan Matematika*, 13(2), 261–272. <https://doi.org/10.24042/ajpm.v13i2.13005>
21. McPeck, J. E. (1981). *Critical thinking and education*. St. Martin's Press.
22. Nur Juhainah Ulfa, & Permatasari, B. D. (2023). Critical thinking ability in solving word problems viewed from learning styles. *Jurnal Geliga Sains*, 2(4), 247–262. <https://doi.org/10.58421/gehu.v2i4.70>
23. Orhan, T. Y. (2022). The relationship between critical thinking and problem-solving skills of university students. *Journal of Pedagogical Research*, 6(3), 53–66.
24. Paul, R. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world*. Foundation for Critical Thinking.
25. Permadi, H., & Zakiyah, I. A. (2021). Building student critical thinking skills through the peer assessment learning model in calculus materials with the open-ended approach. *AIP Conference Proceedings*, 2330(1), 040040. <https://doi.org/10.1063/5.0043358>
26. Polya, G. (1945). *How to solve it*. Princeton University Press. <https://doi.org/10.1515/9781400828678>
27. Rabardel, P. (1995). *Les hommes et les technologies: Approche cognitive des instruments contemporains*. Armand Colin.
28. Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
29. Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). Macmillan.
30. Tall, D. (2013). *How humans learn to think mathematically: Exploring the three worlds of mathematics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139565202>
31. Thomas, M. O. J., & Hong, Y. Y. (2004). Integrating CAS calculators into mathematics learning: Issues of partnership. In M. J. Høines & A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 297–304). PME.
32. Trouche, L. (2004). Managing the complexity of human/machine interactions in computerized learning environments: Guiding students' command process through instrumental orchestrations. *International Journal of Computers for Mathematical Learning*, 9(3), 281–307. <https://doi.org/10.1007/s10758-004-3468-5>
33. Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education*, 52(1), 1–16. <https://doi.org/10.1007/s11858-020-01130-4>
34. Wang, Q., & Abdullah, A. H. (2024). Enhancing students' critical thinking through mathematics in higher education: A systematic review. *SAGE Open*, 14(3). <https://doi.org/10.1177/21582440241275651>

35. Webel, C., & Otten, S. (2015). Teaching in a world with PhotoMath. *Mathematics Teacher*, 109(5), 368–373. <https://doi.org/10.5951/mathteacher.109.5.0368>
36. Zain, S. F. H. S., Yusof, N., & Musa, N. (2023). Use of Photomath applications in helping improving student mathematics learning. *Asian Journal of University Education*, 19(1), 60–71.