HOTS in Calculus Assessment: Examining the Gap between Student Capacity and Assessed Practice

Mthobisi Ntuli*[1,2], Aneshkumar Maharaj[2]

[1]Department of Mathematics, Durban University of Technology, South Africa

[2]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa

HOW TO CITE:

Mthobisi Ntuli*, Aneshkumar Maharaj (2026). HOTS in Calculus Assessment: Examining the Gap between Student Capacity and Assessed Practice. International Journal of Special Education, 41(18s), 1625-1635

ABSTRACT:

The development of Higher-Order Thinking Skills (HOTS) is a critical objective in undergraduate mathematics education, particularly in calculus, which serves as a gateway to STEM disciplines. However, standard assessments often privilege procedural fluency over conceptual depth. This exploratory case study investigates how specific assessment task designs can elicit and evaluate HOTS, specifically transfer, critical thinking, and problem-solving among first-year calculus students. Guided by Brookhart's assessment theory and Lithner's framework of Creative Mathematical Reasoning, the study adopts a qualitative task-based interview approach with students from mainstream calculus modules at a South African university. Findings from online assessments and follow-up interviews show a substantial reliance on Algorithmic Reasoning, where students demonstrated procedural competence but struggled with the Creative Mathematical Reasoning required for interpretation and interdisciplinary transfer. The study argues that to bridge the gap between "knowing how" and "knowing why," assessment tasks must be explicitly designed to disrupt rote formula application and require justification

Keywords: Higher-Order Thinking Skills (HOTS), Creative Mathematical Reasoning (CMR), Algorithmic Reasoning (Imitative Reasoning) , Calculus Assessment Design, Procedural vs. Conceptual Fluency

1. Introduction

In the landscape of higher education, first-year calculus is frequently cited as a barrier to student success in STEM fields (Bressoud et al., 2016). A prevailing issue is the gap between the procedural nature of high school mathematics and the conceptual demands of university-level calculus. While many students can perform algorithmic tasks such as finding a derivative using a rule, they often falter when required to apply these concepts to unfamiliar problems or justify their reasoning (Rasmussen et al., 2019). This discrepancy is often attributed to a lack of HOTS.

Students may perform exceptionally well in routine tasks yet lack the ability to take the further step required for deep mathematical understanding (Dubinsky & McDonald, 2001; Maharaj & Wagh, 2016). When students encounter unfamiliar problems, HOTS must be triggered; if these skills are absent, the student reverts to routine memorisation or gives up (King et al., 2016).

Crucially, current assessment practices often fail to detect this conceptual weakness. Standard assessment tasks frequently prioritise procedural correctness over cognitive depth, allowing students to succeed through what Lithner (2017) terms Imitative Reasoning, the application of memorised solution templates to routine problems. Consequently, conventional tests often mask the lack of HOTS, creating an illusion of competence that fails only when students are confronted with non-routine or interpretative tasks.

This study addresses the need for intentional assessment design. If assessment drives learning, then assessments must be constructed to demand HOTS.

The research was guided by the question: *How can the development of HOTS amongst first-year calculus students are assessed?* Specifically, we investigate what types of assessment questions promote the transition from procedural to conceptual understanding.

2. Literature Review and Theoretical Framework

2.1 Higher-Order Thinking Skills (HOTS) in Mathematics

HOTS are characterized by non-algorithmic thinking, involving uncertainty and requiring mental effort to regulate the thinking process. While early definitions relied on Bloom's Taxonomy, recent mathematics education research frames HOTS through the lens of Mathematical Proficiency, which intertwines conceptual understanding, procedural fluency, and adaptive reasoning (Anderson & Krathwohl; Kilpatrick et al., 2001; Groves, 2012). Brookhart (2010) identifies three overlapping abilities essential for HOTS:

Transfer: The ability to relate learning to new situations or different contexts (e.g., applying calculus to physics or geometry). This bridges the gap between abstract formalism and real-world application (Newman, 1990; Perkins & Salomon, 2012).

Critical Thinking: The process of reasoned judgement, self-regulation, and evaluation. In mathematics, this involves deciding what to believe, reflecting on a solution, and producing a reasoned argument (Ennis, 2018).

Problem-Solving: The mechanism behind all thinking, required when a student encounters a problem where the solution path is not immediately obvious (Madu, 2017).

2.2 Analyzing Reasoning: Lithner's Framework

To explain why students succeed or fail in these categories, we employ Lithner's (2008, 2017) distinction between reasoning types:

Imitative Reasoning (IR): This includes Algorithmic Reasoning, where a student executes a memorized set of steps without understanding the underlying mathematical properties. This is often sufficient for routine exams but fails in HOTS tasks.

Creative Mathematical Reasoning (CMR): This is required for HOTS. It involves:

Novelty: Creating a new solution sequence.

Plausibility: Justifying why the strategy is reasonable.

Anchoring: Basing the argument on intrinsic mathematical properties rather than surface cues.

2.3 The South African Context: The Transition from CAPS to University

The challenges associated with HOTS are particularly acute within the South African educational landscape. The transition from the high school Curriculum and Assessment Policy Statement (CAPS) to first-year university mathematics is characterized by a significant articulation gap (Engelbrecht et al., 2005).

Scholars argue that the secondary school assessment system exerts pressure on teachers to teach the test, resulting in a pedagogy that promotes Algorithmic Reasoning. Engelbrecht and Harding (2008) have observed that while entering students are often procedurally fluent, capable of manipulating symbols and solving standard equations, they frequently lack the conceptual depth required for university calculus. The CAPS curriculum, while rigorous in content coverage, often assesses mathematics through predictable templates. Consequently, students are trained to recognize surface cues (Lithner, 2008) and trigger memorized procedures, a habit of Imitative Reasoning that proves fragile when students encounter the non-routine or interpretative tasks typical of university assessments. This reliance on proceduralism acts as a barrier to the Transfer and Creative Thinking skills necessary for success in STEM degrees at South African institutions.

2.4 Conceptual Barriers in Integration

To appreciate the cognitive demand of the assessment tasks used in this study, specifically Questions 4 and 5, it is necessary to review student difficulties with the definite integral. Research suggests that many undergraduate students view integration exclusively as an algebraic process (finding an anti-derivative) rather than a process of accumulation (Riemann sums).

Orton (1983) and Jones (2013) documented that students struggle when the geometric representation of an integral conflicts with algebraic intuition. For example, when calculating integrals for trigonometric functions that cross the x-axis, students conditioned to believe "area must be positive" often fail to distinguish between geometric area and net signed area (Bezuidenhout & Olivier, 2000). Furthermore, setting up an integral for a complex-bounded region requires visualization, a key component of Problem-Solving. Without understanding the integral as a summation of infinitesimal slices ($(x)dx$), students are unable to construct new integrals for non-standard regions. They resort to memorized formulas or incorrectly mapped bounds, failing to anchor their reasoning in the visual properties of the functions.

2.5 Synthesis for Assessment Design

Using these frameworks, we posit that valid calculus assessment must force students out of Imitative Reasoning. For example, asking a student to "interpret a result" (Critical Thinking) or "link a derivative to a physical graph" (Transfer) requires Creative Mathematical Reasoning because memorized algorithms cannot solve these tasks.

To provide a robust framework for analyzing student struggles, this study integrates Brookhart's (2010) assessment categories with Lithner's (2008, 2017) Theory of Mathematical Reasoning.

Table 1: Analytical Framework for Calculus Assessment

Brookhart Category	**Cognitive Demand (HOTS)**	**Associated Reasoning (Lithner)**	**Calculus Indicator**
Transfer	Apply knowledge to unfamiliar contexts (e.g., Physics, Geometry).	**CMR:** synthesizing different domains.	Translating a word problem about motion into $f'(x)$ and $f''(x)$
Critical Thinking	Evaluate mathematical statements; Justify steps; Detect errors.	**Plausibility & Foundation:** Validating why a theorem (e.g., FTC) applies.	Explaining why an integral yield a negative value (net area vs. total area).
Problem-Solving	Solve non-routine problems without a prescribed algorithm.	**Novelty:** Constructing a solution path where no template exists.	Sketching regions and setting up integrals for non-standard boundaries.

The theoretical framing informs the design of non-routine assessment tasks and interviews aimed at prompting students' reasoning processes. By operationalising HOTS through Brookhart's categories and Lithner's framework, the study systematically examines students' reasoning. This approach enables analysis of the shift between imitative and creative mathematical reasoning in calculus contexts.

3. Methodology

3.1 Research Design and Justification

This study employed a qualitative case study approach to investigate how undergraduate calculus students' higher-order thinking skills are assessed and demonstrated. The case study design was selected to enable intensive examination of student reasoning processes in authentic assessment contexts (Merriam & Tisdell, 2016), allowing for detailed analysis of how students engage with complex calculus tasks.

3.2 Justification for Sample Size

While this study examines three student cases, this sample size is consistent with established precedent in mathematics education research that prioritizes depth of cognitive analysis. Comparable studies in mathematics education have effectively employed similar approaches: Zandieh (2000) analyzed 1625-1635 undergraduate students' understanding of the derivative concept through written assessments and interviews; Weber (2005) examined 1625-1635 students' proof-construction and problem-solving processes; and Lithner (2008) used case studies of 1625-1635 students to investigate reasoning types in mathematics. This intensive case study approach allows for detailed cognitive modeling (Schoenfeld, 1985). The approach is particularly suited to exploratory research examining reasoning processes, as advocated by Tall & Thomas (1991), who demonstrated how three to four student cases can reveal fundamental patterns in mathematical thinking.

Following Merriam's (2009) definition, each case is bounded by the individual student's responses to a specific assessment instrument and corresponding interview. The bound nature of each case allows for within-case analysis (examining each student's reasoning trajectory) and cross-case analysis (identifying patterns across the three students).

3.3 Participants and Context

The participants were three first-year students enrolled in mainstream Calculus modules (MATH 130/140) at a South African university. Participants were selected based on their willingness to engage in an in-depth interview process following an online assessment administered and invigilated via MS Teams. The assessment was conducted under test or examination conditions; consequently, students were not permitted to access notes or any additional materials while responding to the assessment.

3.4 Data Collection Instruments

Written Assessment: A purposeful 5-question assessment was designed. Each question targeted specific Brookhart abilities (Transfer, Critical Thinking, Problem-Solving) and was designed to be non-routine to disrupt algorithmic thinking. The assessment was timed (45 minutes).

Interviews: Semi-structured interviews were conducted based on the written responses. The goal was to uncover the students' cognitive processes specifically, why they made certain errors or how they derived correct answers.

3.5 Data Analysis

Written responses were scored on a 0–2 Proficiency Scale to quantify evidence of HOTS:

Level 0 (No Evidence): Student cannot initiate the task or provide irrelevant algebraic manipulation.

Level 1 (Partial/Procedural Evidence): Student completes the algorithmic components (e.g., finding a derivative) but fails to apply the concept to a non-routine context or provide justification.

Level 2 (Clear Evidence of HOTS): Student successfully navigates the non-routine task, demonstrating Novelty, Plausibility, and Anchoring (Lithner, 2017). To deepen the analysis, the qualitative interview data was subjected to a rigorous coding process using deductive thematic analysis (Braun & Clarke, 2019). This process involved three distinct phases:

3.6 Transcription and Familiarization:

Semi-structured interviews were transcribed verbatim. The researchers read through the transcripts multiple times to become immersed in the data, noting initial patterns in student reasoning.

3.7 Deductive Coding against Theoretical Frameworks

The transcripts were coded using a priori categories derived from the study's dual theoretical framework:

Indicators of Imitative Reasoning (Lithner): We explicitly flagged segments where students referenced routine memory or procedural steps without conceptual justification. Keywords and phrases such as “I remembered from high school,” “I just followed the steps,” or “that is how we were taught” were coded as Algorithmic Reasoning.

Indicators of Creative Mathematical Reasoning (Lithner) and HOTS (Brookhart): Conversely, we looked for evidence of Plausibility (checking if an answer makes sense), Anchoring (referencing mathematical properties like “gradient” or “area”), and Transfer (explicitly linking physics concepts to calculus notation). Segments where students recognized errors through reflection such as Student S1 realizing the bounds $[-1,1]$ were incorrect because of the first quadrant constraint were coded as evidence of emerging HOTS.

3.8 Triangulation and Theme Development

Finally, the coded interview segments were triangulated with the students' written scripts to confirm validity. For instance, if a student's written response was mathematically correct but their interview revealed confusion (e.g., Student S3 correctly finding area but being confused by the negative integral), the code was adjusted to reflect Procedural Reliance rather than genuine Critical Thinking. This recursive process resulted in the final three themes: Procedural Reliance, Conceptual Gaps, and Transfer Challenges.

4. Results

The analysis indicates a dissonance between students’ procedural abilities and their conceptual grasp, consistent with what Lithner (2017) describes as a dependency on Algorithmic Reasoning. Table 2 summarizes the performance.

Table 2: Student Performance by HOTS Category

Question	Targeted HOTS	Performance Summary	Dominant Reasoning Type (Lithner)
1	Problem Solving	Level 1: All students found y' but failed to link $m = \tan 60$	Algorithmic: Stuck once th standard template ended.
2	Transfer (Physics Context)	Level $1-2$: Unable to translate physica description into f, f', and f'' graphs.	Imitative: Knowledge was “siloed” within math symbols.
3	Creative Thinking (FTC)	Level $1-2$: $S1$ succeeded; $S3$ struggle with the non-standard limit (x^2)	CMR vs. IR: $S3$ relied on the “limi is x” template
4	Critical Thinking (Interpretation)	Level 1: Correct calculation of the area but viewed as an “error”.	Algorithmic: Trusted th procedure over the concept.
5	Problem Solving (Visualization)	Level 1: $S1$ used roots $[-1,1]$ ignoring the first quadrant constraint.	Surface Cues: Relied on algebr over geometric reality.

The following subsections detail the analysis of these obstacles.

Question 1: Critical Thinking and Problem-Solving

1. Find the points P and Q on the parabola $y = 1 - x^2$ so that the triangle ABC formed by the x-axis and the tangent lines at P and Q is an equilateral triangle. (See Figure 2)

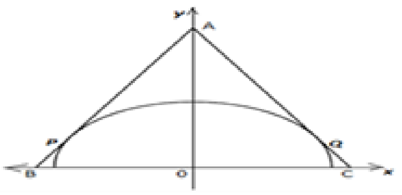

Figure 2: Showing parabola $y = 1 - x^2$ and triangle ABC

All students achieved Level 1. They correctly derived $y' = 2x$ (Imitative Reasoning) but could not link this algebraic result to the geometric property of an equilateral triangle.

Qualitative Insight: Student *S3* noted**:** *"I knew that I had to use the derivative... but I was not sure if the derivative gave me the equation or the slope... The problem started when I derive the equation."*

Analysis: The student mastered the procedure (differentiation) but lacked the conceptual anchor that the derivative represents the gradient of the tangent, which equals $ta(\theta)$. Without a template for equilateral triangles in calculus, the student could not construct a solution.

Question 2: The Challenge of Transfer

2. Suppose that a car is stationary at a robot. As soon as the robot turns green the driver of the car accelerates, moving the car for a particular time. Upon noticing that the next robot is red the driver applies brakes causing the car to decelerate until it stops. Assume that the car starts at $t = 0$ and stops at $x = 6$.

2.1 On the same set of axes, show the graphs of three functions which represent the position of the car, velocity of the car and its acceleration, respectively.

2.2 In term of derivatives, which of your functions will represent the original function f, f' and f''. Explain your choices.

[Ability in Table 1 focused on: Identify possible applications of mathematics in their surroundings and work systematically through cases in an exhaustive way, with a target on Transfer and Critical Thinking]

Students struggled significantly. *S1* and *S2* scored 0 on the graphs. They were unable to sketch the graphs of velocity and acceleration based on a description of a car's motion.

Qualitative Insight: Student *S1* stated: *"The question asked for two graphs representing the velocity and acceleration of the car. I was not sure how to use the information given to draw the graphs. I did not know where to start or what the graphs would look like."*

Analysis: This is a classic Transfer deficit. As noted by Planinic et al. (2018), students often compartmentalize knowledge. The student treated this as a "Physics" problem and could not transfer their Calculus knowledge of derivatives (slopes) to solve it.

Question 3: Creative Thinking and Non-Standard Notation

3.1 If $x \sin(\pi x) = \int_0^2 (t)\, dt$, were f is a continuous function, find $f(4)$. *[Ability in Table 1 focused on: Identify possible applications of mathematics in their surroundings and work systematically through cases in an exhaustive way, with a target on*

S3 demonstrated Level 1 performance, showing difficulty differentiating x sin πx, whereas *S1* achieved Level 2 by correctly applying the Fundamental Theorem of Calculus (FTC) together with the chain rule.

Qualitative Insight: *S1* admitted: *"Normally, we are given a simple problem, and the upper limit is just x."*

Analysis: This confirms Lithner's warning about algorithmic teaching. Students memorize the "template" for FTC. When the task changed slightly (requiring Creative Thinking to adapt the method), they could not proceed because their knowledge was rigid, not flexible.

. **Question 4:** Procedural Reliance vs. Critical Thinking

4. Let $(x) = \cos x - \sin x,\ x \in [0, \frac{}{4}]$ 9.1 Sketch the graph f. 9.2 Evaluate $\int_0^{4} (x)\ d(x)$ and interpret the results in terms of area. *[Ability in Table 1 focused on: Interpret a general definition or statement in the context of a given model and critically evaluate their and others' presented solutions to a problem, with a target on Critical Thinking.]*

Students correctly calculated the definite integral (finding a negative value) but failed the interpretation step.

Qualitative Insight: Student *S1* remarked: "*I did not understand why the area was negative, because my calculation was correct, I even did it twice just to check if I did not make a mistake.*"

Analysis: This finding is significant. It demonstrates that the student equates integral with geometric areas without Critical Thinking. They could not evaluate the validity of their result against the concept of net signed area. They trusted the procedure (calculation) more than the mathematical concept, a hallmark of Imitative Reasoning.

Question 5: Problem-Solving and Visualisation

5. Draw sketches and set up a definite integral that gives the area of the region in the first quadrant bound by the x-axis, the line, $y = \sqrt{3}x$ and the circle $x^2 + y^2 = 4$. *[Ability in Table 1 focused on: Work systematically through cases in an exhaustive way, with a focus on Critical Thinking, Creative Thinking, and Problem-Solving]*

All students scored at Level 1, demonstrating only partial understanding. Although they could complete isolated algebraic steps such as finding the intersection points, they were unable to construct the definite integral correctly. Specifically, S1 utilized the calculated roots directly as limits $[-1,1]$, ignoring the first quadrant constraint. S3 identified the correct domain [0,1] but misidentified the bounding functions for the integrand.

Qualitative Insight: Interview data reveals a prioritization of calculation over geometric verification. *S1* justified the incorrect limits by citing the algebraic result rather than the visual region: *"When I calculated the point of intersection, I found that the graphs meet at the point $x = \pm 1$, so I used that as my limits since I am integrating with respect to x."*

S3 attempted to visualize the region but failed to map the text to the graph:

"The question stated that the area is bounded by the $x - axis$... So, my region was the bottom half [pointing to the wrong area]."

Analysis: Successful reasoning here requires Plausibility (checking if a limit of -1 is valid in the first quadrant) and Anchoring (grounding the algebra in intrinsic geometric properties). *S1*'s reasoning was anchored only in surface algebraic cues. The student trusted the algorithm over the geometric reality, demonstrating that while they possess the procedural tools of calculus, they lack the reasoning skills to synthesize them into a novel solution.

4.2 Thematic Analysis of Interviews

The interview data revealed three dominant themes regarding why HOTS development is stalled:

Procedural Reliance (8/12 coded segments): Students defaulted to "following steps from class." When steps didn't work, they stopped. This aligns with surface-level learning approaches.

Conceptual Gaps (6/12 segments): Concepts like "net area" vs "total area" were missing. Students focused on how to integrate, not what integration means.

Transfer Challenges: Students viewed calculus as isolated from real-world contexts (kinematics) or other mathematics domains (geometry).

5. Discussion

The results reveal concerning patterns that warrant investigation with larger cohorts: students possess the procedural tools of calculus (e.g., differentiating polynomials, calculating definite integrals) yet lack the HOTS to apply these tools when standard templates are removed. The prevalence of Imitative Reasoning (Lithner, 2017) observed across the assessment tasks is not merely a reflection of student ability, but an indicator of an instructional culture that may unintentionally prioritize algorithmic performance over conceptual understanding.

The following discussion outlines how these findings should inform the broader teaching and learning landscape to foster CMR

5.1 Shifting from Algorithmic Demonstration to Concept Construction:

The failure of students to adapt the Fundamental Theorem of Calculus in Question 3 (involving the upper limit x^2) suggests that teaching often emphasizes the replication of standard prototypes. When students are exposed primarily to routine examples during instruction, they develop solution templates rather than deep conceptual networks.

5.2 Implication for Teaching and Learning

Instructional strategies must shift from the demonstration of algorithms to the construction of concepts. To disrupt the reliance on Algorithmic Reasoning, lectures and learning materials should adopt a Productive Struggle approach (Warshauer, 2015). Instead of presenting solutions to standard problems, lecturers should expose students to tasks that lack immediate solution paths such as the integration of functions over undefined regions. Teaching should value the process of deriving the mathematical model (Plausibility) as much as the final calculation. This involves making the limitations of algorithms visible during lectures, explicitly showing why a standard formula fails when conditions change, thereby forcing students to engage in Creative Mathematical Reasoning to adapt.

5.3 Teaching for Transfer through Interdisciplinary Modeling

In Question 2, students' difficulty in graphing velocity and acceleration from a physical description may indicate a tendency to engage with mathematical ideas predominantly in symbolic form. This pattern points to a possible compartmentalization of knowledge, where mathematical procedures are more readily activated in explicitly algebraic contexts (e.g., tasks framed with cues such as "find") but are less accessible when embedded in descriptive or applied scenarios. While not conclusive, this observation raises important questions about the integration of mathematical understanding across representational forms and contexts.

5.4 Implication for Instruction

To facilitate Transfer, a core component of HOTS teaching strategies must bridge the gap between abstract formalism and applied context. Calculus instruction should move beyond standard variable notation (x and y) to incorporate multi-representational modeling (Rasmussen et al., 2019). Lecturers should introduce physical scenarios as primary instructional tools, requiring students to translate these into mathematical notation before any calculation occurs. By routinely varying variables (e.g., utilizing t for time or P for population) and context during concept introduction, lecturers can prevent students from associating calculus concepts solely with algebraic manipulation, thereby fostering the flexibility required for transfer.

5.5 Prioritizing Interpretation and Visualization

The results from Question 4, where students correctly calculated a negative integral but failed to interpret it as net signed area, highlight a significant deficit in Critical Thinking. Students viewed the negative sign as a calculation error to be fixed rather than a meaningful representation of displacement. This indicates a calculational orientation (Thompson, 2008) where the goal of learning is seen as producing a number rather than generating meaning.

5.6 Implication for Classroom Culture

To develop Critical Thinking, the teaching and learning culture must prioritize Visualization and Interpretation over computation. Pedagogical approaches should incorporate Predict-Verify cycles. For example, before teaching integration techniques, students should be tasked with sketching regions and predicting whether a result will be positive or negative based on visual inspection. By requiring students to justify the plausibility of an answer before computing it, lecturers can help anchor algebraic procedures in geometric reality. This shifts the learning objective from simply getting the right answer (Imitative) to understanding what the answer means (Creative).

6. Conclusion

This study demonstrates that while first-year calculus students often possess high levels of procedural fluency, this competence frequently masks a profound deficit HOTS. By utilizing an assessment framework integrated with Lithner's theory of reasoning, we revealed that students rely heavily on Imitative Reasoning successfully executing memorized algorithms while struggling to transfer these concepts to physical contexts, visualize geometric regions, or critically interpret mathematical results. Standard assessments that prioritize calculation often fail to detect these gaps, creating a "false positive" regarding student readiness.

Therefore, the design of assessment tasks is not merely a technical matter of grading but a pedagogical imperative. To effectively foster HOTS, assessment tasks must systematically disrupt standard solution templates, neutralize the advantage of rote memorization, and explicitly demand justification and interdisciplinary transfer.

6.1 Implications for the STEM Pipeline

Ultimately, these findings have urgent implications for the integrity of the STEM pipeline. First-year calculus serves as a foundational course not just to filter students, but to equip them with the conceptual tools necessary for advanced Physics, Engineering, and Economics. If students pass this module through algorithmic impression without acquiring the ability to reason creatively or transfer knowledge, they are set up for failure in subsequent years where calculus becomes a tool for modeling complex reality, not just an exercise in symbol manipulation. Bridging the gap between "knowing how" and "knowing why" is therefore essential not only for mathematical success but for ensuring that graduates are truly prepared for the cognitive demands of the modern scientific workforce.

7. Recommendations for Assessment Design

To foster HOTS, assessment tasks must evolve:

Disrupt Templates: Assessments must intentionally vary standard problems (e.g., changing limits of integration, using graphical prompts instead ofalgebraic ones) to prevent students from using Imitative Reasoning. Valuation of Justification: Marks should be awarded explicitly for the reasoning process. Questions such as "Explain why this result is negative" force Critical Thinking. Interdisciplinary Tasks: Calculus assessments should include non-routine problems or physical scenarios (like the car problem) to enforce Transfer skills, requiring students to build the mathematical model themselves.

8. Limitations

The primary limitation of this study is the small sample size, which prevents broad statistical generalization. However, as an exploratory case study, it demonstrates the efficacy of using Lithner's and Brookhart's frameworks to diagnose specific stalling points in student reasoning. Future research should apply this framework to a larger cohort using a mixed-methods approach to determine the prevalence of these reasoning gaps across the STEM pipeline.

9. Ethics Statements

To ensure the ethical conduct of the study and obtain permission to involve students, approval was sought from the module coordinators responsible for the relevant modules. Additionally, ethical clearance was granted by the Humanities and Social Sciences Research Ethics Committee of a university in South Africa. All participants provided informed written consent after receiving detailed information about the study's purpose, procedures, and their right to withdraw. All interview recordings and transcripts will be stored in a secure hard drive and destroyed 5 years after study completion.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Bezuidenhout, J., & Olivier, A. (2000). Students' conceptions of the integral. *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education, 2*, 73–80.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597.

Bressoud, D., Mesa, V., & Rasmussen, C. L. (2016). *Insights and recommendations from the MAA national study of college calculus*. MAA Press.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria, VA: ASCD.

Engelbrecht, J., & Harding, A. (2008). The impact of the transition to outcomes-based teaching on university preparedness in mathematics in South Africa. *Mathematics Education Research Journal, 20*, 57–70. https://doi.org/10.1007/BF03217477

Engelbrecht, J., Harding, A., & Potgieter, M. (2005). Undergraduate students' performance and confidence in procedural and conceptual mathematics.

Ennis, R. H. (2018). Critical Thinking Across the Curriculum: A Vision. *Topoi, 37*(1), 165–184.

Groves, S. (2012). Developing mathematical proficiency. *Journal of Science and Mathematics Education in Southeast Asia, 35*(2), 119–145.

Jones, S. R. (2013). *Understanding the definite integral as a sum of infinitesimal pieces.*

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding It Up: Helping Children Learn Mathematics*. Washington, DC: National Academy Press.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics, 67*(3), 255–276.

Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM Mathematics Education, 49*(6), 937–949.

Madu, A. (2017). Higher Order Thinking Skills (HOTS) in Math Learning. *IOSR Journal of Mathematics (IOSR-JM), 13*(5), 70–75.

Maharaj, A., & Wagh, V. (2016). Formulating tasks to develop HOTS for first-year calculus based on Brookhart abilities. *South African Journal of Science, 112*(11/12).

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Jossey-Bass.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Newman, F. M. (1990). Higher Order Thinking in Teaching Social Studies: A rationale for the assessment of Classroom Thoughtfulness. *Journal of Curriculum Studies, 22*(1), 41–56.

Orton, A. (1983). *Students' understanding of integration.*

Perkins, D. N., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist, 47*(3), 248–258.

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measure? *Physical Review Physics Education Research, 15*(2).

Rasmussen, C., et al. (2019). Student understanding of the relation between tangent line and rate of change. *International Journal of Research in Undergraduate Mathematics Education, 5*(1), 5–26.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic Press.

Tall, D., & Thomas, M. (1991). Encouraging versatile thinking in algebra using the computer. *Educational Studies in Mathematics, 22*(2), 125–147.

Thompson, P. W., & Harel, G. (2018). *Ideas revealed: A comprehensive view of the vital role of notation in mathematical understanding*. Springer.

Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education, 18*(4), 375–400.

Weber, K. (2005). Problem-solving, proving, and learning. *Journal for Research in Mathematics Education, 36*(2).

Zandieh, M. (2000). *A theoretical framework for analyzing student understanding of the concept of derivative*. In *CBMS Issues in Mathematics Education* (Vol. 8, pp. 103–127).