Journal on Mathematics Education
Volume 16, No. 2, 2025, pp. 559-580

A SOLO Taxonomy-based rubric for assessing conceptual understanding in applied calculus

Patrick John Martinez Fernandez*, Angela Fatima Hilado Guzon

Department of Mathematics, Ateneo de Manila University, Quezon City, Philippines
*Correspondence: patrick.fernandez@student.ateneo.edu

Received: 22 January 2025 | Revised: 25 March 2025 | Accepted: 18 April 2025 | Published Online: 21 April 2025

Abstract

Assessing conceptual understanding in mathematics remains a persistent challenge for educators, as traditional assessment methods often prioritize procedural fluency over the complexity of connections between mathematical ideas. Consequently, these methods frequently fail to capture the depth of students' conceptual understanding. This paper addresses this gap by developing and applying a novel rubric based on the Structure of Observed Learning Outcomes (SOLO) Taxonomy, designed to classify student responses according to demonstrated knowledge capacity and cognitive complexity. The rubric introduces transitional levels between the main SOLO categories and includes provisions for evaluating unconventional solutions, enabling a more nuanced assessment of student work based on knowledge depth and integration. The rubric was constructed through an analysis of the conceptual knowledge components required to solve each problem, validated by expert review, and guided by criteria aligned with SOLO level classifications. It also incorporates qualitative feedback to justify each SOLO level assignment. Using this rubric, the study analyzed responses from 57 first-year undergraduate students—primarily chemistry and computer science majors at a private university in the Philippines—to test items on linear approximations and the Extreme Value Theorem. Interrater reliability was established through weighted Cohen's kappa coefficients (0.659 and 0.667 for the two items). The results demonstrate the rubric's capacity to differentiate levels of conceptual understanding and reveal key patterns in student thinking, including reasoning gaps, reliance on symbolic manipulation, and misconceptions in mathematical logic. These findings underscore the value of the SOLO Taxonomy in evaluating complex and relational thinking and offer insights for enhancing calculus instruction. By emphasizing the interconnectedness of mathematical ideas, the study highlights the potential of conceptually oriented assessments to foster deeper learning and improve educational outcomes. Furthermore, the rubric's adaptability suggests its applicability beyond calculus, supporting a broader shift toward concept-focused assessment practices in higher education.

Keywords: Applied Calculus, Conceptual Understanding, SOLO Taxonomy

How to Cite: Fernandez, P. J. M., & Guzon, A. F. H. (2025). A SOLO Taxonomy-based rubric for assessing conceptual understanding in applied calculus. *Journal on Mathematics Education, 16*(2), 559–580. https://doi.org/10.22342/jme.v16i2.pp559-580

The effective evaluation of conceptual understanding in foundational university-level mathematics courses such as applied calculus is a challenge for educators. While mathematical proficiency comprises multiple dimensions, conceptual understanding remains particularly difficult to evaluate systematically. Kilpatrick et al. (2001) identified five strands of mathematical proficiency: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition toward mathematics—these are all interwoven and interdependent. In higher education mathematics, the necessity to develop all these strands are more prominent: a student in higher mathematics should be able to relate mathematical concepts, to execute mathematical procedures fluently, to represent and solve mathematical problems, to reason logically, and to maintain a positive disposition toward mathematics and learning mathematics. This demand for mathematical proficiency is particularly crucial in foundational but conceptually challenging courses such as differential calculus (Santos-Trigo et al., 2024).

Traditionally, mathematics teaching has predominantly focused on developing procedural skill, perhaps due to mathematics teachers' "conception of mathematics as a static body of knowledge, involving a set of rules and procedures that are applied to yield one right answer" (Stipek et al., 2001). This procedural focus is reinforced when teachers generally point out procedural errors when giving feedback on students' solutions (Stovner et al., 2021). However, procedural skill is not enough to be successful in learning mathematics in a higher education context, as higher education mathematics demand more competencies: mathematical thinking, problem handling, modelling, reasoning, representation, and communication (Büchele & Feudel, 2023).

An emphasis on procedural skill creates students who are fluent in executing procedures (e.g., how to calculate derivatives) but possibly lack depth in understanding concepts (e.g., why the derivative gives the slope of a line tangent to a curve). This also leads to students' inability to see the connection between related concepts (e.g., why a curve is not differentiable at a sharp corner, or why differentiability at a point implies continuity at that point) (Arslan, 2010). Therefore, the need for mathematics teaching to also place an emphasis on conceptual understanding is imperative. This call for an increased focus on conceptual understanding in mathematics teaching is not new—it has been recognized as early as 25 years ago (Rittle-Johnson & Alibali, 1999). However, the persistence of research on the interplay between conceptual understanding and procedural fluency, as well as on concepts-first versus procedures-first approaches, indicates that it remains a pressing need in mathematics education.

Kilpatrick et al. (2001) define conceptual understanding as "an integrated and functional grasp of mathematical ideas," which entails more than knowing facts in isolation. It extends to a connected network of concepts, where "the degree of students' conceptual understanding is related to the richness and extent of the connections they have made" (Kilpatrick et al., 2001).

While frameworks like Bloom's Taxonomy (Bloom et al., 1964) have been widely used for categorizing learning goals and designing assessments, they are not always suitable for evaluating open-ended responses in a systematic way. According to Biggs and Collis (2014), Bloom's Taxonomy "is really intended to guide the selection of items for a test rather than to evaluate the quality of a student's response to a particular item" and is "based on judgments about quality, which may be arbitrary" (p. 13). In other words, the levels in Bloom's Taxonomy are *a priori*, set by the teacher as standards against which student outcomes are measured. For example, an evaluator using Bloom's Taxonomy might classify a solution to an optimization problem at the "application" level, depending on the student's use of derivatives to find extrema. However, this does not capture whether the student understands the connection of the procedure to underlying concepts.

In response to this limitation of Bloom's Taxonomy as a rubric for evaluating student responses to assessments, Biggs and Collis (2014) developed the Structure of the Observed Learning Outcome (SOLO) Taxonomy, whose levels "arise naturally in the understanding of the material" (p. 13). Unlike Bloom's Taxonomy, which classifies types of cognitive processing, the SOLO Taxonomy evaluates the structural quality of a response—both breadth and depth of conceptual integration. Thus, the SOLO Taxonomy is aligned for assessing conceptual understanding in mathematics, where the "richness and extent of connections" is its defining characteristic.

Unlike Piaget's levels of cognitive development which aims to categorize a student's general cognitive structure, the SOLO Taxonomy aims to evaluate a particular response of a student to a learning task. This is important because a student may demonstrate differing levels of learning in different subjects, and they may also even demonstrate differing levels of learning within the same subject in different time periods. In other words, the labels are shifted "from the student to his response to a particular task" (Biggs & Collis, 2014). In the SOLO Taxonomy, the complexity of a response to a learning task may be classified to one of several levels, based on the number of relevant concepts demonstrated (capacity) and the connections made between these concepts and the cue to form a logical, consistent response (relating operation):

1. Prestructural: Response represents the use of no relevant concept;
2. Unistructural: Response represents the use of one relevant concept;
3. Multistructural: Response represents the use of several disjoint concepts;
4. Relational: Response represents the use of all related concepts in an integrated whole; and
5. Extended abstract: Comprehensive use of all relevant concepts together with related hypothetical constructs and abstract principles (Radmehr & Drake, 2019).

Movement from the prestructural to the unistructural and to the multistructural levels indicate a development in knowledge quantity—that is, a student knows more things, but they may not be well-integrated. On the other hand, movement from the multistructural to the relational level and to the extended abstract level indicate a development in knowledge quality—that is, a student is able to integrate the isolated pieces of information they know.

Because student responses to a specific task may vary greatly, it is possible that a particular response may not fit exactly in one of these five levels. The SOLO Taxonomy allows for the possibility of a response to be classified into a transitional level. In general, transitional responses

> *… tend to be marked by confusion or inconsistency. It is as if the student is handling more information than he can cope with in his working memory, and he loses track of his argument. Typically, transitional responses carry more information than is usual in the level the student is emerging from, but he is forced to give up before reaching the complexity of structure that is expected at the next SOLO level* (Biggs & Collis, 2014, p. 29).

While studies such as Bisson et al. (2016) and Crooks and Alibali (2014) developed frameworks to assess conceptual understanding, these studies have mainly focused on K-12 education or theoretical aspects of assessing conceptual understanding. Additionally, the few which use the SOLO Taxonomy in mathematics education mainly focused on the K-12 levels (Claudia et al., 2020; Mukuka et al., 2020; Putri et al., 2017; Sudihartinih, 2019), and even fewer have higher education participants (Mulbar et al., 2017; Ramos et al., 2024). None of these have specifically looked into applied calculus courses in higher education, where conceptual understanding is particularly important. Thus, even with the recognition of the importance of conceptual understanding, there is a gap in assessment frameworks for applied calculus contexts: there is an absence of structured rubrics based on the SOLO Taxonomy specifically designed to evaluate the structural complexity of conceptual connections.

This paper addresses the question: How can the SOLO Taxonomy be adapted and applied to evaluate conceptual understanding in university-level applied calculus? The study has a twofold focus: first, to develop a SOLO Taxonomy-based rubric specifically for evaluating conceptual understanding in applied calculus; and second, to explore its effectiveness through its application to solutions from 57 first-year undergraduate students to test items. The rubric's effectiveness was measured with respect to interrater reliability, its ability to distinguish between levels of conceptual understanding, and its ability to identify patterns in conceptual understanding. While the primary context of the study is undergraduate mathematics, the findings may have broader implications for mathematics education in various levels.

METHODS

Development of the Rubric

The development of the SOLO-Taxonomy based rubric began with a literature review to search for frameworks to evaluate conceptual understanding that aligned with Kilpatrick et al.'s (2001) definition of conceptual understanding. Finding no such existing framework, the SOLO Taxonomy (Biggs & Collis, 2014) was identified as a promising foundation. The SOLO Taxonomy was then deconstructed into its core components of capacity and relating operation. In this adaptation, capacity was operationalized as the number of relevant conceptual knowledge components demonstrated in a solution, and relating operation was interpreted as the degree of integration among these conceptual knowledge components. This bridges the aspects central to classifying a response in the SOLO Taxonomy to Kilpatrick et al.'s (2001) definition of conceptual understanding.

Instead of creating a rubric tailored to each item, a general rubric which could be applied to various problems was developed. The initial rubric comprised SOLO level descriptors specific to mathematics and expounded on transitional levels to capture nuanced differences. This is particularly important in mathematics where conceptual development happens in small steps rather than large jumps (Wilson, 2009). For example, a prestructural response is marked by fragmented or incorrect concepts that lack any meaningful connection to a problem or by a lack of engagement with a problem. In contrast, a relational response demonstrates a well-integrated understanding of calculus concepts and the ability to correctly reason through the connections among calculus concepts. After developing the initial rubric, it was tested informally with some of the collected solutions, leading to iterative refinements until the rubric could effectively distinguish between different levels of conceptual understanding. This process ensured that the rubric remained theoretically grounded while being practically applicable to evaluating conceptual understanding. While this approach is not as rigorous as formal validation procedures, the substantial interrater reliability provides post-hoc evidence supporting the rubric's validity for its intended purpose.

To meaningfully use this rubric to evaluate student solutions, the following evaluation procedure was utilized.

1. Identifying conceptual knowledge components
   For each item, the researchers identified the conceptual knowledge components necessary to answer the problem. A conceptual knowledge component is a concept present in an expected solution based on the course content. This list of knowledge components is then validated by a teacher who has been teaching university-level calculus and real analysis for over ten years.
2. Identifying demonstrated knowledge components
   A solution is thoroughly read to classify whether each conceptual knowledge component identified in the list from the previous step is explicitly expressed, implied, is demonstrated incorrectly, or not demonstrated at all by the solution. A conceptual knowledge component is explicitly expressed if the student clearly states or articulates it by stating it directly. This shows that the student correctly understands and can communicate the concept. Otherwise, if the solution shows that the

conceptual knowledge component is present without being explicitly stated, it is classified as implied. If a solution attempts to demonstrate a conceptual knowledge component but does not do so correctly, then the conceptual knowledge component is classified as incorrectly demonstrated. Finally, a conceptual knowledge component is not demonstrated if it is not present, whether explicitly demonstrated or otherwise.

Note here that a solution may not demonstrate all the conceptual knowledge components listed in the previous step. This is especially true if a student uses a different solution than expected. In these cases, the conceptual knowledge components not in the list are taken note of, including whether each of these is explicitly expressed, implied, or incorrectly demonstrated.

3. Identifying the solution's SOLO level

   Based on the number of conceptual knowledge components demonstrated by the solution and how well these demonstrated conceptual knowledge components are integrated, the solution is then classified into a SOLO level using Table 1. Again, care must be taken when evaluating solutions with unconventional approaches. A solution that does not demonstrate the expected solution must not be penalized; instead, the evaluation focuses on the relevance, quantity, and quality of the demonstrated conceptual knowledge components and their connections. Moreover, because the presence of verbal explanations is a crucial indicator of being able to integrate multiple conceptual knowledge components into a coherent solution, a solution that does not have verbal explanations (or if the lack of verbal explanations hinders the clarity of the integration) will be penalized. In particular, if a solution consists of mostly mathematical expressions without explicit verbal explanations, it will be evaluated at the highest SOLO level demonstrated within the mathematical expressions but adjusted downward by one level only if the lack or absence of verbal explanations decreases the clarity of integration among the knowledge components implied by the mathematical expressions.

4. Providing qualitative feedback

   Qualitative feedback was given in order to further justify the choice of SOLO level for the solution. This may include whether the solution is unconventional or lacks verbal explanations. In addition, solutions which are classified in the transitional MR level are given feedback on whether the demonstrated conceptual knowledge components are connected with low integration (i.e., some weak or incomplete connections are made between components, showing an emerging understanding of relationships between concepts) or high integration (i.e., logical and consistent connections are present among the demonstrated conceptual knowledge components, but there are some gaps in relating them or the solution lacks full cohesion). In an analogous manner, it is understood from the qualities of each SOLO level that solutions classified from the prestructural level up to the multistructural level lack any form of integration among conceptual knowledge components, and that solutions classified from the relational level up to the extended abstract level are marked by a full integration of conceptual knowledge components.

In this rubric, transitional levels are included to address solutions that fall in between the five main SOLO levels summarized in Table 1. These levels capture the nuances present in responses that do not quite reach one level's characteristics but also exhibit more than the previous level's characteristics. The inclusion of these transitional levels enables a closer examination of solutions that may otherwise be classified ambiguously.

Table 1. Criteria for classifying a solution based on capacity and relating operation

SOLO Level	Description	Indicators
Prestructural	The solution lacks any relevant knowledge of the topic or demonstrates clear misconceptions. There is no attempt to connect information.	Solution lacks understanding, with irrelevant or incorrect information. May resort to denial (e.g., refusing to engage), tautology (repeating the question), or transduction (providing a perceptual guess).
Transitional: PU	The solution attempts to recognize a single relevant concept but lacks comprehension or meaningful development.	Attempts to identify a relevant concept but does not demonstrate understanding. Minimal engagement with the problem.
Unistructural	The solution demonstrates understanding of only one relevant concept, with no integration of other ideas.	Accurately identifies and addresses one knowledge component. Lacks other relevant knowledge components, so no connections are present.
Transitional: UM	The solution attempts to recognize more than one concept but lacks comprehension or meaningful development.	Unsuccessful demonstration of more than one knowledge component, so in the end, only one knowledge component is completely used in the solution.
Multistructural	The solution demonstrates multiple distinct pieces of relevant information but lacks any integration between them.	Correctly addresses several knowledge components in isolation. Shows awareness of multiple knowledge components but treats them independently.
Transitional: MR	The solution shows emerging understanding of connections between multiple concepts, but these connections are incomplete or unclear.	Begins to relate multiple knowledge components but lacks full logical coherence. Partial, emerging relational understanding.
Relational	The response demonstrates a fully integrated understanding of relevant concepts, applying them coherently to solve the problem.	Accurately integrates necessary knowledge components in a cohesive solution. Solution is logically consistent with no inconsistencies.
Transitional: RE	The solution shows awareness of the potential for abstraction or extension beyond the problem context but lacks complete clarity in this generalization.	Attempts abstraction or generalization beyond the given problem but is unclear.
Extended Abstract	The solution demonstrates abstract reasoning, applying integrated concepts in innovative or generalized ways that extend beyond the problem.	Extends solution by moving beyond problem constraints to form new hypotheses. Demonstrates clear resolution of inconsistencies, even beyond expected knowledge boundaries.

Participants and Setting

Participants of the study are students enrolled in two classes of an applied differential calculus course offered to first-year undergraduate students in a private university in Quezon City, Philippines in the second semester of Academic Year 2023-2024. Each of the two classes were almost homogeneous: one class consisted mostly of chemistry majors, and the other mostly of computer science majors. While the participants came from two different majors, several factors minimized potential confounding effects. Because the participants are first-year undergraduate students, they have not taken any university-level mathematics courses yet. Thus, they have similar mathematics backgrounds from senior high school. Moreover, both classes were taught by the same instructor using identical course materials and teaching strategies. Lastly, since the focus of the study is on conceptual understanding rather than discipline-specific applications, the likelihood of performance differences between majors is reduced.

The study was conducted with approval from the university's Research Ethics Office, with approved protocol code SOSEREC_23_014. All 57 participants were briefed about the study prior before voluntarily giving informed consent. Data were anonymized prior to analysis and were kept in a secure storage location.

Data Collection

The primary data for this study consisted of written student solutions to two conceptual items from the researcher-developed Derivatives Concepts and Skills Test (DCST). The DCST was developed collaboratively with the experienced course instructor to ensure alignment with course objectives, establishing content validity through expert judgment. However, as the DCST served as a summative assessment of the course, formal pilot testing was not conducted to maintain test security, a common limitation in classroom-based research.

The test items were designed to assess conceptual understanding and procedural fluency on topics relating to the derivative (implicit differentiation, related rates of change, linear approximations and differentials, local and absolute extrema of functions, how derivatives affect the shape of a graph, indeterminate forms and L'Hôpital's Rule, and optimization). For this paper, full solutions to two items in the DCST of 57 students were collected and digitized. These two items are as follows:

> *Item 5: In physics and engineering, the approximation $sin\, x \approx x$ is sometimes used. Under what condition/s is this approximation reasonable?*
> *Item 7: The extreme value theorem states that a continuous function f on an interval $[a, b]$ attains attains an absolute minimum and an absolute maximum in $[a, b]$. Will the theorem still be true if we replace the interval $[a, b]$ by the interval (a, b)? If so, provide proof or an explanation. Otherwise, explain and provide a counterexample.*

These two items were selected since they specifically target conceptual understanding with minimal procedural components. Unlike the other DCST items which generally combine conceptual and procedural elements (e.g., problems on related rates of change and optimization), these two items focus almost exclusively on students' ability to relate mathematical concepts. In particular, Item 5 requires an understanding of the relationship between a function and its linear approximation, and Item 7 requires an understanding of why a closed interval is necessary for the extreme value theorem.

The DCST was administered in person in two instances and served as two summative assessments of the course. For each instance, students were given 80 minutes to answer five items, which vary in difficulty and focus. The time constraints likely affected some aspects of student performance, especially for Item 5 which appeared at the end of the first testing session. This time pressure can affect the completeness of responses and the quality of explanations, which may result in some solutions being classified at lower SOLO levels than might have been achieved under ideal problem-solving conditions (Biggs & Collis, 2014). However, this time-constrained situation reflects actual assessment conditions in higher education.

Data Analysis

The study employed a primary evaluator with a secondary validation approach for methodological consistency. One researcher classified all 57 solutions for both items, while two additional raters (one co-researcher and one mathematics faculty member) each evaluated three solutions per item, with no overlap between their assigned solutions. To minimize potential bias, the primary evaluator used a structured protocol that requires the explicit identification of demonstrated conceptual knowledge components in a solution before classifying it to a SOLO level. Moreover, the interpreters were trained with the rubric using a detailed document containing sample solutions not included in the study, and interrater samples were strategically selected to represent different types of solutions. For each item, reliability in classifying the solutions to SOLO levels was measured by calculating weighted Cohen's kappa for the six total solutions given to the interpreters. Although there were two interpreters, each of whom evaluated three solutions, their consolidated ratings were treated as coming from a single interrater in the calculation of Cohen's kappa. Qualitative interpretation of the resulting value of Cohen's kappa followed the nomenclature put forth by Landis and Koch (1977, p. 165).

For each item, student responses were analyzed using frequency counts of the demonstrated SOLO levels. In addition, using the qualitative feedback from the evaluation procedure, student responses were analyzed qualitatively to identify patterns in demonstrated conceptual understanding, including common conceptual misconceptions or interesting solutions. To illustrate these findings, representative examples of student responses at selected SOLO levels were identified. These examples highlight the characteristics associated with each SOLO level. This mixed-methods approach provides a detailed view of student understanding by capturing both big-picture patterns and more nuanced insights into the solutions themselves.

RESULTS AND DISCUSSION

Distribution of SOLO Levels

Figure 1 illustrates the distribution of SOLO levels for each exam item. For Item 5, 63% of the responses were classified at the multistructural level or lower, indicating that students were able to identify and demonstrate the conceptual knowledge components to answer the item without integration. Although one-third of the responses were classified at the transitional MR level, the degree of integration varied. In contrast, responses for Item 7 showed a different distribution of SOLO levels, with 86% achieving the transitional MR or the relational levels. Similar to Item 5, responses at the transitional level for Item 7 also varied in the degree of integration of the knowledge components.

For Item 7, 74% of the responses were able to arrive at the correct answer for item 7 (i.e., that the extreme value theorem does not hold in an open interval) than for item 5 (i.e., that the approximation is valid for values of x near zero). This difference in performance may be attributed to several factors beyond the effects of learning materials. Item 5 required students to explore the linear approximation of $\sin x$, an open-ended task which they have not encountered in the class materials. In contrast, the explanations needed to answer Item 7 have been briefly discussed in class. The higher number of MR-level responses for Item 7 suggests that familiarity with the topic allowed students to integrate multiple conceptual knowledge components, albeit in different degrees. Moreover, the open-ended nature of Item 5 required constructive reasoning, which may have contributed to lower success rates. In contrast, evaluating the validity of the theorem in Item 7 may have been facilitated by identifying a counterexample, which may have contributed to higher success rates. Research suggests that undergraduate students usually find constructive proving more difficult that finding counterexamples to a statement (Bedros, 2003; Pasani, et al., 2021), which may further explain the difference in performance between the two Items.

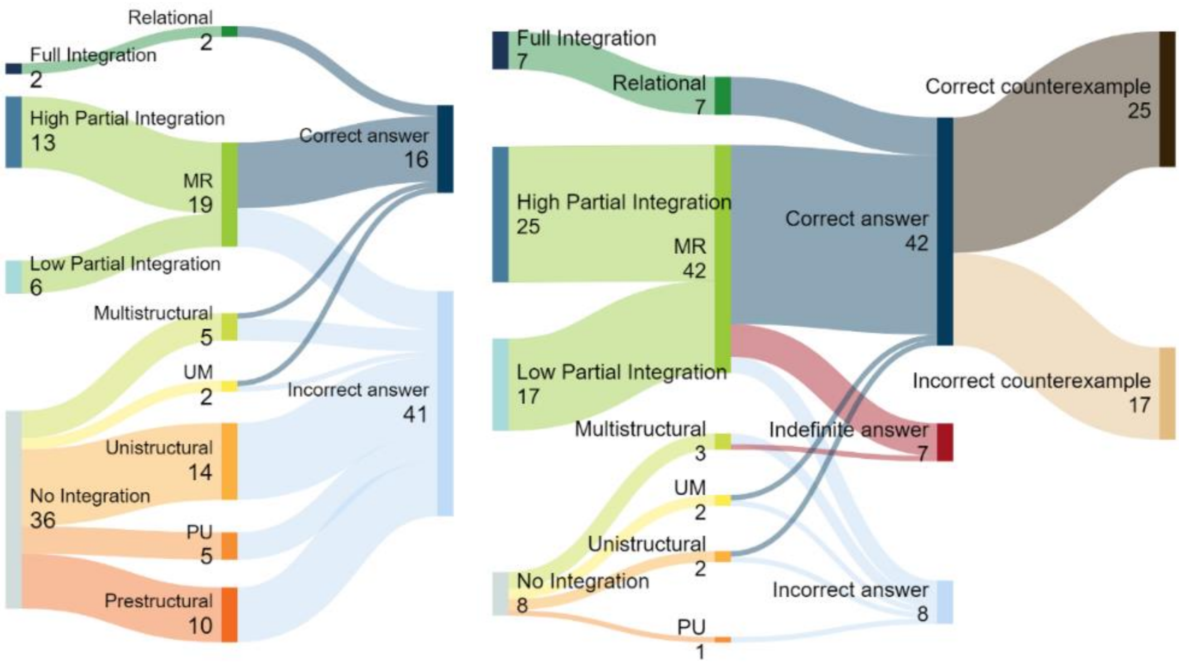


Figure 1. Distribution of evaluations for Item 5 (Left) and distribution of evaluations for Item 7 (Right)

For both items, there is a significant relationship between a solution's SOLO level and whether or not it arrives at the correct answer: for item 5, $\chi^2(6, N = 57) = 28.7, p < .001$, and a value of 0.709 for Cramer's V; and for item 7, $\chi^2(10, N = 57) = 15.4, p = .008$, and a value of 0.459 for Cramer's V. For Item 5, this indicates a strong relationship between a solution's SOLO level and correctness of answer. For Item 7, while the relationship is significant, the moderate effect size measured by Cramer's V suggests that there may be other factors beyond conceptual understanding—such as familiarity with the particular topic or with constructing counterexamples—that affected the correctness of a solution's answer.

Among the transitional MR responses to Item 5, there is a significant relationship between the degree of integration of demonstrated knowledge components in a solution and whether it arrives at the correct answer: $\chi^2(1, N = 19) = 8.15, p = .004$, and a value of 0.655 for Cramer's V. This means that among the transitional MR responses to Item 5, a solution which demonstrates a high degree of integration of demonstrated knowledge components strongly suggests that it arrives at the correct answer. Interestingly, for Item 7, among the transitional MR responses, there is no significant relationship between the degree of integration of demonstrated knowledge components in a solution and whether or not it arrives at the correct answer): $\chi^2(2, N = 42) = 4.77, p = .092$, and a value of 0.337 for Cramer's V. While there is still a relationship between these two variables among transitional MR responses to Item 7, it is not as strong as that exhibited among transitional MR responses to Item 5.

However, among the 33 responses in the transitional MR level and which arrived at the correct answer for Item 7, there is a significant relationship between the degree of integration of demonstrated knowledge components in a solution and whether or not it is able to provide a correct counterexample: $\chi^2(1, N = 33) = 16.2, p < .001$, and a value of 0.702 for Cramer's V. This suggests that for the transitional MR responses to Item 7 which arrived at the correct answer, being able to integrate knowledge components strongly suggests the ability to provide a correct counterexample.

The dominance of lower SOLO level responses and incorrect answers for Item 5 may also be attributed to the manner in which the DCST was administered. The DCST was administered as two summative assessments of the course: Items 1 to 5 as Quiz 3, and Items 6 to 10 as Quiz 4, each timed at 75 minutes. It is possible that most students ran out of time in answering Item 5 while taking Quiz 3, and so they were not able to explore enough or explain their answers enough to warrant a higher SOLO level or arrive at the correct answer.

While these trends reveal that student solutions to Items 5 and 7 have different distributions, the most common SOLO level across both items was the transitional MR level, which suggests students are able to demonstrate, whether explicitly or implicitly, the relevant conceptual knowledge components and integrate these, albeit with varying depth, to form a conclusion. Very few responses reached the relational level, suggesting that only a few students are able to demonstrate and fully integrate the relevant conceptual knowledge components, as well as to come up with correct answers.

Interrater reliability for SOLO level classification was measured at a weighted Cohen's kappa of 0.659 for Item 5 and 0.667 for Item 7, both considered as substantial agreement according to Landis and Koch (1977). This agreement shows the effectiveness of the rubric. These values for the weighted Cohen's kappa indicate that despite its effectiveness, there may have been potential biases in the evaluation process. To mitigate this, evaluators cited particular details in the solutions to justify their SOLO classifications. Raters encountered some difficulties with solutions that are between adjacent levels, since interpreting these solutions entail more nuance. For example, determining whether a solution is in the prestructural or the PU level requires nuanced judgment as it requires an evaluator to determine whether a solution genuinely attempts to engage with a conceptual knowledge component. This level of interrater reliability suggests that additional refinement, perhaps more detailed and clearer delineations between levels, could enhance interrater reliability by minimizing subjective interpretation in boundary cases. Implementing a collective moderation process where solutions that are difficult to classify are discussed by multiple raters collectively may also address boundary cases. Nevertheless, the substantial overall agreement suggests that the evaluation rubric and procedure is able to objectively classify student solutions while being sensitive to small nuances in students' demonstrated conceptual understanding.

Insights into Students' Conceptual Understanding

Analysis of solutions to Items 5 and 7 revealed significant insights into students' conceptual understanding. For Item 5, which examines students' comprehension of the approximation $\sin x \approx x$ in the context of linear approximations, responses reflected varied levels of understanding. A pattern emerged where many solutions were classified into either the unistructural level or the transitional MR level of understanding.

Student solutions to Item 5 were generally marked by the ability to recognize the need to construct the linear approximation of $\sin x$. However, unistructural-level solutions were either not able to go beyond this step, used other concepts that were not relevant to the problem, or exhibited an incorrect understanding of concepts relevant to the problem. Figure 2 shows two solutions with this characteristic. Note that knowing that the derivative of $\sin x$ is not a conceptual knowledge component; rather, it is a procedural knowledge component.

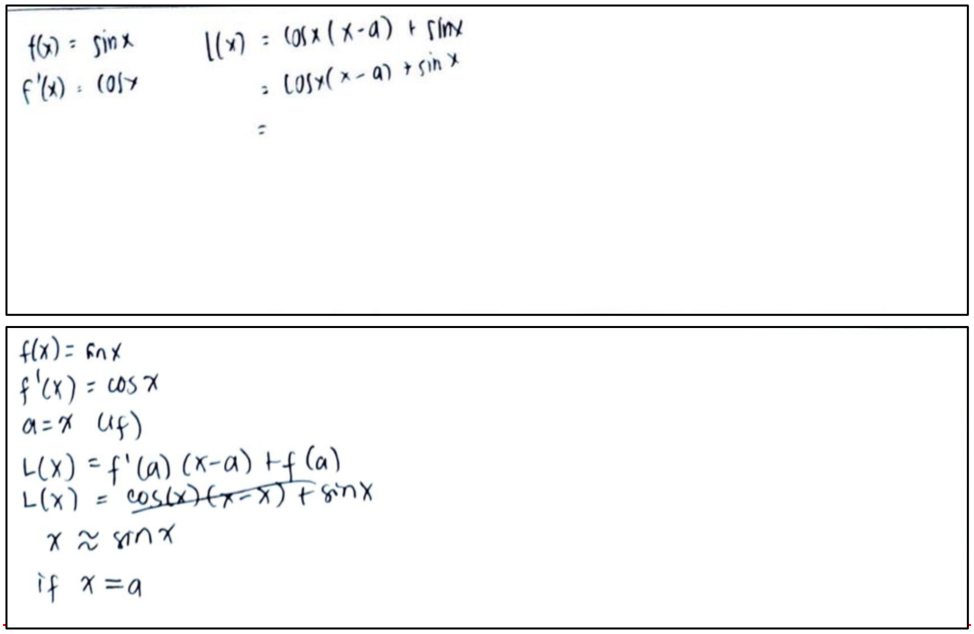


Figure 2. Two unistructural responses to Item 5

Another interesting insight comes from some multistructural-level and most MR-level solutions to Item 5. These solutions were able to construct the linear approximation of $\sin x$ at a number a, that is, $L(x) = f'(a)(x-a) + f(a) = \cos x\,(x-a) + \sin a$. However, almost all these solutions substituted $a = 0$ without any justification. Naturally this arrives at $L(x) = x$, prompting the students to conclude that the approximation is valid for values of x near zero. This error of substituting $a = 0$ is not limited to MR-level solutions; some multistructural-level solutions also exhibit this error. Figure 3 shows two solutions with this error.

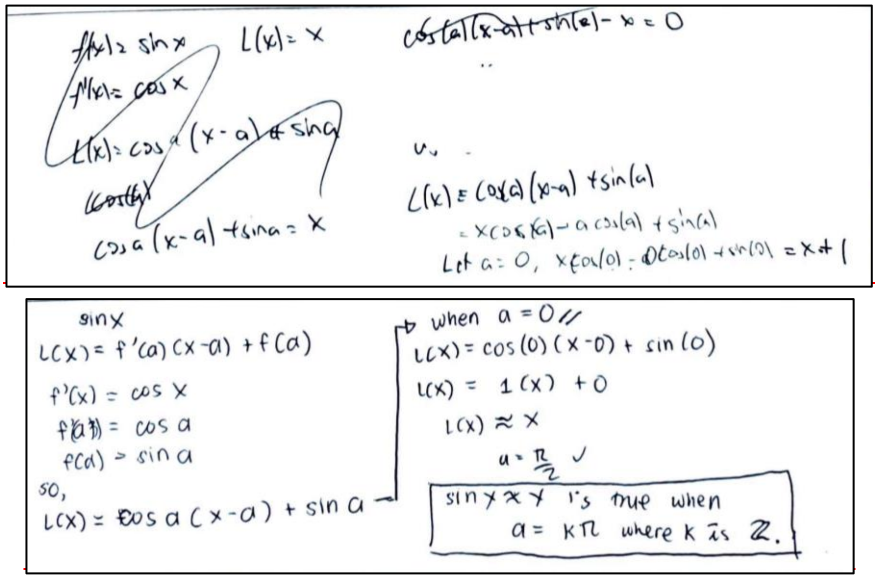


Figure 3. Two MR-level solutions with low partial integration (top) and high partial integration (bottom) to Item 5

These errors indicate a lack of sufficient depth in conceptual understanding and suggest that students may be applying procedural skills without fully understanding why the said procedures work. This tendency to rely on symbolic manipulation alone, or "proceduralizing," may be because students are used to using mathematical procedures to answer mathematics tests—that is, typical questions require students to “solve, sketch, find, graph, evaluate, determine, and calculate in a straightforward fashion” (Ferrini-Mundy & Graham, 1991)—or because of a personal belief that this is how to do mathematics “properly” (Engelbrecht et al., 2009).

Interestingly, there were two solutions which compared coefficients between the expressions in the two sides of the equality $L(x) = x$, leading to the correct answer. This technique of comparing coefficients was not introduced in the course, and thus its presence in some student solutions was a pleasant surprise. Assuming that these students have not encountered this technique in the past, their ability to come up with this line of reasoning is a demonstration of their ability to connect concepts and “think outside the box,” an essential skill for mathematical proficiency. Figure 4 shows these two solutions which used the technique of comparing coefficients.

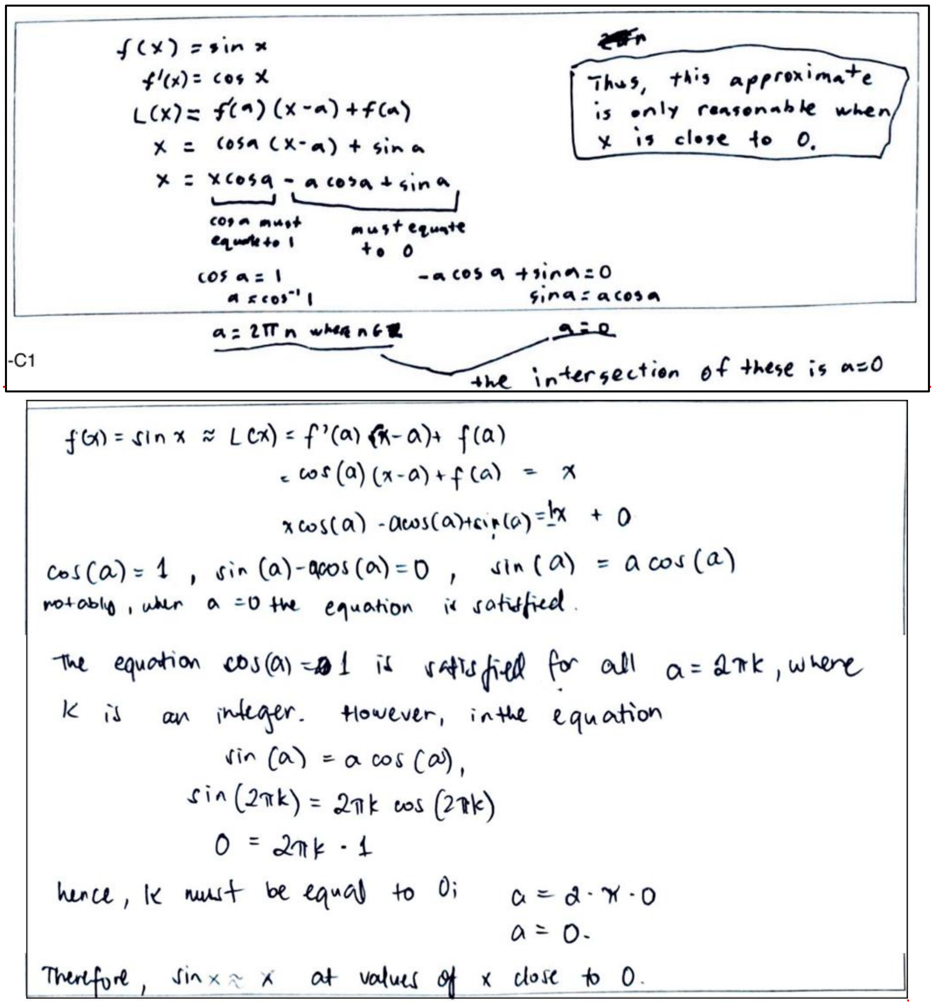


Figure 4. An MR-level solution (top) and a relational-level solution (bottom) to Item 5

For Item 7, which evaluates understanding of the generalizability of the extreme value theorem to open intervals, solutions were more homogeneous, with almost three-quarters of the solutions classified in the transitional MR level. Among all the items in the DCST, Item 7 has the greatest number of solutions in the transitional MR level as well as the greatest number of solutions that arrived at the correct answer.

Most responses at this level correctly concluded that the extreme value theorem will not hold in general for open intervals, but they are also characterized by an incorrect counterexample or an inability to fully articulate the proper mathematical reasoning. In general, MR-level responses argue that in a closed interval, an absolute extremum of a function may occur at an endpoint of the interval, and so excluding the endpoints will no longer guarantee that the function has both absolute extrema on that open interval. The level of articulation and depth, of course, varies among the MR-level solutions. However, this line of reasoning does not fully explain why the function will no longer have an absolute maximum or an absolute minimum on an open interval; this is the most common gap in explanation present in MR-level responses. It is thus apt to look at some MR-level solutions with different degrees of integration among demonstrated knowledge components.

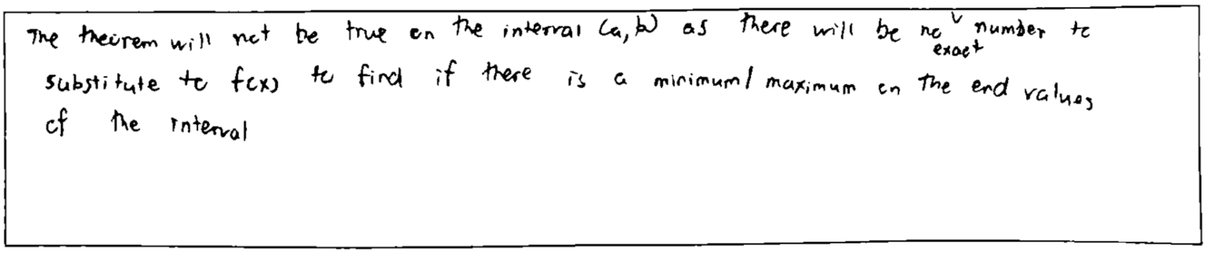

The theorem will not be true on the interval (a,b) as there will be no exact number to substitute to f(x) to find if there is a minimum/maximum on the end values of the interval

Figure 5. A transitional MR response to Item 7 with low partial integration

The MR-level solution in Figure 5 demonstrates and integrates several conceptual knowledge components: that the theorem will not be true on an open interval because of an implied demonstration of the difference between open and closed intervals. However, this is classified as having a low partial integration because the reasoning lacks details and depth: it is not clear why "there will be no exact number to substitute to $f(x)$ to find if there is a minimum/maximum on the end values of the interval." The solution also fails to provide a counterexample to show that the extreme value theorem is false on an open interval.

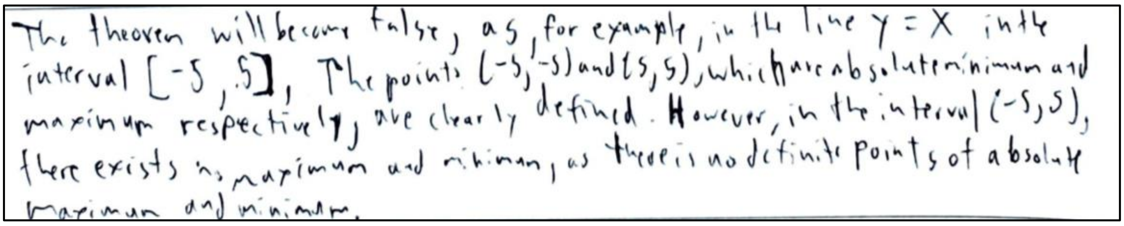

The theorem will become false, as, for example, in the line y = X in the interval [-5, 5], The points (-5,-5) and (5,5), which are absolute minimum and maximum respectively, are clearly defined. However, in the interval (-5,5), there exists no maximum and minimum, as there is no definite points of absolute maximum and minimum.

Figure 6. A transitional MR response to Item 7 with high partial integration

The MR-level solution in Figure 6 also demonstrates not only the conceptual knowledge components as the solution in Figure 5 but also an understanding of what a counterexample is, and so it is able to provide a correct counterexample. The construction of the explanation also aids the reader to deduce, albeit with some effort, what the clause "in the interval $(-5,5)$, there exists no maximum and minimum, as there is [sic] no definite points of absolute maximum and minimum" means.

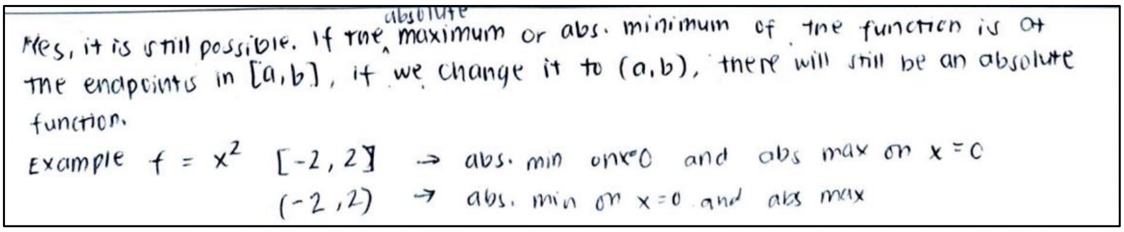
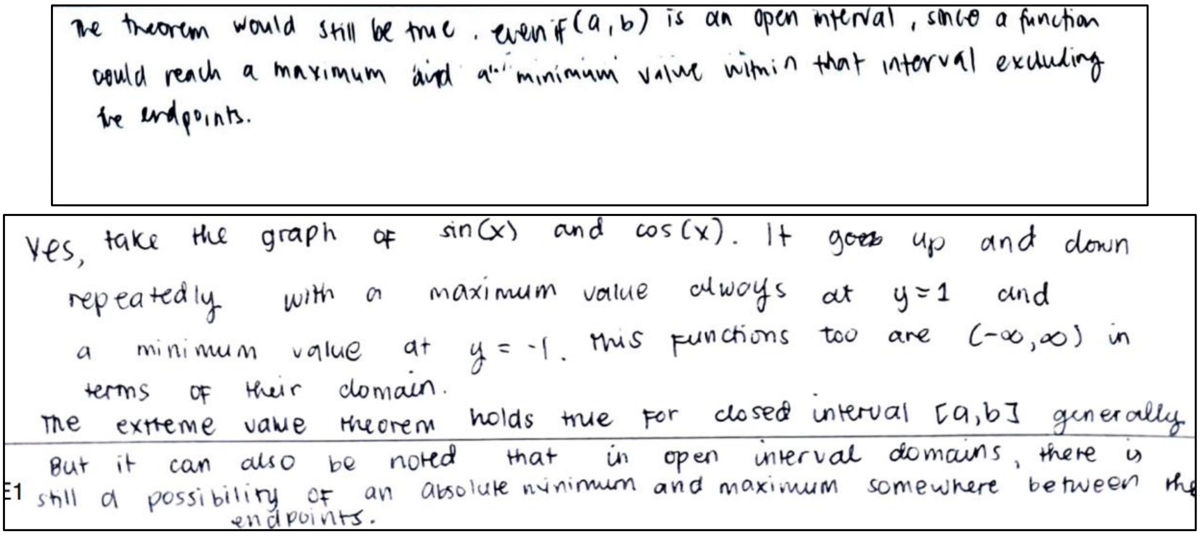

Yes, it is still possible. If the absolute maximum or abs. minimum of the function is at the endpoints in [a,b], if we change it to (a,b), there will still be an absolute function.
Example f = x² [-2,2] → abs. min on x=0 and abs max on x=c
(-2,2) → abs. min on x=0 and abs max

Figure 7. A transitional MR response to Item 7 with high partial integration

One notable misconception in the solutions stems from an incorrect understanding of mathematical logic. A number of solutions interpreted the negation of the conclusion of the extreme value theorem (i.e., the function has both an absolute maximum and an absolute maximum on the interval) as the function not having both an absolute maximum and absolute minimum. This misconception leads to the following line of reasoning: the extreme value theorem is not true on an open interval because it is possible for a function to still have at least one absolute extremum on an open interval. This is usually followed by an erroneous counterexample of $y = x^2$ on an open interval containing zero. Figure 7 shows one such solution.


The theorem would still be true. even if (a, b) is an open interval, since a function could reach a maximum and a minimum value within that interval excluding the endpoints.

Yes, take the graph of sin(x) and cos(x). It goes up and down repeatedly with a maximum value always at y = 1 and a minimum value at y = -1. This functions too are (-∞, ∞) in terms of their domain.

The extreme value theorem holds true for closed interval [a, b] generally. But it can also be noted that in open interval domains, there is still a possibility of an absolute minimum and maximum somewhere between the endpoints.

Figure 8. Responses to Item 7 which show a logical misconception

Another notable misconception is related to determining the truth value of the statement: "*A continuous function on an open interval attains an absolute maximum and an absolute minimum on that interval.*" Figure 8 shows solutions exhibiting this misconception. In particular, these solutions exhibit a belief that particular examples are sufficient to deduce the truth value of a universally quantified statement.

This misconception concerning mathematical logic seems to arise from several factors. Students often have difficulties in understanding universally quantified statements, especially the idea that a single counterexample invalidates a universally quantified statement. A study by Knuth et al. (2019) shows that students typically resort to reasoning through examples instead of understanding the logical structure of mathematical statements. This tendency to believe that examples constitute proof may stem from students' experience with inductive reasoning in other contexts. This inappropriate transfer happens even though

> *… from a logical point of view, to understand that in order for a universal statement to be true it must hold for all elements in the statement's domain and that a single counterexample is sufficient for refuting a false statement, implies that confirming examples are insufficient for proving and that a general justification is needed* (Buchbinder & Zaslavsky, 2013, p. 135).

This logical misconception may also be tied to confirmation bias and a thinking that the statement is true unless proven otherwise. In fact, Alacaci and Pastor (2005) found that such logical errors involving incorrectly negating quantified statements and statements involving logical connectives are not limited to students but appear across various populations.

Across both items, these patterns reveal three characteristics of student solutions. The first and most common is the tendency to resort to mathematical expressions alone, with no or insufficient verbal explanations to relate consecutive mathematical expressions. This harkens back to students' tendency to proceduralize answers and solutions. This is troublesome because Habre (2002) posits the rule of four—that an effective calculus learner is capable of "communicating ideas through algebraic, graphical, numerical, and verbal means." The second is the tendency to miss small gaps in reasoning. The third is the presence of misconceptions in mathematical logic.

Utility and Challenges of the SOLO-Based Rubric

The SOLO-based rubric demonstrated several strengths in evaluating students' responses based on their conceptual understanding of the applications of derivatives. The rubric's structure, especially with its inclusion of transitional levels, facilitated differentiation between levels of understanding. This feature allowed evaluators to identify and categorize responses that showed either partial comprehension or emerging relationships between conceptual knowledge components. The rubric's inclusion of how to handle special cases also proved to be beneficial in evaluating partial solutions, solutions with early errors, solutions with unconventional approaches, and solutions with no or insufficient verbal explanations. These features of the rubric enabled a more nuanced evaluation of student solutions.

One particularly effective aspect of the rubric was its ability to capture different levels of integration among demonstrated conceptual knowledge components. In cases where solutions were classified in the transitional MR level, which indicates partial integration of knowledge components, the rubric provided a framework to further classify partial integration as either low partial integration or high partial integration. This allowed it to be extremely beneficial since for both items, the transitional MR level had the highest number of solutions. This ability to differentiate the degree of integration among MR-level solutions made the rubric especially useful for identifying varying abilities in relating conceptual knowledge components that might otherwise have gone unrecognized. This echoes the work of Burnett (1999), who also added sublevels to the SOLO Taxonomy and used it to assess learning outcomes from counselling, and Chan et al. (2002), who did the same in the context of an advanced practice subject in mental health. However, both studies did not use the transitional levels originally proposed by Biggs and Collis (2014), and it is unclear why these studies did not do so (Stålne et al., 2016).

Despite these strengths, the application of the rubric also presented certain challenges. A primary difficulty arose in cases where student responses did not clearly align with characteristics of a single SOLO level. For example, some responses exhibited full integration of all necessary knowledge components for an item, but not explicitly. This is most apparent in Item 5. Some solutions implied a full understanding of how to reason and solve the problem but lacked sufficient verbal explanations. Such solutions made it challenging to assign a definitive level, despite the rubric having a note regarding solutions without verbal explanations. This highlights an area where future refinements to the rubric could add clarity.

In addition, while the rubric aimed for consistency, subjectivity in interpretation affected rater agreement, especially when categorizing responses that exhibited ambiguous reasoning or lack of clarity. To address these challenges, it may be valuable in future iterations to provide additional guidance or to refine certain criteria, which would help clarify the handling of special cases. These adjustments would likely improve interrater agreement and reliability.

Implications for Educational Practice

Addressing Gaps and Misconceptions

Insights from the patterns and observations provide important implications for calculus education, specifically to address challenges related to students' conceptual understanding. Although these findings primarily concern undergraduate calculus education, they also suggest broader considerations for higher education mathematics instruction. One key trend observed among the solutions was students' tendency to omit verbal explanations to connect mathematical expressions in their solutions, which hinders a full demonstration and articulation of relationships between mathematical concepts and procedures.

Addressing this issue requires targeted strategies that emphasize the importance of explicitly articulating these relationships (Kågesten & Engelbrecht, 2006; Norqvist, 2018). Encouraging students to verbalize reasoning during problem-solving can help solidify their understanding and better illustrate the connections between concepts and procedures. To achieve this, instructors can integrate techniques such as think aloud exercises, where students are prompted to verbalize their reasoning and connections between ideas as they work through problems (Hicks & Bostic, 2021; Kani & Shahrill, 2015), and guided questioning, which can prompt students to reflect on the underlying reasoning behind each step (Franke et al., 2009). These methods not only help clarify students' thinking but also guide them to develop a deeper understanding of mathematical concepts.

While these strategies provide approaches to develop conceptual understanding, their implementation may present challenges. In content-heavy courses such as calculus, time constraints may hinder the full use of these strategies and create a tension between covering course material and developing conceptual understanding. Effective implementation of these strategies also requires specific pedagogical expertise that not all instructors may possess think aloud activities need skillful facilitation to elicit students' meaningful verbalization without leading their thinking, and effective questioning needs careful planning and adaptation to student responses. Lastly, these approaches may become impractical when implemented in large lecture settings in undergraduate classes. Despite these, instructors may adopt a gradual implementation of these activities. Integrating brief think aloud activities in class, requiring conceptual explanations in course requirements, and using peer discussion to practice students' verbalization contribute to an emphasis on conceptual understanding alongside procedural fluency.

To address specific conceptual misconceptions, instructors could design tasks that explicitly address common errors and misconceptions. One possible approach is to present an example solution that contains a common misconception and ask students to identify whether the solution is correct or incorrect, and to explain why (Adams et al., 2013). This presents a natural opportunity for students to actively engage in the correction process, either individually or in groups, thus promoting deeper learning and understanding. Through these methods, students can better articulate their reasoning, which, in turn, reinforces their conceptual understanding. Encouraging verbal explanations and guided reflection on problem-solving processes will help bridge the gap between procedural fluency and conceptual depth, fostering a more integrated understanding of calculus.

Curriculum Integration

On a broader level, curriculum design in applied calculus courses should not neglect the importance of conceptual understanding in students' mathematical proficiency. Integrating deeper explanations of foundational concepts in calculus, such as the definitions of the derivative, the definite integral, and the fundamental theorem of calculus, can positively impact not only students' procedural skills but also their ability to more effectively relate mathematical concepts. Such adjustments require systemic effort, but these are essential to developing students' mathematical proficiency. As evidenced by the findings of this paper, teaching procedural skills without fostering an understanding of the underlying concepts may lead to shallow learning, where students can execute procedures but fail to fully grasp why they are using them or how they connect to broader mathematical ideas. Therefore, the findings highlight the importance of balancing conceptual and procedural instruction in applied calculus education. By adopting strategies that address observed gaps and challenges, instructors can support students in developing a more integrated mathematical understanding that allows them to apply both conceptual insights and procedural knowledge effectively.

Broader Applications of the Rubric

Although the method and rubric presented in this study were developed within the context of a first-year undergraduate applied differential calculus course, they possess the flexibility to be adapted for use in a range of mathematical domains. By employing this rubric, instructors can obtain deeper insights into the levels of conceptual understanding exhibited by their students, thereby informing instructional strategies and supporting the achievement of desired learning outcomes.

The structure of the SOLO Taxonomy-based rubric, with its emphasis on conceptual understanding, renders it adaptable to various applied calculus topics beyond the initial focus on linear approximations and the extreme value theorem. Topics such as limits, the definition of the derivative, definite and indefinite integrals, and the Fundamental Theorem of Calculus may also benefit from this rubric-based approach. Through its application, instructors can assess students' conceptual understanding in ways that extend beyond procedural competence. However, effective implementation requires careful design of assessment tasks. Items must be constructed to probe the depth and structure of students' understanding, rather than merely procedural accuracy, and the rubric should be employed formatively throughout the course to monitor and support the development of conceptual understanding over time.

Moreover, the rubric has potential applicability beyond calculus. It can be adapted to support the evaluation of conceptual understanding in other areas of undergraduate mathematics, including abstract algebra and linear algebra. In the context of abstract algebra, the rubric could be modified to include conceptual knowledge components related to algebraic structures and may be useful for analyzing students' understanding of properties of groups, rings, homomorphisms, and isomorphisms. In linear algebra, the rubric could be tailored to assess how students integrate and coordinate algebraic, geometric, and abstract representations of vector spaces and linear transformations—an essential dimension of conceptual understanding in this domain (Hillel, 2000).

Adapting the SOLO-based rubric to these broader domains, however, may present certain challenges. First, the development of conceptual understanding in each mathematical field may follow a progression that does not align neatly with the SOLO Taxonomy levels. Second, different content areas emphasize distinct modes of mathematical thinking, which may necessitate substantial revisions to the SOLO level descriptors. Third, the identification of relevant conceptual knowledge components for specific problems requires domain-specific expertise to ensure alignment with disciplinary expectations.

In addition to its role in assessment, the SOLO-based rubric holds promise as a tool for course and curriculum design in higher mathematics education. When integrated into course-level assessments, it enables instructors to identify patterns of conceptual understanding and common misconceptions, thereby facilitating targeted instructional interventions. This stands in contrast to conventional mathematics assessments, which often focus on routine exercises and procedural proficiency. The rubric thus provides a structured and theoretically grounded framework for enhancing students' conceptual understanding not only within calculus but also across a broad spectrum of mathematical disciplines.

CONCLUSION

This study aimed to develop and implement a rubric based on the SOLO Taxonomy to evaluate students' conceptual understanding as demonstrated in their written responses to applied calculus test items. The proposed rubric provides a novel assessment tool that enables instructors to evaluate not only the capacity of students' mathematical thinking but also the complexity and depth of their reasoning.

The findings indicate that the rubric is effective in capturing varied levels of conceptual understanding and offers a nuanced framework for identifying patterns in students' reasoning, including conceptual strengths, misconceptions, and incomplete understandings. Notably, the inclusion of transitional levels and explicit criteria for evaluating unconventional or incomplete solutions underscores the rubric's capacity to assess demonstrated understanding in a rigorous and flexible manner. Beyond its role as an assessment instrument, the rubric shows potential to enhance student learning outcomes when integrated into instructional practices. By offering structured feedback based on SOLO levels, rather than relying solely on procedural accuracy or correctness of answers, the rubric can guide students in identifying specific conceptual gaps. This process fosters the development of metacognitive awareness and supports students' self-regulation in learning mathematics.

The contributions of this work extend beyond applied calculus education. By adapting the SOLO Taxonomy to assess conceptual understanding through demonstrated levels of thinking and cognitive complexity, this study offers an assessment framework that emphasizes the integration and depth of mathematical knowledge—core elements of conceptual understanding. The rubric can be readily embedded in mathematics assessments to support the diagnostic identification of conceptual gaps and to inform targeted instructional interventions. For instance, if an instructor identifies that a majority of students demonstrate responses at the multistructural level, instructional activities can be designed to promote the integration of isolated ideas and facilitate progress toward relational understanding. The adaptability of the rubric to other mathematical domains further supports a broader shift toward concept-focused assessment and instruction in undergraduate mathematics education. Elements of the rubric may be modified to align with domain-specific knowledge structures, with expert consultation used to define the key conceptual components relevant to particular content areas. Indicators for each SOLO level (e.g., as presented in Table 1) can be refined to reflect the characteristics of conceptual understanding in distinct domains. Provisions for evaluating unconventional responses can also be expanded, particularly in proof-based or open-ended tasks, where diverse reasoning paths may yield valid mathematical conclusions.

Finally, future research may focus on further testing and refining the rubric, particularly in addressing challenges such as borderline cases where student responses lie between two SOLO levels. Empirical validation across a range of mathematical contexts—such as linear algebra, abstract algebra, or statistics—is necessary to examine the generalizability and robustness of the rubric. Additionally, studies may explore the impact of using the rubric on students' conceptual development and the extent to which it supports the alignment between conceptual understanding and procedural fluency in mathematics learning.

Acknowledgments

The authors would like to express gratitude to the study participants, the validator of the test items' conceptual knowledge components, and the interpreters for their time and effort to be part of this research endeavor.

Declarations

Author Contribution	:	PJMF: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – Original Draft, Visualization, Supervision, Project administration, Funding acquisition. AFHG: Conceptualization, Methodology, Validation, Resources, Writing – Review & Editing, Supervision, Funding acquisition.
Funding Statement	:	This paper was made possible by the Department of Education – Science Education Institute (DOST-SEI).
Conflict of Interest	:	The authors declare no conflict of interest.
Additional Information	:	Additional information is available for this paper with the corresponding author.

REFERENCES

Adams, D. M., McLaren, B. M., Mayer, R. E., Goguadze, G., & Isotani, S. (2013). Erroneous examples as desirable difficulty. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 803–806). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_117

Alacaci, C., & Pasztor, A. (2005). On people's incorrect either-or patterns in negating quantified statements: A study. *Proceedings of the Annual Meeting of the Cognitive Science Society, 27*(27), 1714-1719. https://escholarship.org/uc/item/5fz0p6m1

Arslan, S. (2010). Traditional instruction of differential equations and conceptual learning. *Teaching Mathematics and Its Applications, 29*(2), 94–107. https://doi.org/10.1093/teamat/hrq001

Bedros, V. (2003). An exploratory study of undergraduate students' perceptions and understandings of indirect proofs. *Doctoral Dissertation*. The University of Montana. https://scholarworks.umt.edu/etd/9473

Biggs, J., & Collis, K. (2014). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the observed learning outcome)*. Academic Press.

Bisson, M.-J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 141–164. https://doi.org/10.1007/s40753-016-0024-3

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1964). *Taxonomy of educational objectives (Vol. 2)*. Longmans, Green New York.

Buchbinder, O., & Zaslavsky, O. (2013). Inconsistencies in students' understanding of proof and refutation of mathematical statements. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the*

37th Conference of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 129–136). 37th Annual Conference of the International Group for the Psychology of Mathematics Education, Kiel, Germany.

Büchele, S., & Feudel, F. (2023). Changes in students' mathematical competencies at the beginning of higher education within the last decade at a German university. *International Journal of Science and Mathematics Education, 21*(8), 2325–2347. https://doi.org/10.1007/s10763-022-10350-x

Burnett, P. C. (1999). Assessing the structure of learning outcomes from counselling using the SOLO taxonomy: An exploratory study. *British Journal of Guidance & Counselling, 27*(4), 567–580. https://doi.org/10.1080/03069889908256291

Chan, C. C., Tsui, M. S., Chan, M. Y. C., & Hong, J. H. (2002). Applying the Structure of the Observed Learning Outcomes (SOLO) Taxonomy on student's learning outcomes: An empirical study. *Assessment & Evaluation in Higher Education, 27*(6), 511–527. https://doi.org/10.1080/0260293022000020282

Claudia, L. F., Kusmayadi, T. A., & Fitriana, L. (2020). The SOLO taxonomy: Classify students' responses in solving linear program problems. *Journal of Physics: Conference Series, 1538*(1), 012107. https://doi.org/10.1088/1742-6596/1538/1/012107

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*(4), 344–377. https://doi.org/10.1016/j.dr.2014.10.001

Engelbrecht, J., Bergsten, C., & Kågesten, O. (2009). Undergraduate students' preference for procedural to conceptual solutions to mathematical problems. *International Journal of Mathematical Education in Science and Technology, 40*(7), 927–940. https://doi.org/10.1080/00207390903200968

Ferrini-Mundy, J., & Graham, K. G. (1991). An overview of the calculus curriculum reform effort: Issues for learning, teaching, and curriculum development. *The American Mathematical Monthly, 98*(7), 627–635. https://doi.org/10.1080/00029890.1991.11995769

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education, 60*(4), 380–392. https://doi.org/10.1177/0022487109339906

Habre, S. (2002). Writing in a reformed differential equations class. *International Conference on the Teaching of Mathematics*.

Hicks, T., & Bostic, J. D. (2021). Formative assessment through think alouds. *Mathematics Teacher: Learning and Teaching PK-12, 114*(8), 598–606. https://doi.org/10.5951/MTLT.2020.0245

Hillel, J. (2000). Modes of description and the problem of representation in linear algebra. In J.-L. Dorier (Ed.), *On the Teaching of Linear Algebra* (pp. 191–207). Springer Netherlands. https://doi.org/10.1007/0-306-47224-4_7

Kågesten, O., & Engelbrecht, J. (2006). Supplementary explanations in undergraduate mathematics assessment: A forced formative writing activity. *European Journal of Engineering Education, 31*(6), 705–715. https://doi.org/10.1080/03043790600911803

Kani, N. H. A., & Shahrill, M. (2015). Applying the thinking aloud pair problem solving strategy in mathematics lessons. *Asian Journal of Management Sciences and Education, 4*(2), 20–28. http://www.ajmse.leena-luna.co.jp/AJMSEPDFs/Vol.4(2)/AJMSE2015(4.2-03).pdf

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). The strands of mathematical proficiency. In N. R. Council, *Adding It Up: Helping Children Learn Mathematics*. The National Academies Press. https://doi.org/10.17226/9822

Knuth, E., Zaslavsky, O., & Ellis, A. (2019). The role and use of examples in learning to prove. *The Journal of Mathematical Behavior, 53*, 256–262. https://doi.org/10.1016/j.jmathb.2017.06.002

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Mukuka, A., Balimuttajjo, S., & Mutarutinya, V. (2020). Applying the SOLO taxonomy in assessing and fostering students' mathematical problem-solving abilities. *Problems of Education in the 21st Century, 77*(6), 13.

Mulbar, U., Rahman, A., & Ahmar, A. (2017). Analysis of the ability in mathematical problem-solving based on SOLO taxonomy and cognitive style. *World Transactions on Engineering and Technology Education, 15*(1), 68-73. http://wiete.com.au/journals/WTE&TE/Pages/Vol.15,%20No.1%20(2017)/12-Ahmar-AS.pdf

Norqvist, M. (2018). The effect of explanations on mathematical reasoning tasks. *International Journal of Mathematical Education in Science and Technology, 49*(1), 15–30. https://doi.org/10.1080/0020739X.2017.1340679

Pasani, C. F., Kusumawati, E., & Suryaningsih, Y. (2021). The ability of mathematics education students to build counterexamples in solving cyclic group problems. *AIP Conference Proceedings, 2330*(1), 040023. https://doi.org/10.1063/5.0043223

Putri, U. H., Mardiyana, M., & Saputro, D. R. S. (2017). How to analyze the students' thinking levels based on SOLO taxonomy? *Journal of Physics: Conference Series, 895*(1), 012031. https://doi.org/10.1088/1742-6596/895/1/012031

Radmehr, F., & Drake, M. (2019). Revised Bloom's taxonomy and major theories and frameworks that influence the teaching, learning, and assessment of mathematics: A comparison. *International Journal of Mathematical Education in Science and Technology, 50*(6), 895–920. https://doi.org/10.1080/0020739X.2018.1549336

Ramos, R. A., Gutierrez, M. C., Carang, E. S. P., & Caguete, R. R. (2024). Measuring student performance in mathematics in the modern world course using bloom's and SOLO taxonomies. *International Journal of Management and Accounting, 6*(3), 78–84. https://doi.org/10.34104/ijmms.024.078084

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*(1), 175–189. https://doi.org/10.1037/0022-0663.91.1.175

Santos-Trigo, M., Camacho-Machín, M., & Barrera-Mora, F. (2024). Focusing on foundational calculus ideas to understand the derivative concept via problem-solving tasks that involve the use of a dynamic geometry system. *ZDM – Mathematics Education, 56*(6), 1287–1301. https://doi.org/10.1007/s11858-024-01607-6

Stålne, K., Kjellström, S., & Utriainen, J. (2016). Assessing complexity in learning outcomes – A comparison between the SOLO taxonomy and the model of hierarchical complexity. *Assessment & Evaluation in Higher Education, 41*(7), 1033–1048. https://doi.org/10.1080/02602938.2015.1047319

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17*(2), 213–226. https://doi.org/10.1016/S0742-051X(00)00052-4

Stovner, R. B., Klette, K., & Nortvedt, G. A. (2021). The instructional situations in which mathematics teachers provide substantive feedback. *Educational Studies in Mathematics, 108*(3), 533–551. https://doi.org/10.1007/s10649-021-10065-w

Sudihartinih, E. (2019). Facilitating mathematical understanding in three-dimensional geometry using the SOLO taxonomy. *Erudio Journal of Educational Innovation, 6*(1), 11–18. https://doi.org/10.18551/erudio.6-1.2

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*(6), 716–730. https://doi.org/10.1002/tea.20318