REPLICATING THE DERIVATIVE KNOWLEDGE MODEL: EVIDENCE FROM TAIWANESE HIGH SCHOOL STUDENTS

Pin-Chen Guo[1], Kai-Lin Yang[1], Kristin Litteck[2], Aiso Heinze[2]

[1] National Taiwan Normal University

[2] IPN – Leibniz Institute for Science and Mathematics Education Kiel

Using Sfard's (1991) description of operational (process-oriented) and structural (object-oriented) knowledge, Litteck et al. (2024) empirically validated a derivative knowledge model with German students' data. This study replicated their work with a Taiwanese sample (N = 373) to examine whether the model holds across educational systems. Operational items were found to be consistently more difficult than structural ones; however, the relative difficulty among structural sub-concepts differed from the German results. Model-comparisons showed that a three-factor model (ratio, limit, function) best represented the Taiwanese data, whereas Litteck et al. identified a two-dimensional structure (operational, structural). These results suggest that while the process-object distinction is robust, its internal organization may vary across contexts.

INTRODUCTION

Operational Versus Structural Knowledge in Three Derivative Sub-Concepts

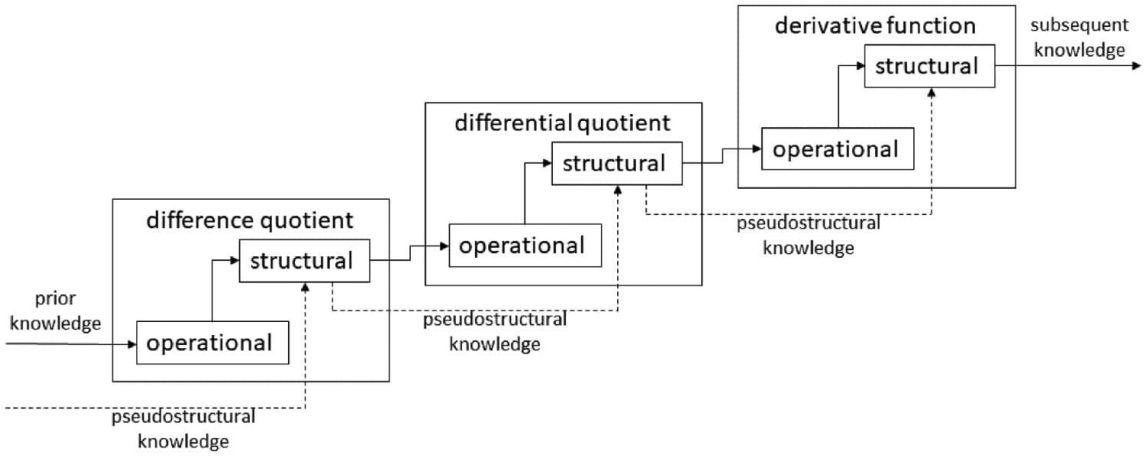
The relationship between operational (process-oriented) and structural (object-oriented) knowledge has long been a central concern in mathematics education (Dubinsky, 1991; Sfard, 1991). Despite these theoretical accounts describing the transition from operational to structural understanding, quantitative evidence documenting this progression remains scarce (Litteck et al., 2024). The derivative concept, with its three sub-concepts difference quotient, differential quotient, and derivative function, provides a fertile ground for such investigation (Zandieh, 2000). Specifically, learners are assumed to first perform a ratio process with two differences, then apply a limiting process to get the differential quotient, and finally generalize across all points to yield the derivative function. Within each sub-concept, the process is expected to become reified into an object that can be further operated upon (Sfard, 1991; Zandieh, 2000).

Accordingly, Litteck et al. (2024) proposed a 3×2 model of derivative knowledge, distinguishing operational and structural aspects within three sub-concepts (Figure 1). They empirically validated this model using data from German high school students, and the results demonstrated that students' derivative knowledge is best represented as a two-dimensional construct (operational vs. structural). Moreover, operational items were found more difficult than structural ones, suggesting the prevalence of pseudo-structural knowledge (Sfard, 1992), an apparent structural grasp without a solid process

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 539-546. PME.

foundation. This finding challenges the presumed developmental sequence from operational to structural understanding and raises a critical question: *Is this asymmetry a universal pattern or context-specific?* To test the robustness and cross-cultural transferability of this model, a systematic replication study is essential (Aguilar, 2020).

Figure 1: The Acquisition Model for the Derivative Concept (Litteck et al., 2024)



Fundamental Questions and Modes of Exploration in Replication Studies

Replication studies are essential for verifying the stability and generalizability of research findings, ensuring that observed effects are not artifacts of specific samples or contexts (Aguilar, 2020). Despite their importance, such studies remain rare in mathematics education publications, partly due to the field's emphasis on novelty and unclear boundaries between replication and extension (Burkhardt, 2013; Star, 2018). To guide future replication efforts, Aguilar (2020) proposed two fundamental questions accompanied by four modes of exploration. The first question concerns *whether the results hold true beyond the original context*. This line of inquiry can be explored through scaling up (using larger samples), scaling out (using new populations or settings), or using different research methods (e.g., new data collection techniques). The second question asks *whether there are alternative ways to study and explain an identified phenomenon or finding*, which is typically addressed by adopting a different theoretical approach from the one originally employed. Together, these fundamental questions and modes of exploration serve as a guideline for the replication studies.

The present study primarily engages with the first fundamental question, while leaving the second to the discussion. Specifically, by scaling up, we extend the sample from 176 German students (Litteck et al., 2024) to 373 Taiwanese high school students to enhance statistical reliability. By scaling out, we situate the investigation in Taiwan's distinct educational and cultural context, recognizing the well-documented differences in both student performance and teacher orientations between Taiwan and Germany (OECD, 2023; Dreher et al., 2021). Through methodological variation, we supplement the original matrix-design partial-credit modeling (Litteck et al., 2024) with full-test administration and additional confirmatory factor analysis (CFA), allowing comparison of dimensional structures across different psychometric frameworks.

Building on these considerations, this study addresses the following research questions for replicating the German study with Taiwanese data:

RQ1: How do item difficulty patterns compare across operational and structural aspects within the three derivative sub-concepts for Taiwanese students?

RQ2: How well do multidimensional models fit the data from Taiwanese students compared with other competing models?

METHOD

Participants and Test Administration

The participants comprised 373 Grade-12 students (aged 17-18) from four Taiwanese senior high schools, recruited by convenience sampling to represent a range of academic selectivity. Although some participants were enrolled in science-track and others in social-science-track programs (Ministry of Education, 2018), they were treated as a single group in this study to provide an initial overview of Taiwanese students' common response patterns in solving derivative tasks. Considering that Litteck et al. (2024) administered a 14-item version (with matrix-design) within 35 minutes, the present study adopted an extended 18-item version administered as a 50-minute in-class review exercise following instruction on derivatives. Those items were translated into Chinese and reviewed by three experienced secondary mathematics teachers for linguistic accuracy and local conventions. Figure 2 illustrates one example item, which is a constructed-response task targeting the operational aspect of the derivative function. Completed test sheets, including students' written work, were collected for subsequent scoring and analysis.

Figure 2: An example item (targeting operational aspect of derivative function)

Population

Using the formula $P(t) = t^2 + 3t + 50$, the size of a population of frogs can be approximated for the next 20 years. Using the differential quotient $\lim_{h \to 0} \frac{f(x_0+h)-f(x_0)}{h}$, generate a formula that provides the instantaneous rate of change of the size of the population for an arbitrary point in time (in the next 20 years). Write down your calculations.

Data Analysis Process

All students' responses were scored following the coding scheme provided by Litteck et al. (2024), with two trained coders independently coding open responses. Inter-rater reliability ranged from κ = .764 to .977, and any disagreements were discussed and resolved jointly. Statistical analyses were conducted in R (version 4.4.2) using the packages TAM, eRm, lavaan, and psych. The internal consistency of the 18-item scale was high (Cronbach's α = .85). To address RQ1, we followed the procedure of Litteck et al. (2024). A linear partial credit model was fitted to the response data to obtain difficulty estimates and 95% confidence intervals for six item groups (operational and structural aspects within the difference quotient, differential quotient, and derivative function). These group-level estimates allowed comparison of the relative difficulty between operational and structural knowledge in three sub-concepts.

To address RQ2, we first replicated Litteck et al.'s IRT-based model comparisons (unidimensional, 2-dim models in terms of operational and structural aspects, and 3-dim models in terms of sub-concepts). Building on those results, we further extended the inquiry using confirmatory factor analysis: a series of nested CFA models (single-factor, aspect-based, sub-concept-based, and more fine-grained partitions) were estimated and compared. Model comparisons used Akaike and Bayesian information criteria (AIC, BIC) and likelihood-ratio tests (ANOVA) within the lavaan framework; fit indices CFI, RMSEA, and SRMR were examined to evaluate factorial plausibility.

RESULTS

Item-group difficulty patterns (RQ1)

Table 1 summarizes the IRT difficulty estimates and 95% confidence intervals for operational and structural item groups within each of the three derivative sub-concepts. The table also includes the estimates reported by Litteck et al. (2024) to allow cross-study comparison. Nevertheless, because IRT difficulty estimates depend on calibration, scaling, and sample characteristics, the absolute logit values are not directly comparable across the two studies. Interpretation should therefore focus on the relative patterns, which are outlined below.

Table 1: Numerical estimates for the difficulty of item groups

	Results in this study		Results in Litteck et al. (2024)	
	Difficulty estimate (logit)	95%-CI	Difficulty estimate (logit)	95%-CI
Difference Quotient				
Operational	0.77	[0.34, 1.20]	1.02	[0.65, 1.38]
Structural	-1.60	[-1.89, -1.31]	-1.65	[-2.05, -1.26]
Differential Quotient				
Operational	0.58	[0.24, 0.92]	0.97	[0.57, 1.38]
Structural	-1.86	[-2.21, -1.50]	-0.44	[-0.77, -0.12]
Derivative Function				
Operational	2.53	[1.90, 3.17]	2.24	[1.74, 2.73]
Structural	-1.55	[-1.88, -1.22]	1.00	[0.67, 1.34]

For the Taiwanese data, operational items were consistently more difficult than structural ones, indicating that tasks requiring operational knowledge were more challenging than those assessing structural knowledge. This overall trend was also observed in the German data reported by Litteck et al. (2024), showing a similar pattern across both Taiwanese and German contexts. However, the pattern across sub-concepts diverged, particularly for the structural aspect. While Litteck et al. (2024) found clearer

differentiation among the three structural sub-concepts, the Taiwanese data showed relatively similar difficulty levels across them (−1.60, −1.86, and −1.55, respectively). Although the difference was not statistically significant, the structural differential quotient appeared slightly less difficult than the other two in the present study. For the operational aspect, both datasets exhibited a comparable pattern (similar difficulty for the first two sub-concepts, with the derivative function items being the most difficult). Notably, the Taiwanese operational derivative function reached a difficulty estimate above 2.5 logits, representing a substantial gap from their structural counterparts.

Model comparison and dimensionality (RQ2)

Table 2 summarizes the results of IRT model comparisons for different dimensional structures using AIC and BIC. While the AIC favored the three-dimensional (3D) model, the BIC indicated a better fit for the one-(1D) and two-dimensional (2D) model. This discrepancy reflects the criteria's different penalties for model complexity: AIC tends to prefer more complex models, whereas BIC is more conservative. In Litteck et al. (2024), the 2D model clearly outperformed the 1D and 3D alternatives, supporting a distinction between operational and structural dimensions. In contrast, the Taiwanese data did not show such a strong dominance of the 2D model, suggesting that a more nuanced dimensional structure may better capture students' response patterns.

Table 2: Information-criterion comparisons for multi-dimensional models

Model	loglike	Deviance	Npars	AIC	BIC
1-dim	-4402.77	8805.53	31	8867.53	8989.10
2-dim (O/S)	-4398.47	8796.95	33	8862.95	8992.36
3-dim (R/L/F)	-4391.02	8782.03	36	8854.03	8995.21

Note. R: ratio (difference quotient); L: limit (differential quotient); F: function (derivative function); D: limit + function; O: operational; S: structural.

Table 3: CFA fit-index comparisons for the hypothesized models

Models	AIC	BIC	CFI	RMSEA	SRMR
1-factor	68442.0	68583.2	0.947	0.039	0.040
2-factor (O vs. S)	68436.3	68581.4	0.952	0.038	0.040
2-factor (R vs. D)	68433.2	68578.3	0.954	0.037	0.040
3-factor (R, L, F)*	68431.2	68584.1	0.956	0.036	0.039
4-factor (RO, RS, DO, DS)	69431.7	68596.4	0.958	0.036	0.039
6-factor (RO, RS, LO, LS, FO, FS)*	68436.1	68636.1	0.961	0.036	0.037

Note. * with the warning "covariance matrix of latent variables is not positive definite."

To further examine the latent structure, confirmatory factor analyses (CFAs) were conducted on a broader set of competing models, as summarized in Table 3. Beyond the original 2D and 3D models, three additional models were included. This extension was motivated by two considerations. First, multicollinearity between the L (limit process) and F (derivative function) factors led to a non-positive-definite covariance matrix in some models. Second, prior research (Guo et al., under review) indicated that Taiwanese textbooks place less emphasis on the difference quotient (R), implying that a distinction between "derivative-related" (L+F) and "non-derivative" (R) structures might be theoretically meaningful. Under this extended analysis, the 3-factor model again provided the best performance according to AIC. Other fit indices varied across models, likely reflecting their differing sensitivity to model complexity. Overall, the 3-factor solution appeared to represent the Taiwanese data most adequately.

Finally, scaled χ^2 difference tests can be used to compare competing nested CFA models (Table 4). Compared with the 1D model, both the two-factor (operational-structural) and the three-factor (R/L/F) models provided significantly better fit ($\Delta\chi^2(1) = 4.96$, $p = .026$; $\Delta\chi^2(3) = 13.64$, $p = .003$). The six-factor model showed only a marginal improvement over the two-factor model ($\Delta\chi^2(14) = 23.77$, $p = .049$) and no significant advantage over the three-factor model ($\Delta\chi^2(12) = 15.85$, $p = .198$). These results indicate that, while additional factors slightly increase model fit, the three-factor solution remains the most parsimonious representation of the Taiwanese data (a χ^2 difference test is not appropriate for a direct comparison of the 2D and 3D model).

Table 4: Scaled χ^2 difference tests for nested CFA model comparisons

Comparison	Δdf	$\Delta\chi^2$ (S-B)	*p*
1-factor (uni.) vs. 2-factor (O/S)	1	4.96	.026 *
1-factor (uni.) vs. 3-factor (R/L/F)	3	13.64	.003 **
2-factor (O/S) vs. 6-factor (all)	14	23.77	.049 *
3-factor (R/L/F) vs. 6-factor (all)	12	15.85	.198

Note. $\Delta\chi^2$ values are based on Satorra-Bentler scaled difference tests (MLR estimator).

DISCUSSION

Interpretation of results

Considering the theoretical assumption that students' understanding of mathematical concepts develops from operational to structural stages (Dubinsky, 1991; Sfard, 1991), the present study replicated Litteck et al.'s (2024) finding concerning the phenomenon of pseudo-structural knowledge (Sfard, 1992). This suggests that even though learning is expected to proceed through reifying a process into an object, in practice many students can solve structural items successfully but fail to reproduce the underlying processes. A concrete example emerged in our data: when solving the problem in

Figure 2, many Taiwanese students first wrote the correct answer, 2t+3, and only afterwards attempted to justify it using the differential-quotient formula and often producing an incorrect derivation. This raises a broader question of how far educational goals should emphasize instrumental versus relational understanding (Skemp, 2012).

With regard to the ordering of estimated item difficulty, the Taiwanese pattern diverged from the hierarchical structure of the three sub-concepts described by Litteck et al. (2024) and Zandieh (2000). Taiwanese students performed almost equally across the three sub-concepts on structural items, indicating that their mastery of derivative concepts may reflect curricular familiarity rather than the intrinsic logical progression within the mathematics. As Guo et al. (under review) observed, Taiwanese textbooks often treat the difference quotient only briefly before moving quickly to the differential quotient and then to applied topics such as tangent-line equations, whereas German textbooks typically follow the canonical sequence of the three sub-concepts. These curricular differences likely shape how the conceptual structure of derivatives develops in each educational context and highlight that finer-grained aspects of concept formation and application must be considered alongside the sub-concepts themselves.

Such contextual variation may also explain why the preferred latent-structure models differed between this study and Litteck et al. (2024). For the German data, a structure distinguishing knowledge types fits best, while for the Taiwanese data such a model seems to fit just as well as the 3D structure based on the sub-concepts. This outcome may partly stem from the curricular emphasis described above and partly from instructional orientations: Dreher et al. (2021) reported that Taiwanese preservice teachers tend to emphasize dynamic and static aspects of mathematics to a similar extent, whereas German teachers often stress process aspects more strongly. This raises an interesting issue of to what extent the textbooks instruction and teachers' orientation may shape students' cognitive structures of derivative understanding.

Limitations and future research directions

While this study followed the principles of replication research (Aguilar, 2020), as noted in the Results, our analysis and that of Litteck et al. (2024) were conducted independently, so the absolute parameter estimates cannot be directly compared. Even so, meaningful cross-study patterns were identified, suggesting that independently conducted replications can generate valuable insights and motivate collaborative international research. Combining data from multiple countries would enable a clearer characterization of students' cognitive profiles. Besides, the present sample included students from different academic tracks (Ministry of Education, 2018); analyzing them as a single group obscures potential differences across learning needs. Future studies should therefore compare students systematically across tracks and educational levels to identify specific weaknesses and inform targeted instructional interventions.

To capture cross-context cognitive structures more precisely, further refinement of the instrument is also needed. Each of the six sub-aspects was represented by only three

items, which may limit measurement stability, and potential translation effects could have altered item functioning. In addition, operational items were mostly constructed-response, possibly affecting dimensional balance. Developing cross-nationally equivalent items with balanced formats and sufficient coverage would provide a stronger basis for modeling students' understanding of derivatives across contexts.

References

Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education, 18*(Suppl 1), 37–50.

Burkhardt, H. (2013). Methodological issues in research and development. In Y. Li & J. N. Moschkovich (Eds.), *Proficiency and beliefs in learning and teaching mathematics: Learning from Alan Schoenfeld and Günter Törner* (pp. 203–236). Sense.

Dreher, A., Lindmeier, A., Feltes, P., Wang, T.-Y., & Hsieh, F.-J. (2021). Do cultural norms influence how teacher noticing is studied in different cultural contexts? A focus on expert norms of responding to students' mathematical thinking. *ZDM, 53*(1), 165–179.

Dubinsky, E. (1991). Reflective abstraction in advanced mathematical thinking. In D. Tall (Ed.), *Advanced Mathematical Thinking* (pp. 95–126). Springer.

Litteck, K., Rolfes, T., & Heinze, A. (2024). The structure of knowledge about the concept of derivative – a study investigating a process-object framework. *International Journal of Mathematical Education in Science and Technology, 56*(11), 2164–2184.

Ministry of Education (2018). *Curriculum Guidelines for the 12-Year Basic Education Elementary School, Junior High School, and Upper Secondary School: The Domain of Mathematics.*

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education.* OECD-Publishing.

Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on processes and objects as different sides of the same coin. *Ed. Studies in Mathematics, 22*(1), 1–36.

Sfard, A. (1992). Operational origins of mathematical objects and the quandary of reification: The case of function. In E. Dubinsky & G. Harel (Eds.), *MAA notes: Vol. 25. The concept of function: Aspects of epistemology and pedagogy* (pp. 59–84). Mathematical Association of America.

Skemp, R. R. (2012). *The psychology of learning mathematics: Expanded American edition.* Routledge.

Star, J. R. (2018). When and why replication studies should be published: Guidelines for mathematics education research journals. *Journal for Research in Mathematics Education, 49*(1), 98–103.

Zandieh, M. (2000). A theoretical framework for analyzing student understanding of the concept of derivative. In E. Dubinsky, A. H. Schoenfeld, & J. Kaput (Eds.), *Research in Collegiate Mathematics Education* (Vol. IV, pp. 103–127). Am. Mathematical Society.