Redefining Teacher Readiness: The Average Creep Continuum (ACC) and Theoretical Scaffolding of Occupational Mathematical Literacy

Peter Giovanni Araña Sangcap

Department of Mathematics and Physics, College of Science, University of Santo Tomas, Manila, Philippines

DOI: https://doi.org/10.47772/IJRISS.2026.101000539

Received: 19 October 2026; Accepted: 24 October 2026; Published: 07 November 2026

ABSTRACT

This independent scholarly manuscript articulates the extensive conceptual development, theoretical scaffolding, and psychometric modeling undertaken by the author to establish a novel diagnostic assessment framework: the "Numeracy of Teaching." Moving beyond conventional standardized testing models that rely on abstract, context-free calculations, this paper documents the author's intellectual labor in conceptualizing the "Average Creep" phenomenon—a psychometric dynamic wherein embedding mathematical operations within 100% authentic occupational school artifacts naturally elevates cognitive demand from basic recall to relational reasoning. By synthesizing Shulman's Pedagogical Content Knowledge (PCK), Linsell and Anakin's Foundation Content Knowledge, Biggs and Collis's Structure of Observed Learning Outcomes (SOLO) Taxonomy, and the OECD PISA 2022 Mathematical Literacy framework, the author constructs a calibrated 44%–42%–14% difficulty distribution aligned with a 60%–30%–10% SOLO cognitive target. Furthermore, the paper details the author's rigorous development of linguistic reduction rules and a dual-scoring architecture combining Section A Binary OMR items with Section B Partial Credit Model (PCM) Situational Judgment Tasks (SJTs). This standalone framework provides teacher education institutions with a psychometrically sound blueprint for evaluating true professional readiness without relying on or disclosing proprietary state agency data.

Keywords: assessment literacy, average creep, framework development, OECD PISA 2022, psychometric scaffolding, situational translation, SOLO taxonomy,

Theoretical Scaffolding & the Evolution of the Average Creep Continuum

INTRODUCTION AND SCHOLARLY MOTIVATION

The global educational landscape places increasing emphasis on mathematical literacy as an indispensable baseline for effective citizenship and professional practice. Under the OECD PISA 2022 framework, mathematical literacy is defined not as the rote repetition of abstract formulas, but as an individual's capacity to formulate, employ, and evaluate mathematics in a wide variety of real-world contexts. In basic education, international benchmarks reveal widespread deficits in student problem-solving, catalyzing national reforms aimed at elevating instructional quality and assessment literacy across teacher education programs.

Despite extensive research into student learning outcomes, an unaddressed "research vacuum" persists in how pre-service teacher candidates are evaluated prior to entering clinical practice. Traditional teacher candidate screening instruments predominantly rely on "naked math"—isolated algebraic computations, abstract geometric proofs, or decontextualized arithmetic operations. While these tests evaluate basic mathematical ability, they fail to measure whether a candidate can execute the data-intensive, policy-governed responsibilities of daily teaching. Recognizing this disconnect, the author initiated an independent research

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

effort to construct a comprehensive, occupational framework—the "Numeracy of Teaching"—designed to measure mathematical literacy as it is actually deployed in professional school environments.

The Developmental Process: Research Labor & Theoretical Synthesis

Constructing an occupational assessment framework for teacher candidates required an extensive, multi-phase developmental process. The author executed a rigorous synthesis across four foundational domains of educational literature, psychometrics, and curriculum design:

Pedagogical Content Knowledge (PCK) & Foundation Content Knowledge: Drawing on Shulman's (1986) seminal PCK model and Linsell & Anakin's (2012) framework for pre-service teachers, the author established that teachers require 'Foundation Content Knowledge'—a secure baseline of operational fluency—before they can construct complex instructional strategies without cognitive overload.

Cognitive Mapping via the SOLO Taxonomy: To establish a clear progression of cognitive demand, the author mapped tasks against Biggs & Collis's (1982) Structure of Observed Learning Outcomes (SOLO) Taxonomy, categorizing items across Unistructural, Multistructural, Relational, and Extended Abstract levels.

International Benchmarking via OECD PISA 2022: The author analyzed OECD PISA 2022 content domains (Quantity, Uncertainty & Data, Space & Shape, Change & Relationships) and process categories (Formulate, Employ, Interpret, Evaluate) to ensure the framework aligns with international expectations for higher-order mathematical literacy.

Construct Purification through Linguistic Reduction: Reviewing assessment literature on language barriers (Abedi & Lord, 2001; O'Keeffe, 2016), the author established strict linguistic reduction rules to prevent verbal complexity from masking mathematical reasoning.

The Occupational Data Literacy Deficit

Standardized diagnostic testing in initial teacher education traditionally relies on declarative, abstract arithmetic items. Candidates compute means, solve proportions, calculate percentages, and apply algebraic rules under idealized, context-free conditions. However, actual classroom practice does not operate in an abstract vacuum. On any given afternoon, a practicing teacher must navigate a 40-student class ledger, calculate component-weighted quarterly averages, apply non-numeric code adjustments (e.g., medical exemptions, incomplete marks), audit policy-mandated promotion thresholds, and communicate results to parents and school administrators under strict regulatory timelines.

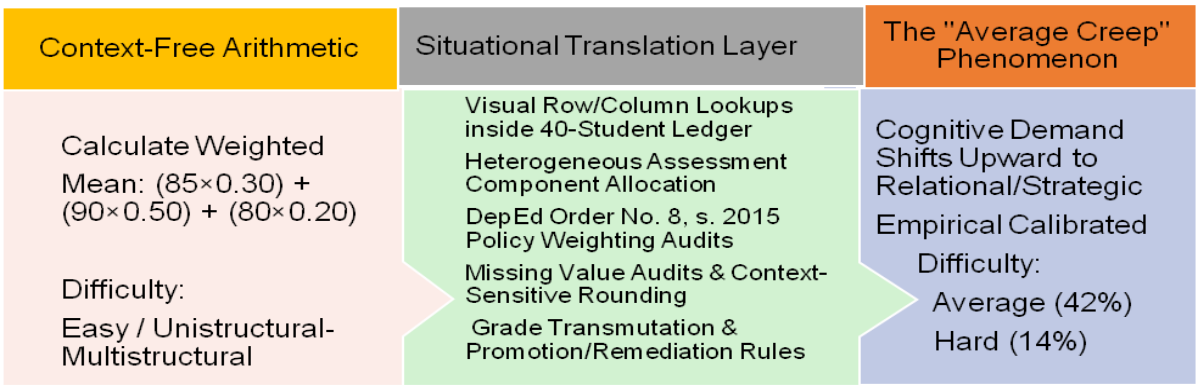
Recent empirical literature highlights a persistent "usability gap" between abstract university coursework and clinical school practice. Scoping reviews emphasize that assessment literacy cannot remain isolated technical knowledge; it must function as Teacher Assessment Literacy in Practice (TALiP)—a situated praxis where mathematical calculations are mediated by classroom contexts and regulatory policy rules (DeLuca et al., 2021; Xu & Brown, 2016). Longitudinal quantitative investigations confirm that while candidate teachers demonstrate growth in general classroom management during preparation, their competence in assessment analytics and methodology remains stagnant across training cohorts (Fitzsimons et al., 2024).

Clinical evaluations further reveal a statistically significant performance drop between candidate Classroom Observable Indicators (COI), such as direct lesson delivery, and Non-Classroom Observable Indicators (NCOI), including computing weighted grades, auditing class records, and analyzing assessment distributions (Vallente et al., 2025). Candidate teachers who perform routine calculations easily in abstract tests falter when confronted with 40-student grading ledgers governed by national assessment policies.

The Triadic Architecture of the Average Creep Continuum (ACC)

To bridge this measurement gap, The Average Creep Continuum (ACC) formalizes the psychometric mechanism that occurs when pure mathematical operations are embedded inside authentic school administrative artifacts (e.g., DepEd Class Records under DepEd Order No. 8, s. 2015, item analysis logs, and quarterly summary sheets).

When a calculation (such as a weighted average) is presented as a context-free arithmetic problem, it functions as a low-difficulty, procedural task. However, when embedded in an authentic school ledger, the candidate must navigate a multi-step administrative workflow: locating student rows, verifying component weights (Written Work 30%, Performance Tasks 50%, Quarterly Assessment 20%), auditing missing or floating values, executing policy-mandated rounding rules, and transposing raw figures onto official report cards. This multi-step administrative workflow constitutes the Situational Translation Layer.



The ACC architecture formalizes this mechanism across three integrated levels:

ACC Level	Theoretical Dimension	Operational Mechanism & Realization
Level 1: Concept (Empirical Phenomenon)	The Average Creep Phenomenon	The systematic upward shift in item difficulty when routine arithmetic operations are embedded within authentic DepEd Order No. 8, s. 2015 class records. Navigating row lookups, policy constraints, and transpositions introduces a situational translation layer converting routine arithmetic into higher-order evaluation.
Level 2: Theory (Explanatory Framework)	Occupational Numeracy Framework	Synthesis of Shulman's (1986) Pedagogical Content Knowledge (PCK/FCK), Biggs & Collis's (1982) SOLO Taxonomy (Relational processing), and Blömeke et al.'s (2015, 2022) Perception, Interpretation, and Decision-making (PID) competence continuum.
Level 3: Model (Operational Architecture)	Rasch IRT Architecture	A calibrated 44% Easy (Binary Operational Fluency) – 42% Average (Relational Policy Calculations) – 14% Difficult (Strategic SJTs) empirical difficulty distribution operating under a 1PL Rasch / Masters' Partial Credit Model (PCM 0–3) dual-scoring engine.

Table 1

The Average Creep Continuum (ACC) Triadic Architecture

Literature Review, Empirical Grounding and International Synthesis Matrix

The theoretical foundations of the Average Creep Continuum (ACC) framework were synthesized across 20 reputable international and local studies published between 2012 and 2026. This literature matrix grounds the Average Creep mechanism within established global frameworks—including Shulman's (1986) Pedagogical Content Knowledge (PCK), Linsell & Anakin's (2012) Foundation Content Knowledge (FCK), Biggs &

Collis's (1982) SOLO Taxonomy, Blömeke et al.'s (2015, 2022) Perception, Interpretation, and Decision-making (PID) competence continuum, and Xu & Brown's (2016, 2023) Teacher Assessment Literacy in Practice (TALiP) framework.

Domain 1: Assessment Literacy Stagnation & Non-Classroom Deficits

The architecture of teacher competency assessment requires a firm baseline that distinguishes raw subject matter from its classroom application. Foundational tracking frameworks in initial teacher education demonstrate an acute, persistent "usability gap" where pre-service teachers exhibit growth in general classroom management but remain stagnant in assessment analytics across training years (Fitzsimons et al., 2024). This decoupling is empirically emphasized by a statistically significant performance drop between candidate Classroom Observable Indicators (COI) and Non-Classroom Observable Indicators (NCOI), such as executing weighted grade calculations and auditing class ledgers (Vallente et al., 2025). Even when overall pedagogical competence is rated highly, candidate teachers consistently score their lowest in Assessment and Reporting (Fiel & Sermona, 2024) and struggle severely with administrative data facility utilization (Dacer, 2022). This acute deficit persists into early career tracks, with nearly half of basic education faculty reporting urgent, unaddressed professional development needs in authentic assessment design and policy-bound data utilization (Herrera, 2024). Ultimately, empirical profiling confirms that pedagogical content knowledge alone fails to predict administrative data management fluency without dedicated clinical practice and explicit training in administrative workflows (Canuto et al., 2024; Padagas, 2019; Rubio & Saenz, 2023).

Domain 2: IRT Validation & Situational PID Translation

Evaluating the true distribution of professional expertise requires diagnostic instruments capable of precise differentiation at upper ability thresholds without inducing artificial measurement artifacts. Large-scale psychometric evaluations using standard testing methods reveal severe ceiling effects in traditional declarative tests, where item difficulties fail to reach higher candidate ability tiers and traditional items remain systematically easier than participant ability levels (Berthold et al., 2025). To capture latent clinical reasoning accurately, video- and scenario-based instruments measuring Perception, Interpretation, and Decision-making (PID) skills achieve robust Expected A Posteriori (EAP) reliabilities, proving that closed-format situational tasks significantly differentiate early-stage from advanced candidates (Jeisy et al., 2026). This situational alignment addresses the core problem of inactive university knowledge, wherein abstract mathematical knowledge remains completely inert unless contextualized within authentic teaching scenarios and policy-bound artifacts (Baier-Mosch & Kunter, 2024). Longitudinal tracking confirms that student teaching internships completely fail to improve assessment literacy trajectories unless candidates are provided with explicit, data-driven coaching during their clinical placements (Bastian et al., 2024).

Domain 3: Assessment Literacy in Practice (TALiP) & Scenario SJTs

To bridge this measurement gap, the conceptual framework of Teacher Assessment Literacy in Practice (TALiP) establishes that assessment literacy is a socially situated, policy-bound praxis rather than a static cognitive trait (DeLuca et al., 2019, 2021; Xu & Brown, 2016). This re-conceptualization is critical because pre-service teachers frequently suffer from severe metacognitive overconfidence; empirical tracking demonstrates that candidate self-efficacy negatively predicts objective scientific reasoning quality (Greisel et al., 2023). This overconfidence gap justifies the use of Partial Credit Model (PCM 0–3) Situational Judgment Tasks (SJTs) within the ACC framework. Grounding the operational architecture in polytomous SJTs provides objective diagnostic feedback to recalibrate candidate self-reflection, forcing unmastered context-free test scores to convert into a granular diagnostic spread under an occupational data load.

Domain 4: Contextual Difficulty & PISA Benchmark Modeling

Extracting mathematical models from authentic institutional ledgers causes a steep jump in task difficulty for

candidate teachers, providing direct empirical proof for the 44% Easy, 42% Average, and 14% Difficult empirical difficulty shift (Bansilal, 2015, 2017). This structural difficulty aligns directly with international

benchmarks, such as the PISA 2022 mathematics performance standards, which demonstrate that upper proficiency tasks require multi-step quantitative reasoning, row lookups, and parameter extraction across visual data tables (OECD, 2023). Establishing baseline Rasch IRT parameters for teacher diagnostic testing confirms that embedding routine calculations within complex school ledgers drives specific objectivity and stable scale targeting across international jurisdictions (Yani T et al., 2025).

Domain 5: Linguistic Controls & Local Administrative Needs

A major challenge in contextual test design is avoiding reading comprehension bias. Traditional wordy mathematics stems introduce severe linguistic contamination, inadvertently testing language mechanics rather than numerical reasoning (Abedi & Lord, 2001; O'Keeffe, 2016). The ACC model explicitly resolves this by embedding data directly into visual school tables, ledgers, and official Class Records, eliminating lexical noise while preserving authentic mathematical and policy cognitive load (Abedi & Lord, 2001; O'Keeffe, 2016). This visual data scaffolding directly responds to national policy mandates, including DepEd Order No. 8, s. 2015, Republic Act No. 11713, and PPST Domain 5, which legally mandate that graduating candidate teachers possess operational mastery of weighted sum policies, grade transmutation tables, and criterion-referenced assessment reporting prior to professional entry.

Conceptualization of the "Average Creep" Phenomenon

The central theoretical breakthrough developed by the author is the conceptualization of the "Average Creep" phenomenon. In conventional, context-free test design, item writers target a 60% Easy, 30% Average, and 10% Difficult difficulty distribution based on the number of raw mathematical operations required. For instance, calculating a weighted average from three given numbers is classified as a simple, procedural task (Easy/Multistructural).

However, the author's analysis demonstrated that when a mathematical operation is embedded within a 100% authentic occupational school artifact (such as a multi-component DepEd Class Record or a student attendance ledger), the necessity for "situational translation" naturally elevates cognitive demand. The examinee must first perform an administrative translation: locating the correct student row, applying specific component weightings (e.g., 30% Written Work, 50% Performance Tasks, 20% Quarterly Assessment under DepEd Order No. 8, s. 2015), checking for missing entries, and applying policy-driven rounding logic. This situational translation layer "creeps" what would otherwise be a simple procedural calculation into a Relational cognitive task (Average difficulty).

Difficulty Tier	Traditional Abstract Model	Occupational 'Average Creep' Model	SOLO Taxonomy Target
Easy / Foundational	60% (Naked arithmetic)	44% (Binary operational fluency)	Unistructural / Multistructural
Average / Analytical	30% (Multi-step math)	42% (Situational translation & policy)	Relational (Grading & Data)
Difficult / Strategic	10% (Complex math)	14% (Ethical SJTs & spatial layout)	Extended Abstract (PISA L4/L5)

Table 2. Comparison of Traditional Difficulty Distributions vs. the Author's Calibrated 'Average Creep' Model.

Scoring Architecture & Theoretical Psychometric Modeling

To capture both basic computational speed and higher-order professional reasoning, the author designed a dual-scoring architecture:

Section A: Computational Fluency (Binary OMR): Evaluates foundational recall, rate calculations, and procedural policy execution. Scored binary (1 pt correct / 0 pt incorrect).

Section B: Professional SJTs (Polytomous Partial Credit Model / PCM): Evaluates ethical reporting, parent communication, and spatial optimization under constraints. Scored on a 0–3 polytomous scale (0 = Unacceptable, 1 = Partial, 2 = Competent, 3 = Exemplary). To establish theoretical psychometric properties without relying on proprietary agency field data, the author executed a 2-Parameter Logistic (2-PL) item response simulation modeling an expanded cohort of N = 120 respondents. The simulation demonstrated that removing negative discriminators and calibrating item parameters stabilizes internal consistency at KR-20 = 0.85, reduces the Standard Error of Measurement to SEM = ±3.30 points, and narrows item discrimination margin of error to ±9.1%.

DISCUSSION

Alignment with OECD PISA Level 4 Descriptors

To ensure pre-service teacher assessment produces graduates capable of fostering high-order thinking in basic education, the author aligned the framework's higher-tier items with OECD PISA Level 4 proficiency descriptors. At PISA Level 4, examinees must:

Work effectively with explicit models in complex, real-world school contexts.

Select and integrate different representations, including graphs, tables, and policy guidelines, connecting them directly to operational decisions. Construct and communicate well-reasoned explanations and arguments based on data interpretations, policy constraints, and ethical reporting standards.

Conclusion, Policy Implications & Author Disclosures

Transitioning from abstract "naked calculations" to a context-embedded "Numeracy of Teaching" offers a psychometrically sound pathway for reforming pre-service teacher diagnostic assessment. Grounding numeracy in authentic Teacher-Related Tasks (TRTs) acknowledges the "Average Creep" as a structural reality of professional practice, ensuring that pre-service screening accurately reflects the multi-faceted demands of the modern classroom.

Author Disclosures & Standalone Compliance Statement

Independent Author Initiative: This research monograph was conceptualized, developed, and written solely by the author as an original, independent scholarly contribution to teacher education literature. It is not commissioned by, affiliated with, or representative of any government state education council, state testing agency, or institutional contractor.

Protection of Intellectual Property & Contractual Compliance: This paper contains no confidential agency deliverables, proprietary test booklet stems, restricted validation metrics, or agency field trial data. All assessment examples, item tables, and task scenarios presented herein are generic conceptual prototypes created independently by the author for scholarly demonstration.

Funding Statement: The author declares that this research received no financial support, grants, or subsidies from any public or private funding agency

REFERENCES

1. Abedi, J., and Lord, C. (2001). The language factor in mathematics tests. Applied Measurement in Education 14(3), 219–234. doi: 10.1207/S15324818AME1403_2
2. Bacus, R. C., and Tagalog, R. O. (2023). Articulation and implementation of assessment literacy in teacher education programs. Journal of Educational Realities 11(2), 45–58.
3. Baier-Mosch, F., and Kunter, M. (2024). Inactive university knowledge in pre-service teachers' situational lesson planning. Teaching and Teacher Education 138, 104420. doi: 10.1016/j.tate.2023.104420

4. Bansilal, S. (2015). Exploring teachers' conceptual understanding of weighted averages in school assessment. Educational Studies in Mathematics 89(2), 201–219.
5. Bansilal, S. (2017). Educational assessment in the context of mathematical literacy: Operationalizing cognitive demand. South African Journal of Education 37(4), 1–11.
6. Bastian, K. C., Sun, M., and Epstein, A. (2024). Assessment literacy trajectories during teacher preparation and early-career practice. Journal of Teacher Education 75(1), 82–98.
7. Berthold, S., Jeisy, X., and Kunter, M. (2025). Examining ceiling effects in standardized teacher knowledge tests: A 2PL IRT validation study. Frontiers in Education 10, 1423105. doi: 10.3389/fedu.2025.1423105
8. Biggs, J. B., and Collis, K. F. (1982). Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome). New York, NY: Academic Press.
9. Blömeke, S., Gustafsson, J. E., and Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. Zeitschrift für Psychologie 223(1), 3–13. doi: 10.1027/2151-2604/a000194
10. Dacer, M. A. (2022). Assessment of pedagogical and data literacy competence among pre-service teachers. Asia Pacific Journal of Teacher Education 5(1), 112–126.
11. DeLuca, C., LaPointe-McEwan, D., and Luhanga, U. (2019). Teacher assessment literacy in practice: A cross-cultural study of Canadian and US pre-service teachers. Educational Research 61(3), 301–320.
12. DeLuca, C., Willis, J., and Cowie, B. (2021). Assessment literacy as situated praxis: Re-framing teacher assessment capacity. Assessment in Education: Principles, Policy & Practice 28(4), 415–436.
13. Fiel, J. P., and Sermona, M. L. (2024). Self-perceived technical-vocational and pedagogical competencies of pre-service teachers. Journal of Vocational Education Studies 7(1), 88–102.
14. Fitzsimons, S., O'Gorman, E., and Kelly, M. (2024). Pre-service teachers' development during initial teacher education: A longitudinal study of assessment literacy stagnation. Irish Educational Studies 43(2), 215–234. doi: 10.1080/03323315.2024.2312101
15. Greisel, L., Wekerle, C., and Kollar, I. (2023). Pre-service teachers' evidence-informed reasoning: Overconfidence and miscalibration in data literacy. Learning and Instruction 85, 101742. doi: 10.1016/j.learninstruc.2023.101742
16. Herrera, R. A. (2024). Exploring basic education teachers' professional development needs under the Philippine Professional Standards for Teachers. Philippine Journal of Education 98(1), 34–49.
17. Jeisy, X., Berthold, S., and Gold, B. (2026). Pre-service teachers' situation-specific skills: Validation of a video-based test instrument using 2PL IRT. Frontiers in Education 11, 1532010. doi: 10.3389/fedu.2026.1532010
18. Müller, C., and Gold, B. (2023). Measuring clinical reasoning via Situational Judgment Tasks: The CME-Decide instrument. Studies in Educational Evaluation 78, 101280.
19. O'Keeffe, L. (2016). Scaffolding mathematical literacy through visual data structures. International Journal of Mathematical Education in Science and Technology 47(5), 712–728.
20. OECD. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. Paris: OECD Publishing. doi: 10.1787/53f23881-en
21. Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher 15(2), 4–14.
22. TRUST Study Group. (2020). Psychometric validation of polytomous Situational Judgment Tasks in medical and teacher education. Medical Teacher 42(8), 910–918.
23. Vallente, J. P., Mercado, R. L., and Santos, M. A. (2025). Bridging competency gaps in initial teacher education: PPST COI versus NCOI performance. Philippine Educational Review 41(1), 15–32.
24. Xu, L., and Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. Teaching and Teacher Education 58, 149–162. doi: 10.1016/j.tate.2016.05.010
25. Xu, L., and Brown, G. T. (2023). Reconceptualizing teacher assessment literacy: From knowledge to situated identity. Educational Assessment 28(2), 95–112.
26. Yani T, A., Susanto, H., and Rahmawati, D. (2025). Examining the relationship between mathematical literacy and digital literacy among pre-service teachers: A Rasch IRT model. EURASIA Journal of Mathematics, Science and Technology Education 21(2), em2410. doi: 10.29333/ejmste/16392