The Proficiency Paradox: A Critical Analysis of Tom Schimmer and Standards-Based Assessment Models

Journal of Educational Standards & Accountability | Vol. 14, Issue 2, 2026
Authors: Mot Remmihcs, Erelyt Erutuoc, and Rachel Friederich
Date: February 12, 2026

Abstract

Modern educational reform, notably championed by practitioners advocating for "standards-based" mindsets, seeks to transition grading from point accumulation to pure proficiency reporting. These principles—specifically the use of integer scales, the decoupling of behavior from achievement, and the mandate for unlimited reassessment—aim to increase equity and accuracy. However, emerging critiques suggest a "proficiency paradox," where the removal of traditional accountability structures inadvertently diminishes student performance and creates a preparation gap for higher education. This article examines the theoretical and empirical shortcomings of these principles by synthesizing educational theory with advanced computational models of learning behavior. By leveraging insights from knowledge tracing, engagement propagation, and agentic simulation frameworks, we argue that the reductionist nature of current standards-based grading (SBG) models fails to capture the complexity of student learning trajectories. We propose that without the integration of granular data and behavioral accountability, standards-based models may obscure rather than reveal true academic competence.

I. Introduction

The landscape of secondary education assessment has undergone a radical transformation over the last decade, moving away from the traditional 100-point percentage scale toward standards-based grading (SBG) and reporting. Proponents of this shift, such as Tom Schimmer, argue that grades must be pure representations of content mastery, stripped of "non-academic" factors like timeliness, participation, or effort. This philosophy advocates for a coarse-grained integer scale (e.g., 1–4), the elimination of zeros for missing work, and a focus on "learning over time" through continuous reassessment. While the stated goal is to improve the accuracy of communicating what a student knows, a growing body of evidence suggests a paradoxical outcome: as grades become more "accurate" in theory, the actual rigor of student output and their readiness for post-secondary environments appears to decline.

The core problem lies in the oversimplification of the learning process. By collapsing complex performance metrics into limited categories and removing the feedback loops provided by behavioral accountability, schools risk creating a "soft skills vacuum." Furthermore, the mathematical argument against the percentage scale often ignores the nuance required to differentiate high-level competency from bare proficiency. Existing critiques have largely been philosophical; however, there is a need to apply rigorous data-analytic perspectives to understand why these reforms often fail in practice. If we view a student's learning trajectory as a complex probabilistic distribution rather than a static integer, the flaws in the SBG model become statistically apparent.

This paper addresses these issues by juxtaposing SBG principles with modern computational learning sciences. Specifically, we argue that the "Proficiency Paradox" arises from three methodological failures:

- Granularity Loss: The shift to integer scales exacerbates data sparsity and prevents the detection of subtle shifts in student cognition, a problem mirrored in knowledge tracing tasks where robust representations are difficult to learn without sufficient variance (Zhang et al., 2023).
- Behavioral Isolation: Decoupling behavior from grades ignores the social contagion of engagement, where the lack of accountability for one student can negatively propagate through a peer network (Fan et al., 2025).
- The Reassessment Fallacy: Unlimited unstructured reassessment creates administrative bottlenecks and moral hazard, whereas effective competency training requires structured, agentic feedback loops rather than mere repetition (Marez et al., 2025).

II. Related Work

A. Knowledge Tracing and Representation Granularity

The fundamental premise of SBG is that a 1–4 scale is superior to a 100-point scale because the latter implies a false precision. However, in the field of educational data mining, specifically

Knowledge Tracing (KT), the reduction of data granularity is often viewed as a hindrance to accurate prediction. Zhang et al. highlight that learning robust representations of student knowledge is already challenging due to "data sparsity," particularly for students with few practice records (Zhang et al., 2023). When schools artificially restrict the variance of assessment data by capping scores at "4" or removing the weight of formative attempts, they inadvertently increase this sparsity. The Cognition-Mode Aware Variational Representation Learning Framework (CMVF) demonstrates that accounting for uncertainty and distribution in student data is essential for accurate modeling (Zhang et al., 2023). By forcing student performance into rigid integer buckets, SBG models discard the "probabilistic" nature of learning, effectively masking the uncertainty that educators need to address.

B. Student Engagement and Behavioral Dynamics

A controversial tenet of Schimmer's philosophy is the separation of "soft skills" (behaviors) from academic grades. The argument is that penalizing late work distorts the measurement of content mastery. However, recent research in hypergraph convolutional networks suggests that engagement is not an isolated variable but a complex, contagious phenomenon. Fan et al. propose the DS-HGCN model, which demonstrates that student engagement is multi-dimensional (behavioral, emotional, cognitive) and propagates through social networks in a classroom (Fan et al., 2025). By treating behavior as irrelevant to the "grade," SBG ignores the "social contagion" effect where low behavioral standards (e.g., widespread lateness allowed by policy) reduce the overall engagement energy of the class. This indicates that behavioral metrics are not merely "noise" to be filtered out, but critical features that predict academic success and influence the learning environment of peers (Fan et al., 2025).

C. Simulation and Assessment Scalability

The "reassessment loop"—allowing students to retake assessments until mastery is demonstrated—is theoretically sound but logistically fragile. It often leads to teacher burnout and "strategic procrastination" by students. In professional training contexts, such as medical education, valid reassessment is not achieved by simply repeating a test, but through highly structured simulations. De Marez et al. introduce an agentic AI framework for training skills, utilizing "virtual simulated patients" to provide standards-based assessment and feedback (Marez et al., 2025). Their work highlights that effective competency-based assessment requires "scenario control" and "consistent roleplaying" to be valid (Marez et al., 2025). This contrasts sharply with the typical K-12 SBG implementation, where reassessments are often lower-stakes versions of the original test. The literature suggests that without the rigorous architecture of an agentic framework, the "learning over time" principle becomes a loophole for reduced rigor rather than a mechanism for mastery.

III. Methodological Analysis: The Proficiency Paradox Framework

To critically evaluate the Standards-Based Assessment model, we propose a theoretical evaluation framework: the Multi-Dimensional Proficiency Verification (MDPV) model. This approach utilizes the computational concepts found in recent learning analytics literature to audit the validity of SBG principles.

Phase 1: Granularity Impact Assessment

We first analyze the "Myth of Mathematical Accuracy" regarding integer scales. In traditional grading, a score of 89% vs. 91% provides a variance signal. In SBG, both might be labeled "Proficient (3)."

- Methodology: We apply the logic of the Cognition-Mode Aware Variational Representation Learning Framework (Zhang et al., 2023). If we treat a student's true ability as a latent variable with a probability distribution, the "integer scale" acts as a high-compression filter.
- Hypothesis: By applying a variational inference model to hypothetical gradebook data, we posit that the "posterior distribution" of student ability becomes flatter (less informative) under a 1–4 scale compared to a percentage scale. The integer scale fails to distinguish between a student who barely meets the standard and one who approaches mastery, increasing the "risk of model overfitting" where a student is falsely classified as secure in their knowledge (Zhang et al., 2023).

Phase 2: Behavioral Contagion Modeling

This phase critiques the "Behavioral Decoupling" principle. We utilize the construct of social contagion in student engagement (Fan et al., 2025).

- Methodology: We model a classroom as a hypergraph where students are nodes and "work habits" are the edges of influence.

In an SBG system, the "cost" of poor behavior (e.g., turning in work weeks late) is zeroed out.

- Mechanism: Using the propagation mechanisms described in DS-HGCN (Fan et al., 2025), we simulate the effect of removing deadlines. The model suggests that when the "engagement signal" is decoupled from the reward system (grades), the "contagion" of disengagement spreads more rapidly.
- Implication: The "soft skills vacuum" is not just an individual deficit but a systemic network failure. The removal of penalties removes the feedback signal necessary for the hypergraph attention mechanism to weigh positive behaviors effectively (Fan et al., 2025).

Phase 3: Reassessment Rigor Evaluation

Finally, we examine the "Reassessment Loop." Schimmer's model assumes that the latest evidence is the only relevant evidence.

- Methodology: We compare manual classroom reassessment against the "Agentic AI Framework" standard (Marez et al., 2025). True mastery requires transfer learning—applying skills to novel scenarios.
- Critique via Framework: Current SBG reassessments often lack "evidence-based vignette generation" (Marez et al., 2025). Students often memorize the test format rather than the content. A valid reassessment model would require the "agentic separation of scenario control" (Marez et al., 2025), ensuring that every retake is a distinct, dynamically generated problem instance. Without this technological infrastructure, the manual reassessment policy leads to inflated scores that do not reflect transferrable skills.

IV. Discussion

Practical Implications and Deployment

The analysis suggests that while the ethical motivations of Standards-Based Grading are laudable, the implementation creates significant data fidelity issues. Schools adopting these models must realize that "simplifying" the gradebook actually destroys valuable information regarding student cognition and behavior. The "Proficiency Paradox" implies that by trying to be more "accurate" (by focusing only on content), we become less predictive of future success. Real-world environments—universities and workplaces—operate on competitive, high-granularity feedback loops where "soft skills" like timeliness are inextricably linked to performance evaluations.

Limitations and Failure Modes

The critique presented here identifies several failure modes in the current SBG paradigm:

1. Grade Compression: The integer scale creates a "ceiling effect," making it impossible to distinguish between the top 10% and top 1% of students, a distinction crucial for elite university admissions.
2. The "Safety Net" Hazard: Infinite reassessment creates a moral hazard. Students delay studying, relying on the safety net, which disrupts the consolidation of memory.
3. Algorithmic Blindness: As noted in engagement research, analyzing students based on single-dimensional features (content only) fails to capture the holistic engagement state required for deep learning (Fan et al., 2025).

Ethical Considerations and Future Work

Ethically, there is a risk that SBG creates a "preparation gap." Students from schools with rigorous behavioral accountability may be better prepared for the executive function demands of higher education than those from "pure" SBG environments. Furthermore, relying on teacher discretion for 1–4 scaling introduces subjective bias that percentage scales, while imperfect, help mitigate through aggregation.

Future work should focus on developing "Hybrid Assessment Models." These would integrate the diagnostic precision of Knowledge Tracing (Zhang et al., 2023) to track content mastery while maintaining a separate, weighted vector for behavioral engagement, potentially modeled through hypergraph networks (Fan et al., 2025). Additionally, the administrative burden of reassessment could be solved by deploying Generative AI agents to create unlimited, unique, standards-based practice scenarios (Marez et al., 2025), ensuring that "retakes" represent true learning rather than memorization.

V. Conclusion

The transition to Standards-Based Assessment represents a significant philosophical shift in education, yet it is fraught with methodological perils. By analyzing these pedagogical strategies through the lens of computational learning science, we uncover a "Proficiency Paradox": the very mechanisms designed to enhance accuracy and equity—integer scales, behavioral decoupling, and unlimited reassessment—may ultimately degrade the quality of education. The loss of data granularity obscures true student trajectories (Zhang et al., 2023), the dismissal of behavioral metrics ignores the contagious nature of engagement (Fan et al., 2025), and the lack of structured rigor in reassessment fails to meet the standards of valid simulation (Marez et al., 2025). To resolve this paradox, educational leaders must move beyond the simplistic binary of "traditional vs. standards-based" and embrace a more sophisticated, data-aware approach that values both content mastery and the executive behaviors necessary to achieve it

References

Zhang, Moyu, Zhu, Xinning, Zhang, Chunhong, Pan, Feng, Qian, Wenchen, & Zhao, Hui (2023). *Cognition-Mode Aware Variational Representation Learning Framework for Knowledge Tracing*. 2023 ICDM. https://arxiv.org/pdf/2309.01179v1

Fan, Ziyang, Tao, Li, Wang, Yi, Qu, Jingwei, Wang, Ying, & Jiang, Fei (2025). *DS-HGCN: A Dual-Stream Hypergraph Convolutional Network for Predicting Student Engagement via Social Contagion*. https://arxiv.org/pdf/2512.20059v1

Marez, Victor De, Nooten, Jens Van, Bruyne, Luna De, & Daelemans, Walter (2025). *An Agentic AI Framework for Training General Practitioner Student Skills*. https://arxiv.org/pdf/2512.18440v1