From Vibe Coding to Metacognition: An AI-Generated Framework for Data-Driven Student Reflection

Track: "Artificial Intelligence in Education"

Abstract

Conventional digital assessment platforms often conclude the learning process with a simple score, failing to bridge the crucial gap between summative evaluation and formative, metacognitive development. This paper presents a novel framework, implemented in a web-based learning and assessment application, designed to transform digital testing from a mere measurement tool into a powerful engine for data-driven student reflection. Our approach is built on a multi-layered feedback architecture that provides learners with a holistic view of their competencies, enabling them to move from basic self-assessment to profound metacognitive awareness.

The core of the framework is a three-tiered analysis system presented to the user upon test completion. First, on a **macro-level**, it offers a visual breakdown of performance by content topic, allowing students to immediately identify their conceptual strengths and weaknesses. Second, a sophisticated **comparative dimension** benchmarks the individual's performance against the cohort's average, using a methodologically sound approach that considers only the best completed attempts of peers. This comparison is further granulated by overall score and by question difficulty, providing a nuanced context for self-evaluation. The third and deepest layer facilitates analysis based on **cognitive stages** aligned with Bloom's Taxonomy (e.g., Reproduction, Application, Analysis). This enables students to gain insights not only into *what* they know, but *how* they think, identifying potential gaps between knowledge retention and higher-order application skills.

Beyond this analytical core, the platform's design philosophy promotes learner autonomy and ecosystem integration. A "Bring Your Own Tools" (BYOT) approach is realized through extensive export functionalities (e.g., for Anki, Kahoot!), empowering students to transfer learning materials into their preferred personal study environments for sustainable, long-term engagement. Furthermore, we discuss a pragmatic "Bring Your Own AI" (BYOA) strategy for content creation. The project provides highly-engineered, expert-level prompts that enable educators to use any state-of-the-art external Large Language Model (LLM) to generate high-quality, didactically rich, and technically compatible question sets, decoupling the platform from specific AI vendors. Additional scaffolding, such as a real-time pacing helper that trains exam time-management, further supports the development of procedural test-taking skills.

In conclusion, the presented framework offers a comprehensive, replicable model for designing next-generation educational technologies. By integrating multi-dimensional data analysis, contextual benchmarking, and a commitment to open, learner-centric ecosystems, we demonstrate how digital assessment can evolve into a dynamic and integral component of the metacognitive learning cycle, fostering deeper understanding, strategic thinking, and greater learner autonomy.

Keywords: Generative AI, Student-Facing Learning Analytics, Metacognition, Formative Assessment, Privacy by Design, Feedback Literacy, Low-Code Development.

Excursus: Agile Implementation and "Teacher-Led Software Development"

A remarkable aspect of this project is the genesis of the application itself, which outlines a paradigm shift in EdTech towards "Teacher-Led Software Development." The entire multiple-choice test application was conceived and implemented in just a few hours by AI agents (Large Language Models) in a process of "Vibe Coding." This process took place under the expert supervision of a STEM teacher who possessed basic programming knowledge (low-code approach) and steered the development in an intuitive and goal-oriented manner. This approach aligns with recent research on the democratization of software development through AI (Mollick & Mollick, 2023), demonstrating that educators can now create tailored didactic tools rapidly and resource-efficiently, without relying on "black box" commercial solutions.

---

2. The Learning Platform: A System for Fostering Metacognitive Development

2.1. System Architecture and Technological Foundations

The application is designed as a modern web application based on established open-source technologies. Both the backend and frontend are implemented using the Python framework Streamlit, enabling the rapid and interactive development of data-centric user interfaces. Data persistence for user and test data is ensured by a SQLite database, providing a lightweight and serverless storage solution. This technological stack was deliberately chosen to ensure high portability and ease of deployment.

2.2. The Didactic Data Model: Structuring Questions for Deeper Learning

The heart of the platform is a flexible didactic data model defined in JSON format. Each question is more than just text with response options; it is a rich data object containing metadata for multi-layered analysis. Key fields include `topic` for thematic organization, `weighting` to define difficulty levels, and `cognitive_level` for classification according to Bloom's Taxonomy. This structured data capture is the prerequisite for the data-driven reflection detailed in Section 3.

2.3. The Two-Tiered Glossary Concept: From Contextual Hint to Summative Reference

To reduce comprehension barriers and promote vocabulary acquisition, the platform implements an innovative, two-tiered glossary concept:

1. **Just-in-Time Glossary (`mini_glossary`):** Authors can define a `mini_glossary` directly at the question level. Key terms essential for understanding a question are explained immediately in context. This reduces cognitive load, as learners do not need to leave the application to look up terms.
2. **Summative Reference:** Upon completing the test, learners can download a consolidated PDF glossary of the entire question set. This document aggregates all `mini_glossary` entries, serving as a persistent learning artifact for later review.

2.4. Privacy by Design: Anonymity and Enabling Longitudinal Reflection

In response to increasing concerns regarding student data privacy (Drachsler & Greller, 2016), the platform follows a strict "Privacy by Design" approach. Usage is fundamentally anonymous via automatically generated pseudonyms (e.g., names of famous scientists). Registration with personal data is not required. Simultaneously, **longitudinal reflection** (viewing one's development over time) is enabled: Users can "reserve" their anonymous pseudonym using a self-chosen secret word. This allows them to log in across

different sessions and devices to access their personal learning history without compromising anonymity, balancing the need for privacy with the benefits of learning analytics.

---

3. The Framework for Data-Driven Reflection: Implementation and Analysis

The platform's innovation lies in transforming raw test results into actionable insights through a three-tiered analysis system presented to the user. This design follows principles of effective Student-Facing Learning Analytics Dashboards (SFLAD), prioritizing clarity and actionability over data complexity (Jivet et al., 2020).

3.1. Level 1 (Macro-Analysis): Visual Pattern Recognition by Subject Matter

On the first level, learners receive immediate visual feedback on their thematic competencies. Instead of a simple score, the system utilizes **heatmap-like bar charts** that color-code performance per topic block (e.g., "Thermodynamics", "Mechanics") using a traffic-light system (Red-Yellow-Green).

- **Didactic Benefit:** This visualization facilitates rapid pattern recognition. Students can instantly identify whether deficits are isolated or systemic, promoting targeted remediation of specific modules rather than generalized repetition.

3.2. Level 2 (Comparative Analysis): Contextualized Benchmarking and Psychological Safety

Social comparison is a powerful but potentially risky instrument. To prevent demotivation and support feedback literacy (Yan & Carless, 2022), the framework relies on **contextualized fairness**:

- **Best-Attempt Logic:** The calculation of the cohort average considers only the *best* completed attempts of peers. This prevents "fun runs" or aborted tests from skewing statistics, setting a high but achievable standard.
- **Granularity:** Comparison occurs not just globally, but differentiated by difficulty level. A student might realize: "I am average overall, but I perform above average on difficult questions." This fosters a nuanced self-image beyond binary "good" or "bad" evaluations.

3.3. Level 3 (Cognitive Analysis): Radar Charts for Diagnosing Depth of Thought

The deepest level of analysis leverages the metadata of Bloom's Taxonomy. Results are visualized in a **Radar Chart (Spider Web Diagram)** spanning three axes: *Knowledge/Reproduction*, *Application*, and *Analysis/Transfer*.

- **Metacognitive Insight:** This visualization uncovers discrepancies between rote memorization and understanding. A common pattern revealed is "binge learning" (high scores in reproduction, low in analysis). By visualizing this imbalance, learners are encouraged to shift their learning strategies from pure memorization toward deep conceptual understanding, facilitating self-regulated learning (Matcha et al., 2019).

3.4. Scaffolding Tools: The Real-Time Pacing Helper

Complementing content reflection, the system offers procedural scaffolding. A **Real-time Pacing Helper** visually indicates during the test whether the learner is "on track" compared to the recommended completion time. This trains time management skills and reduces exam anxiety through transparency.

---

4. Discussion: Broader Pedagogical Implications

4.1. Fostering Learner Autonomy (BYOT)

The "Bring Your Own Tools" philosophy prevents the "walled garden" effect. Through export features for Anki and Kahoot!, as well as PDF answer keys, the platform acts as a feeder for the student's personal learning ecosystem. This supports the development of sustainable study habits outside the classroom environment.

4.2. A Pragmatic AI Approach (BYOA): Didactic Prompting

The approach decouples didactics from technology. Instead of integrating a "Black Box" AI, the project provides **expert-validated prompts** that enable educators to use *any* state-of-the-art LLM to generate high-quality, JSON-compatible questions including glossaries. This empowers educators to harness the potential of Generative AI (Kasneci et al., 2023) while maintaining didactic control and ensuring the quality of the material.

4.3. Enhancing Motivation through Gamification

Subtle gamification elements, such as scientific pseudonyms with accessible biographies and a public **Leaderboard** (based on points and time), create engagement incentives without distracting from the core goal of formative reflection.

4.4. Synthesis: From Data to Data Literacy

The system implicitly fosters learners' data literacy. By learning to interpret their own performance data (radar charts, distributions), they become experts in their own learning process. This shift from external assessment to data-driven self-regulation is the core of metacognitive maturity (Winstone & Carless, 2019).

---

5. Conclusion and Outlook

5.1. Summary of Contributions

This paper presented a framework delivering two primary contributions: First, a technical model for rapid, AI-assisted software creation ("Vibe Coding"), and second, a pedagogical model that transforms assessment into learning opportunities through multi-dimensional data visualization.

5.2. Scalability and Standards

The presented JSON data model has the potential to serve as a lightweight exchange format for OER (Open Educational Resources). Unlike complex standards such as QTI, it is human-readable and AI-optimized, significantly lowering the barrier for sharing educational materials.

5.3. Outlook: Adaptive Learning Paths

Future iterations of the platform will utilize the data gained from Cognitive Analysis (Level 3) to not only provide feedback but to actively support learning. The roadmap includes generating **AI-based individual study plans** that, based on weaknesses identified in the Radar Chart, proactively suggest specific

exercises or explanations (Bodily et al., 2018). This effectively closes the loop from assessment back to instructional support.

References

Bodily, R., Nyland, R., & Wiley, D. (2018). The Rise of Student-Facing Learning Analytics Systems. *International Conference on Learning Analytics & Knowledge*.

Drachsler, H., & Greller, W. (2016). Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 89–98.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2020). Repository of student-facing learning analytics dashboard design principles. *Proceedings of the 10th International Conference on Learning Analytics & Knowledge*.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

Matcha, W., Gašević, D., & Pardo, A. (2019). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226-245.

Mollick, E., & Mollick, L. (2023). Assigning AI: Seven Approaches for Students, with Prompts. *SSRN Electronic Journal*.

Winstone, N. E., & Carless, D. (2019). *Designing Effective Feedback Processes in Higher Education: A Literacy Perspective*. Routledge.

Yan, Z., & Carless, D. (2022). Self-assessment is about more than self: the enabling role of feedback literacy. *Assessment & Evaluation in Higher Education*, 47(7), 1116-1128.