

InsPIRE Insight

Learning Analytics for Assessment Preparation: Constructing Graphs to Guide Higher-order Question Generation

By

Joel Weijia Lai, Shen Yong Ho & Fun Siong Lim

Centre for the Applications for Teaching & Learning Analytics for Students (ATLAS)

under the Institute for Pedagogical Innovation, Research & Excellence

Nanyang Technological University

Singapore

KEY IMPLICATIONS

- Learning analytics can be used to transform assessment data into structures that guide AI toward generating higher-order, cognitively demanding questions.
- Combining performance-weighted topic graphs with large language models (LLMs) enhances the conceptual alignment and difficulty calibration of AI-generated questions.
- The approach provides educators with scalable tools to produce conceptually rich assessment items aligned to the SOLO taxonomy's relational level.

BACKGROUND

High-quality assessment design in higher education requires questions that challenge learners to integrate and apply concepts, not just recall facts. However, most automated question-generation systems, even those powered by large language models, struggle to produce questions that achieve higher-order cognitive engagement. The study addresses this gap by leveraging learning analytics and graph-based modeling to guide LLMs in producing assessment items of greater conceptual depth.

FOCUS OF INITIATIVE

This initiative introduces a learning analytics pipeline that transforms archival assessment data and student performance into a topic-performance graph—an undirected, weighted graph linking curricular topics based on how often they co-occur in assessments and how students perform on them. The framework integrates:

- Topic tagging: Using both human and AI (GPT-5 Thinking, Gemini 2.5 Pro) raters to label assessment questions with curricular topics.
- Graph construction: Connecting topics through co-occurrence and correlation of performance scores to quantify conceptual relationships and difficulty.
- Question generation: Using GPT-5 Thinking to generate new assessment questions based on topic clusters derived from the graph, thus grounding LLM generation in empirical patterns of student learning and topic relationships.

The aim is to scaffold AI-generated questions that are curriculum-aligned, difficulty-calibrated, and cognitively demanding, particularly at the relational level of the SOLO taxonomy.

KEY OUTCOMES

- Reliability and Model Selection: GPT-5 Thinking achieved stronger inter-rater reliability with human experts ($\alpha = 0.641$ at depth 3) than Gemini 2.5 Pro, validating its use for topic tagging.
- Graph Construction: A topic-performance graph was built with 30 nodes and 53 edges, revealing six conceptual communities within the physics curriculum, showing how topics coalesce through learning and assessment.
- Question Generation Results: When GPT-5 Thinking generated questions using topic clusters derived from the graph, human evaluators rated these as more difficult and conceptually integrated (Relational-level) than those from random topic selection.

SIGNIFICANCE OF OUTCOMES

Implications for practice: The methodology empowers educators to use data-informed assessment design tools. By grounding AI generation in performance-linked topic structures, teachers can create more meaningful assessments that reflect authentic student learning challenges.

Proposed follow-up activities:

1. Extend the dataset across multiple academic years and disciplines to generalize the framework.
2. Integrate multimodal question generation (e.g., diagrams and simulations) to enhance interpretability.
3. Calibrate graph-derived difficulty metrics with expert benchmarks to improve reliability and pedagogical meaning.

PARTICIPANTS/SCOPE

The project analyzed 1,131 undergraduate students from an engineering physics course at NTU, using performance data from two midterm examinations comprising 37 assessment items. It involved one human subject expert and two advanced reasoning LLMs in the topic-tagging process.

METHODOLOGY/APPROACH

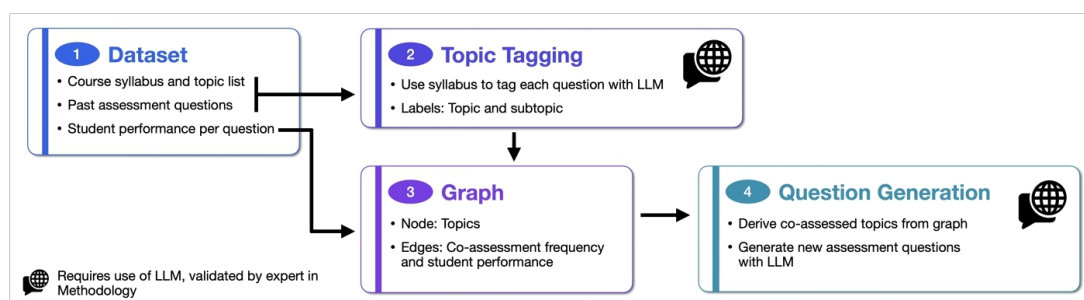


Figure 1: Overview of generic methodology.

The methodology combined learning analytics, graph theory, and generative AI:

- Construction of a topic-performance graph integrating student scores and topic co-occurrence patterns.
- Use of graph-search algorithms to identify cognitively rich and empirically difficult topic clusters.
- Evaluation of LLM-generated questions through blinded expert review for both difficulty and SOLO taxonomy level.

The hybrid approach bridges analytic rigor and educational design, demonstrating how AI can augment human judgment in assessment creation.

References

GitHub Repository for topic-performance-graph-question-generation.

<https://github.com/joellaiwj/topic-performance-graph-question-generation>

Nicy Scaria, Suma Dharani Chenna, and Deepak Subramani. 2024. Automated Educational Question Generation at Different Bloom's Skill Levels Using Large Language Models: Strategies and Evaluation. Springer Nature Switzerland, Cham, Switzerland, 165–179. https://doi.org/10.1007/978-3-031-64299-9_12

Link to full article (available from 27 April 2026): <https://doi.org/10.1145/3785022.3785091>

Contact Joel Weijia Lai at joellai@ntu.edu.sg for a pre-print or for more information about the initiative.