Majumdar, R. et al. (Eds.) (2025). Proceedings of the 1st International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education

Measuring Semantic Fidelity in Student Discussion Posts

Michelle BANAWAN
Asian Institute of Management, Philippines
mbanawan@aim.edu

Abstract: Asynchronous discussion boards are central to many online and blended learning environments, yet assessing the quality of student contributions remains a challenge—particularly in capturing both literal adherence to prompts and deeper conceptual engagement. This paper introduces a layered framework for measuring semantic fidelity in student discussion posts, defined as the extent to which a post aligns with both the instructional prompt (literal fidelity) and the broader course themes (inferential fidelity). Using natural language processing (NLP), we operationalize this framework through two computational lenses: semantic similarity based on sentence-level embeddings to measure prompt adherence, and zero-shot topic modeling using `RoBERTa-large-MNLI` model to assess alignment with course themes. A two-dimensional semantic fidelity quadrant is used to visualize and classify student posts based on their scores across these dimensions. We applied this approach to discussion data from 88 students (44 undergraduate and 44 graduate) across a data-focused undergraduate course and a development-focused graduate cohort. The results revealed clear differences in engagement profiles. Undergraduate posts tended to score higher on literal prompt similarity but lower on inferential alignment, often restating questions without integrating broader concepts. In contrast, graduate student responses were more evenly distributed, with many demonstrating conceptual alignment despite diverging from prompt wording—highlighting a more reflective and applied discourse style. This analytic framework offers educators a scalable and interpretable tool to better understand student thinking, surface off-prompt yet pedagogically rich contributions, and inform more equitable feedback practices. It also advances the field of trustworthy learning analytics by balancing automated insight with human pedagogical goals. The study concludes with implications for instructional design, recommendations for adaptive scoring thresholds, and directions for extending the framework to multi-turn discussions and real-time learning dashboards.

Keywords: semantic fidelity, topic modeling, discussion board analysis, learning analytics

1. Introduction

Student engagement is increasingly assessed not just by participation, but by the quality and relevance of contributions to asynchronous platforms such as discussion boards. This paper introduces a layered framework for measuring semantic fidelity in student posts, combining prompt-level semantic similarity with zero-shot topic modeling to evaluate alignment with broader course themes. Building on prior work (e.g., Banawan et al., 2024) demonstrating the use of NLP pipelines for classroom participation assessment, this study extends the approach to asynchronous discourse and contributes to ongoing research on semantic alignment and topic classification in education.

1.1 Related Work: Semantic Fidelity in Educational Assessments

Embedding models such as word2vec, BERT, and sentence-transformers (Reimers & Gurevych, 2019) have become standard for measuring semantic similarity, with successful applications in educational contexts like open-ended scoring (Wilson et al., 2022) and conceptual understanding assessment (Somers et al., 2021). However, these models often

miss inferential alignment—when students extrapolate meaningfully from lessons without literal overlap. To address this, zero-shot topic modeling tools such as BERTopic (Grootendorst, 2022) and RoBERTa-large-MNLI are used to detect latent topical relevance, offering a proxy for deeper conceptual engagement.

1.2 Current Study

This study introduces a layered framework for assessing semantic fidelity in student discussion posts, defined as alignment with both the instructional prompt and broader course themes. The aim was to examine differences in engagement patterns across academic levels and to support scalable, interpretable feedback beyond participation metrics or keyword matching.

2. Methods

2.1 Participants

The study involved a total of 88 students, equally divided between two academic levels, undergraduate and graduate. The inclusion of two distinct academic profiles—undergraduate learners and professionally experienced graduate students—allowed for comparative insights into patterns of semantic fidelity across different levels of academic and real-world experience. All discussion board posts were collected from a learning management system where students engaged with instructor-designed prompts over the course of a term. Posts were exported, de-identified, and grouped by academic level (undergraduate or graduate).

2.1.1 Semantic Similarity or Literal Fidelity

To measure literal fidelity, we used sentence-level embeddings generated via a MiniLM transformer model (via sentence-transformers). Cosine similarity scores were computed between the prompt and each student's full post. These scores indicate how closely the content of the post matched the language and ideas in the original prompt. Low-alignment posts are flagged as such when cosine similarity dropped below 0.6, indicating potential deviation from the instructional intent.

2.1.2 Inferential Fidelity through Topic Alignment (Zero-Shot Topic Modeling)

To evaluate whether a post demonstrated inferential alignment with broader course themes, even when not lexically close to the prompt, zero-shot topic modeling was implemented. Using a pre-trained entailment model (`roberta-large-mnli`), each student post was evaluated against a curated list of instructional themes and learning concepts derived from the course syllabus and lesson objectives.

2.2 Analysis

To evaluate the semantic fidelity of student discussion posts, we conducted a two-dimensional analysis based on (1) semantic similarity to the instructional prompt and (2) inferential alignment with course themes. Each post was scored along these two dimensions using natural language processing techniques: cosine similarity from sentence-level embeddings for literal prompt adherence, and zero-shot topic modeling for broader conceptual relevance. Both dimensions were evaluated on a five-point ordinal scale, where 1 indicated very low alignment and 5 indicated very high alignment.

To assess semantic fidelity, each student post was evaluated along two dimensions: prompt fidelity and topical relevance. Prompt fidelity was measured using cosine similarity between the post and its associated prompt, based on MiniLM sentence embeddings. Scores were mapped to a 5-point scale: ≥0.80 (score 5), 0.70–0.79 (4), 0.60–0.69 (3), 0.50–0.59 (2), and <0.50 (1, indicating semantic drift). Topical relevance was assessed using zero-shot classification with a RoBERTa-large-MNLI model, comparing each post to curated course themes. Confidence scores were similarly scaled: ≥0.85 (5), 0.75–0.84 (4), 0.60–0.74 (3), 0.45–0.59 (2), and <0.45 (1, minimal alignment). This dual scoring approach captured both

literal and inferential alignment, enabling recognition of surface-level adherence and deeper, conceptually relevant engagement.

3. Results

Student posts were assessed based on their alignment with both the discussion prompt and the broader course topic. Undergraduate responses largely reflected high lexical similarity to the prompt but showed limited topical engagement, often restating questions with surface-level examples. In contrast, graduate responses demonstrated greater variation, with many showing strong conceptual alignment even when they did not closely follow the prompt wording. This pattern suggests that graduate students were more likely to infer, extrapolate, or integrate course material, while undergraduates tended to remain literal and prompt-bound. Only a small portion of responses lacked both prompt and topic alignment, indicating that most students made a conscious effort to connect their posts to the course, either structurally or conceptually.

4. Conclusion

This study proposed a layered framework for assessing semantic fidelity in discussion board posts by combining prompt-level similarity with inferential alignment to course themes. Through embedding-based semantic similarity and zero-shot topic classification, we captured both surface-level and conceptual dimensions of student engagement. The findings revealed that undergraduate responses often reflected literal prompt adherence, while graduate posts showed deeper conceptual integration-even when diverging from prompt wording. These patterns underscore the importance of evaluating both linguistic and thematic alignment in student discourse. The framework supports trustworthy learning analytics by identifying meaningful engagement beyond superficial metrics and subjective judgments of student posts. While effective, the framework has limitations: the semantic similarity threshold was contextually calibrated and may not generalize, and the topic classification relied on a manually curated theme set. Additionally, the analysis focused only on initial posts, excluding replies and multi-turn interactions that could reveal evolving student thinking.

References

Banawan, M., Kukas, E. J., Tano, J. R., Tan, A., & Villegas, R. (2024). Transformative Approach to Fairness and Transparency in Class Participation Assessment. In Proceedings of the Eleventh ACM Conference on Learning@Scale (L@S '24). https://doi.org/10.1145/3657604.3664708

Grootendorst, M. (2022). BERTopic: Neural topic modeling with class-based TF-IDF and BERT embeddings. arXiv preprint arXiv:2203.05794.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing.

Wilson, J., Pollard, B., Aiken, J. M., Caballero, M. D., & Lewandowski, H. J. (2022). Classification of open-ended responses to a research-based assessment using natural language processing. *Physical Review Physics Education Research*, *18*(1), 010141.