

15 - 17 May 2026

Rome, Italy

Development of Automated Discourse Analysis System (ADAS) for Middle & High School Science Classroom Discourse Feedback

Dr. Soon Lee

Kennesaw State University, the United States

Abstract

Teacher feedback on classroom discourse, a proven driver of inquiry-based science learning, requires 3–4 hours of expert coding per lesson, making it inaccessible at scale. NLP-powered automated analysis offers a potential solution, but the conditions under which these systems reliably capture the specific discourse moves central to science learning remain unclear. This matters because dialogic discourse produces substantially stronger learning gains ($d = 0.4$ – 0.8) than transmission-oriented instruction, with particularly pronounced effects for students from low-income backgrounds and students of color, yet most teachers, especially those in under-resourced schools, lack access to expert feedback needed to develop these practices. This study compared four NLP approaches for classifying science classroom discourse: BERT fine-tuning, RoBERTa fine-tuning, GPT-4o feature extraction, and GPT-4o zero-shot prompting. We evaluated these approaches across two classification tasks, Utterance Type (10 categories) and Reasoning Component (6 categories), using cross-context design training on physics and testing on electricity instruction. Results revealed task-dependent performance: GPT-4o embeddings achieved highest Utterance Type accuracy (67.1%, $\kappa = 0.600$), while RoBERTa performed best for Reasoning Component (63.1%, $\kappa = 0.379$). Zero-shot prompting dramatically underperformed (28.7% and 47.8%). Most critically, all approaches exhibited “epistemic collapse,” systematic failure on student-initiated inquiry ($F1 = 0.00$ – 0.07), higher-order reasoning ($F1 = 0.00$), and nuanced feedback ($F1 < 0.23$). These failures were structured: AI reliably identified knowledge reproduction while becoming blind to knowledge construction, students reasoning from evidence and generating explanations, the distinction defining effective science teaching.

Keywords: Automated Discourse Analysis, Science Education, Machine Learning, Classroom Discourse, Teacher Feedback