Automating the Measurement and Assessment of Classroom Discourse
Sidney K. D’Mello1, Andrew M. Olney2, Nathan Blanchard1, Xiaoyi Sun3, Brooke Ward3, Borhan Samei2, Sean Kelly4, Patrick Donnelly1 & Martin Nystrand3
1 University of Notre Dame, 2 University of Memphis, 3 University of Wisconsin-Madison, 4 University of Pittsburgh

Abstract
Dialogic instruction, a style of classroom discourse, is a powerful predictor of student achievement in middle and high-school classrooms. In studies of dialogic instruction involving thousands of students across hundreds of classes, time spent on open discussions and the proportion of teacher authentic questions and questions with uptake were positive predictors on achievement. Unfortunately, the research is difficult to scale because dialogic instruction must be manually coded by trained observers. This project focuses on automatically analyzing classroom discourse using multimodal sensing, automatic speech recognition, natural language understanding, and machine learning. We developed an approach to record audio from classrooms to identify teacher versus student speech and automatically transcribe teacher speech. Ongoing research aims to autonomously code instructional activities (e.g., lecture, discussion), detect teacher questions, and analyze those questions with respect to authenticity and uptake. This approach provides quantifiable variables for teacher education, professional development, and classroom research and is generalizable, allowing other researchers to quantify additional variables of interest.

Overview
Dialogic instruction is a powerful predictor of student achievement:

- **Discussion**: Open exchange of ideas among at least three participants lasting longer than 30 seconds
- **Authentic teacher questions**: Open-ended questions without a pre- scripted answer rather than known-answer test questions
- **Uptake**: Speaker’s incorporation of a previous answer into a subsequent question (e.g., “what did you mean by that?”)

This project focuses on automatic analysis of dialogic instruction

Data Collection

- Teachers wore a headset microphone to capture high fidelity speech
- Classrooms were recorded with a pressure zone microphone
- Dialogic instruction was annotated by trained observers
- 77 sessions collected, totaling 780 hours, averaging 60±20 minutes each session, from 11 teachers, covering 1,600 instructional segments

Technological Approach

- **Identification of teacher speech**
  - F1 = 0.97
- **Identification of student speech**
  - F1 = 0.66
- **Transcription of teacher speech**
  - Accuracy = 69%
- **Identification of Q&A segments**
  - F1 = 0.73
- **Identification of individual questions**
  - F1 = 0.56

Results (work in progress)

- Classroom recordings are analyzed for temporal, acoustic, spectral, and natural language features.
- Machine learning models are trained to identify important components of dialogic instruction and are validated independently of the teacher.

- **Identification of teacher speech**
  - F1 = 0.97
- **Identification of student speech**
  - F1 = 0.66
- **Transcription of teacher speech**
  - Accuracy = 69%
- **Identification of Q&A segments**
  - F1 = 0.73
- **Identification of individual questions**
  - F1 = 0.56

F1 score is a measure balancing precision (relevance) and recall (specificity)

Next Steps

- Classification of question properties (authenticity and uptake)
- Assessing whether automatic coded dialogic instruction predicts student achievement similar to manual coding
- Dissemination of software/hardware to teachers, researchers, and professional development specialists

This research was supported by the Institute of Education Sciences (R305A130030)