SpeakWise: Capturing Real-Time Reasoning Evidence Traces through AI-Mediated Oral Assessment

Jewoong Moon (The University of Alabama) · **Yong Ju Jung** (The University of Oklahoma)

ICOLSEI 2026
Seoul National University · Jun 26–27

Poster Presentation · Work-in-Progress

1 The Problem

Generative AI blurs the line between a learner's own reasoning and AI-assisted production, threatening **assessment validity** for open-ended tasks.

Oral assessment forces real-time articulation (Fenton, 2025), but conventional oral exams **do not scale** and do not capture reasoning as **analyzable evidence**.

2 Approach

SpeakWise is a research-informed intervention grounded in **Evidence-Centered Design** (Mislevy et al., 2003) and assessment design & validation (NRC, 2014), reframing oral assessment as a **process-oriented measurement environment** (cf. Zapata-Rivera et al., 2024).

- Evidence-Centered Design
- Process data
- Human-centred LA

3 Research Questions

- **RQ1** What reasoning evidence traces emerge, and how do they vary across learners and tasks?
- **RQ2** How do dialogic & temporal patterns shape instructors' interpretation of reasoning?
- **RQ3** How do instructors use evidence traces & AI confidence in evaluative judgments?

4 Method

Work-in-progress **design-based research**, synthesizing:

- interaction-log analysis
- automated discourse-pattern detection
- instructor sensemaking interviews

Real-time capture across domains. **Early user feedback (4 informal-learning reviewers)** now feeds the next design iteration.

5 The SpeakWise System

Learners engage in **adaptive oral dialogues** with an AI examiner (Socratic questioning + **tiered scaffolding**). The platform is structured as a four-stage, ECD-aligned pipeline:

1. **Oral Dialogue**
   Student ⇄ AI examiner · voice · scaffolding
2. **Evidence Capture**
   discourse · dialogic · temporal · structure
3. **Reasoning Analysis**
   4-dim rubric + argument graph + AI confidence
4. **Instructor Sensemaking**
   dashboard review · evidence-linked override

Fig 1. SpeakWise system architecture — an Evidence-Centered Design pipeline from oral dialogue to instructor decision.

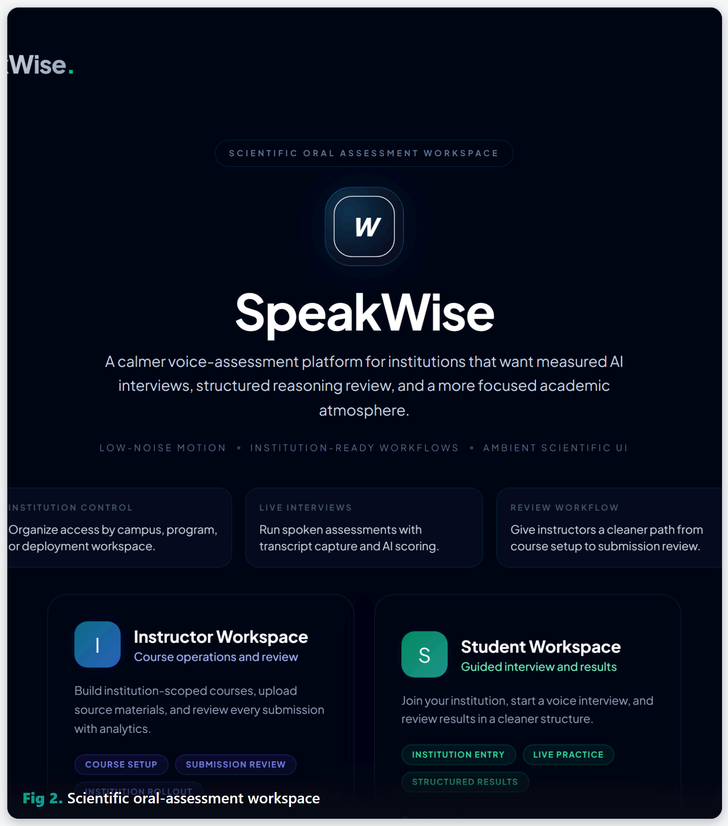

Fig 2. Scientific oral-assessment workspace

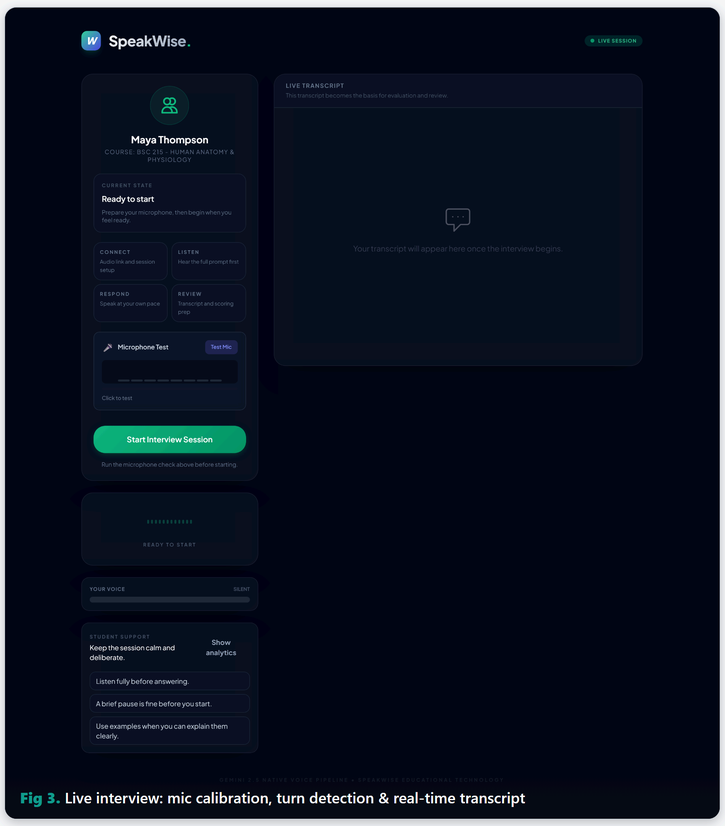

Fig 3. Live interview: mic calibration, turn detection & real-time transcript

6 Reasoning Evidence Traces

The core construct: SpeakWise surfaces **multiple forms of reasoning-relevant evidence** embedded in the interaction (He & Cui, 2025).

- **a Justification & causal markers** ("because", "therefore")
- **b Counter-argument & abstraction** — challenges, generalizing
- **c Dialogic behaviors** — turn initiation, rephrasing
- **d Temporal markers** — response-latency variation

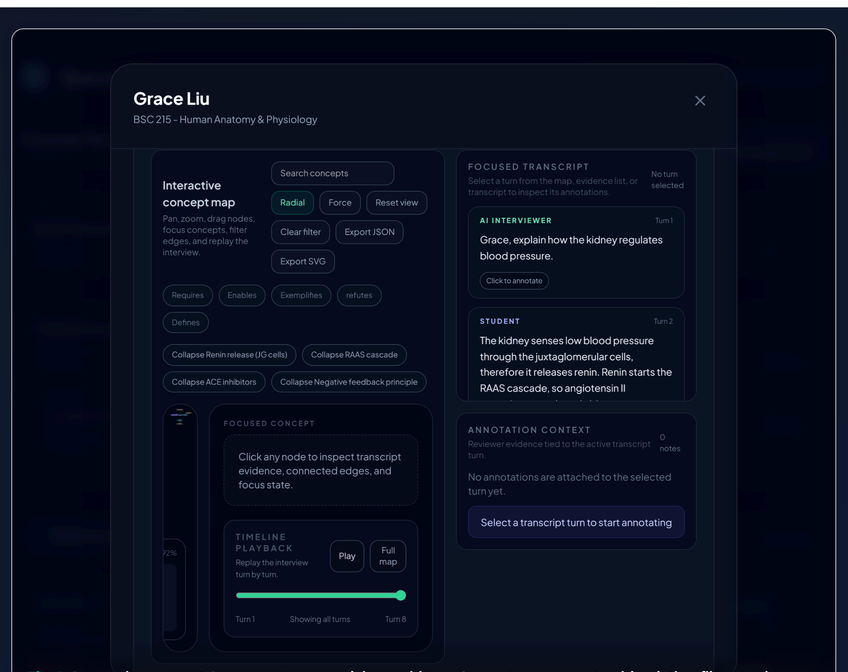

Fig 4. Interactive concept / argument map — claims, evidence & counter-arguments with relation filters and timeline playback

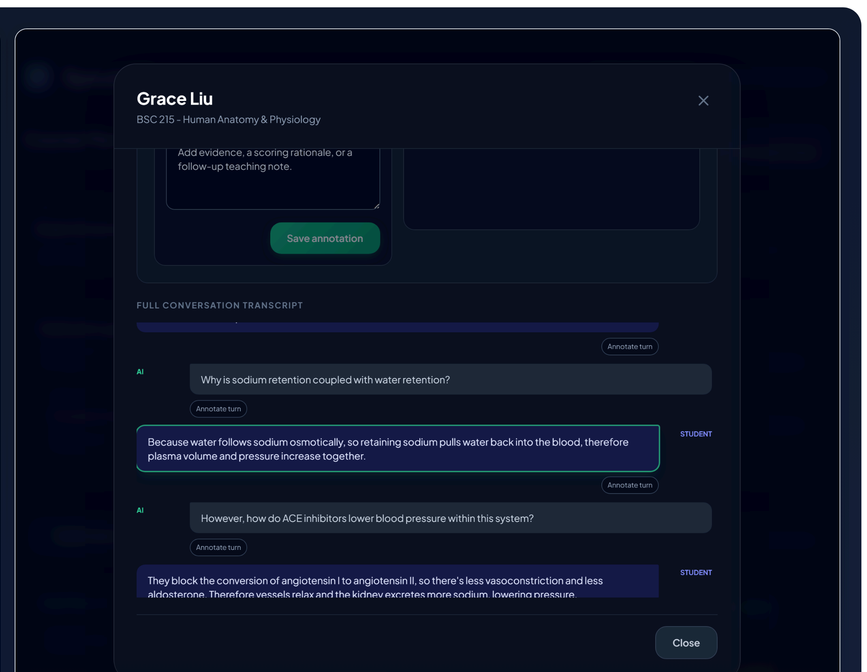

Fig 5. Transcript with captured causal markers ("therefore … releases renin")

7 Instructor Sensemaking

Automated scores are **provisional summaries, not verdicts** (Alfredo et al., 2024). The instructor surface preserves **agency** by showing not only the score, but how it was generated:

- **Confidence-coded feed** — green / yellow / red mastery across submissions
- **Attention flags** — courses & students needing first review
- **Evidence drill-down** — argument graph + 4-dimension reasoning rubric per student
- **Transcript-linked annotations** tied to specific reasoning turns
- **Score override** with instructor notes & audit trail

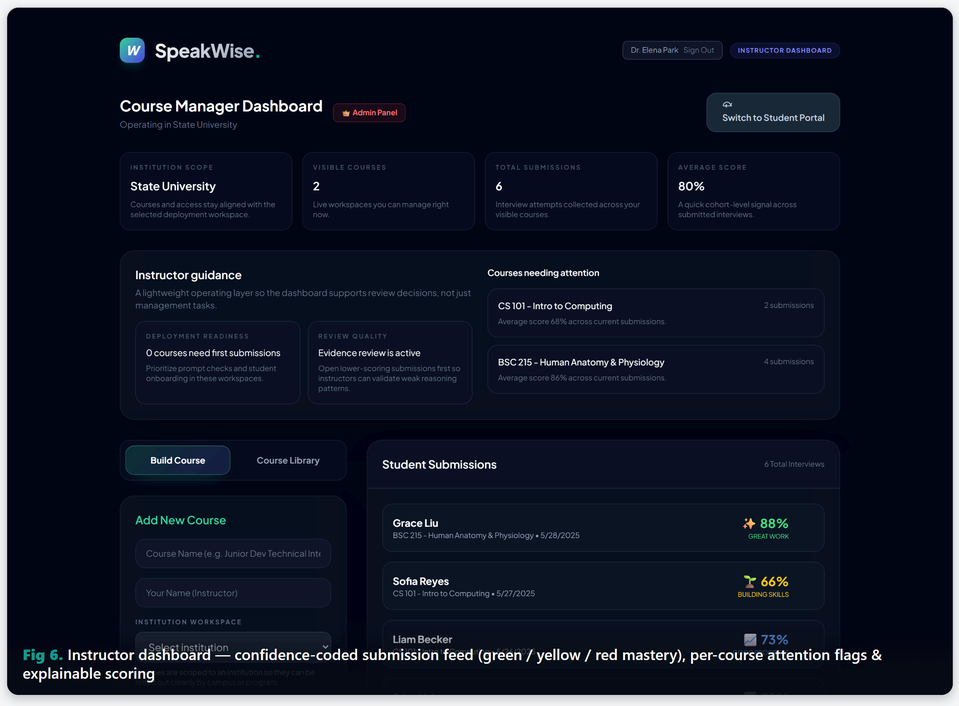

Fig 6. Instructor dashboard — confidence-coded submission feed (green / yellow / red mastery), per-course attention flags & explainable scoring

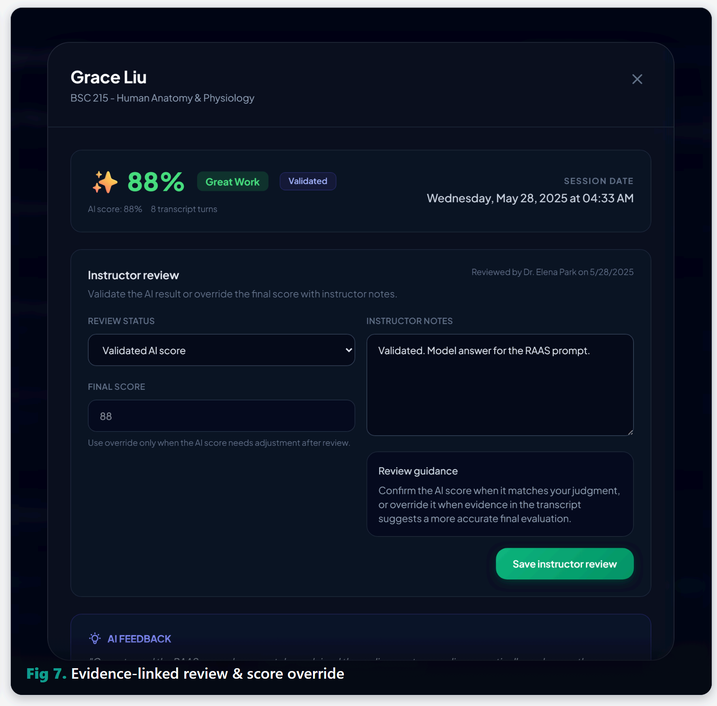

Fig 7. Evidence-linked review & score override

8 Contribution & Next Steps

- **Concept** **Reasoning evidence traces** as a new measurement target for oral assessment, beyond a single holistic score.
- **Method** Real-time capture & fusion of **discourse, dialogic, and temporal** signals into an analyzable argument structure.
- **Practice** An **institution-ready, explainable, human-in-the-loop** assessment environment that keeps instructors in control.

From early user feedback (N = 4): student-facing transparency & explainable scoring · a calm, low-pressure practice mode (prompt preview, pacing, retries) · input robustness & accessibility (ASR + type fallback, multiple voices). **Next:** in-the-wild deployment (UA / OU) · rubric inter-rater reliability · AI-confidence calibration.

References

Alfredo, R., et al. (2024). Human-centred learning analytics and AI in education. *Computers & Education: AI, 6*, 100215.

Fenton, A. (2025). Reconsidering the use of oral exams and assessments. *Educational Researcher, 54*(7), 430–436.

He, S., & Cui, Y. (2025). A systematic review of log-based process data in computer-based assessments. *Computers & Education*, 105245.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS RR, 2003*(1).

National Research Council. (2014). *Developing Assessments for the NGSS*. National Academies Press.

Zapata-Rivera, D., et al. (2024). Designing ECD-based conversations for assessment with LLMs. *EDM 2024 Workshop*.

SpeakWise · AI-mediated oral conversational assessment | **jmoon19@ua.edu** · yongju@ou.edu

Full project page
workflow · architecture · demo videos
educatian.github.io/speakwise

4 evidence-trace types
3 research questions
ECD theoretical anchor