Koh, E., et al. (Eds.) (2026). Proceedings of the 2nd International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education.

# Design and Preliminary Evaluation of an AI-Supported Role Fulfillment System: A Case Study on Delayed Skill Transfer in Collaborative Learning

**Hirotsune TAKAHASHI [a*], Yasuhisa TAMURA [a]**
*Faculty of Science and Technology, Sophia University, Japan*
*tsune0711.main@gmail.com

**Abstract:** While Computer-Supported Collaborative Learning (CSCL) systems effectively structure group interactions, their impact on long-term skill internalization remains underexplored. This exploratory case study examines an AI-supported role fulfillment system that provides real-time, role-specific feedback to individual learners via a low-latency streaming pipeline and LLM-based analysis. Using a within-subjects reversal design (Pre-Test, AI-supported Test, Post-Test) with four university students, we observed two complementary transfer effects: a delayed equalization in participation (44.2% reduction in word-based Gini coefficient only after AI removal) and sustained role fulfillment (+19.3% during intervention, +17.0% maintained post-intervention). These temporally dissociated patterns suggest that a single brief AI intervention can simultaneously support implicit norm acquisition and explicit skill development, though observable effects emerge on different timescales. While limited by sample size, these preliminary findings highlight the importance of post-intervention assessment in CSCL research.

**Keywords:** Collaborative Learning, CSCL, AI Agent, Role Fulfillment, Skill Transfer, LLM

## 1. Introduction

Collaborative learning is essential for deep learning processes such as argumentation, negotiation, and knowledge co-construction (Dillenbourg, 1999). While CSCL scripts effectively structure group interactions (Fischer et al., 2013; Strijbos & De Laat, 2010), a critical gap persists: most systems demonstrate effectiveness only during active intervention, with limited evidence of sustained skill transfer after support removal. Analytics-based feedback improves group performance during intervention (Zheng et al., 2022), yet the trajectory of these skills post-intervention—whether they persist, decay, or manifest with delay—remains underexplored.

This exploratory case study addresses two questions: (RQ1) Does dynamic AI intervention promote behavior aligned with participants' assigned roles during the intervention phase? (RQ2) Does dynamic AI intervention promote skill transfer in role fulfillment capabilities that persists after the removal of AI support? To investigate these, we designed an AI-supported role fulfillment system and evaluated it using a within-subjects reversal design with four university students.

## 2. Method

### *2.1 System Architecture*

We developed a real-time AI-supported system (Figure 1) comprising three core components: (1) a low-latency streaming pipeline using a meeting bot (Recall.ai) that joins

the Google Meet session, captures audio, and produces speaker-labeled transcripts; (2) a context-aware LLM analysis module using GPT-4-turbo that generates role-specific advice based on participant metadata, recent discussion history, and real-time participation statistics; and (3) a role-based targeted delivery mechanism that routes personalized feedback to individual participants' Slack channels based on their assigned roles.

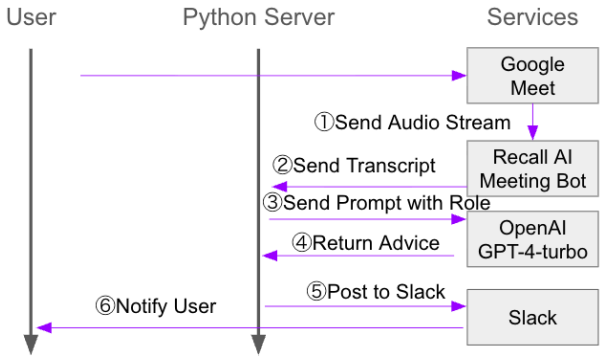


Figure 1. System Architecture

### *2.2 Experimental Design*

We adopted a reversal design with three phases: Pre-Test (10 minutes, no AI intervention), Test (10 minutes, AI-supported), and Post-Test (10 minutes, no AI intervention). Four university students (same group across all sessions) discussed different topics in each phase: a desert island survival scenario (Pre-Test), optimal approaches to group work (Test), and school festival planning (Post-Test). Topics were designed to be comparable in complexity and to elicit diverse discussion strategies.

Participants were randomly assigned one of four roles: Facilitator (manages discussion flow and turn-taking), Idea Provider (proposes new perspectives), Critical Examiner (evaluates proposals and identifies risks), and Consensus Builder (synthesizes viewpoints toward agreement). Role assignments were explained verbally with descriptions of each role's responsibilities. These roles were designed to represent functionally distinct participation patterns that collectively support balanced collaborative learning.

### *2.3 Evaluation Metrics*

We employed three complementary metrics. First, the Gini coefficient was calculated for utterance counts, word counts, and character counts to assess participation equality (lower values indicate more equal participation). Second, the Role Fulfillment Index (RFI) was computed via post-hoc LLM analysis using gpt-5-mini, which evaluated each utterance's alignment with the speaker's assigned role on a 0.0–1.0 scale. To ensure scoring reliability, we conducted three independent evaluation runs and report mean values (SD across runs: 0.002–0.009). The evaluation prompt instructed the model to assess whether each utterance demonstrated behaviors characteristic of the assigned role (e.g., facilitating turn-taking for the Facilitator role, identifying risks for the Critical Examiner role). Third, a 4-point Likert scale measured learners' self-efficacy in role fulfillment.

## 3. Results

The system processed a total of 205 utterances in real-time across three sessions. We present results organized by research question.

### *3.1 RQ1: Behavioral Changes During AI Intervention*

Analysis of participation equality indicated that AI intervention did not produce immediate improvements. As shown in Table 1, the Gini coefficients for word and character counts slightly increased during the Test session (+6.7% and +13.2%), suggesting marginally less equal participation during active intervention. In contrast, the RFI increased by 19.3% from Pre-Test (M=0.257) to Test (M=0.307), suggesting that participants successfully responded to real-time AI prompts regarding role-specific behaviors even while participation balance remained uneven.

Table 1. *Gini Coefficients Across Sessions*

| **Metric** | **Pre-Test** | **Test (AI)** | **Post-Test** | **Pre→Test** | **Test→Post** |
|---|---|---|---|---|---|
| Total Utterances(n) | 76 | 64 | 65 | | |
| Utterances(Gini) | 0.263 | 0.273 | 0.342 | +3.9% | +25.2% |
| Words(Gini) | 0.152 | 0.162 | 0.090 | +6.7% | -44.2% |
| Characters(Gini) | 0.153 | 0.173 | 0.097 | +13.2% | -43.9% |

### *3.2 RQ2: Evidence of Skill Transfer*

The most notable finding occurred in the post-intervention phase. While participation equality did not improve during the Test session, the Post-Test showed a substantial reduction in inequality: the word-based Gini coefficient dropped by 44.2% compared to the Test session (Table 1), suggesting a delayed equalization effect where the desired norm manifested only after support removal.

Notably, utterance-count Gini increased (+25.2%) while word- and character-based Gini decreased sharply (−44.2%). This divergence is explained by Table 2: the Facilitator contributed 49.2% of utterances but only 24.3% of words, indicating many short turn-management utterances, while other roles produced fewer but substantially longer contributions. This pattern suggests that equalization occurred at the level of substantive contribution rather than turn frequency.

Table 2. *Speaker Contribution Ratios (Post-Test)*

| **Speaker** | **Utterance Ratio** | **Word Ratio** | **Character Ratio** |
|---|---|---|---|
| Facilitator | 32 (49.2%) | 299 (24.3%) | 670 (24.2%) |
| Idea Provider | 8 (12.3%) | 287 (23.3%) | 631 (22.8%) |
| Critical Examiner | 6 (9.2%) | 250 (20.3%) | 563 (20.3%) |
| Consensus Builder | 19 (29.2%) | 394 (32.0%) | 908 (32.8%) |

Crucially, the Post-Test RFI (M=0.301) remained nearly as high as the Test phase (M=0.307), representing a 17.0% improvement over the Pre-Test baseline (M=0.257). This sustained elevation suggests that participants internalized role-aligned behaviors during AI intervention and maintained them after support removal. Meanwhile, subjective self-efficacy increased 67% (M=2.25 to M=3.75), peaking alongside the delayed improvement in participation equality.

Qualitative evidence supports these quantitative patterns. In the Post-Test, the Critical Examiner explicitly framed analytical contributions (e.g., identifying risks in proposals) and the Facilitator actively directed turns to specific members—behaviors that were less structured in the Pre-Test, where turn-taking was more incidental.

## 4. Discussion and Conclusion

This exploratory case study observed two complementary transfer patterns following AI-supported role fulfillment intervention. First, a delayed equalization effect: participation equality remained unchanged during AI intervention but improved substantially (44.2% reduction in word-based Gini) immediately after AI removal. Second, sustained role fulfillment: RFI remained 17.0% above baseline after intervention, indicating that role-aligned behaviors were internalized rather than merely scaffolded.

We hypothesize that cognitive load mediation explains the delayed participation effect: during the Test session, participants faced a triple-task demand—participating in discussion, reading AI feedback on Slack, and integrating that feedback—which may have temporarily suppressed balanced participation. The post-intervention surge in participation equality suggests that participants internalized the norm during intervention but could only fully enact it once cognitive resources were freed. Notably, role fulfillment skills did not show this delay, possibly because role-specific behaviors (e.g., facilitating turns, examining proposals) are more directly actionable than the implicit norm of balanced participation. This cognitive load hypothesis requires verification through direct measurement (e.g., NASA-TLX) in future studies.

Together, these results suggest a comprehensive transfer effect from brief AI intervention. Both participation norms and role fulfillment skills showed sustained improvement after support removal, though via different temporal pathways. Participation equality showed delayed transfer (no improvement during intervention, strong improvement post-intervention), consistent with implicit learning that requires cognitive consolidation. Role fulfillment showed immediate and sustained transfer (+19.3% during intervention, +17.0% maintained post-intervention), consistent with explicit learning through direct behavioral prompting. This temporal dissociation implies that a single brief AI intervention can simultaneously support both implicit norm acquisition and explicit skill development, though the observable effects may emerge on different timescales.

This study has several limitations. The sample size (N=4) restricts generalizability; the findings should be interpreted as exploratory observations from a proof-of-concept deployment rather than confirmatory evidence. Although RFI scoring was conducted across three independent LLM runs showing high consistency (SD: 0.002–0.009), the metric lacks human annotator cross-validation (e.g., Cohen's kappa), and cognitive load was not directly measured. Future studies should validate these findings with larger samples, incorporate human-coded reliability checks for the RFI, measure cognitive load explicitly, and employ sequential analysis or epistemic network analysis to capture collaborative dynamics more comprehensively.

## References


Dillenbourg, P. (Ed.). (1999). Collaborative-learning: Cognitive and computational approaches. Elsevier.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. Educational Psychologist, 48(1), 56-66.

Strijbos, J. W., & De Laat, M. (2010). Developing the role concept for computer-supported collaborative learning: An explorative synthesis. Computers in Human Behavior, 26(4), 495-505.

Zheng, L., Niu, J., & Zhong, L. (2022). Effects of a learning analytics-based real-time feedback approach on knowledge elaboration, knowledge convergence, interactive relationships and group performance in CSCL. British Journal of Educational Technology, 53(1), 130-149.