

# WHEN ROBOTS BRAINSTORM WITH US

How robot facilitation style shapes social comparison and self-evaluation in creative group ideation

Author

Musy Ayoub

Supervisors

Catharine Oertel, Ruben Weijers

TU Delft

## 01 BACKGROUND

- People evaluate their abilities partly through social comparison (Festinger, 1954).
- Capable robots can become comparison targets when their contribution is visible (Yaar et al., 2024; Yaar & Erel, 2025).
- Creative ideation makes ideas visible, subjective, and easy to compare.

## 02 RESEARCH QUESTION

**How do different Pepper facilitation configurations shape social comparison, self-efficacy, perceived competence, and perceived contribution?**

- RQ1 social comparison with Pepper
- RQ2 self-efficacy + competence
- RQ3 usefulness, engagement, contribution

## 03 STUDY SNAPSHOT

N = 20

10 dyads

5 dyads / condition

- Between-subjects design: dynamic collaborative vs. pre-generated intervention.
- 20-min brainstorm: campus wellbeing + academic engagement.
- Descriptive analysis of surveys, sliders, and experiment notes.

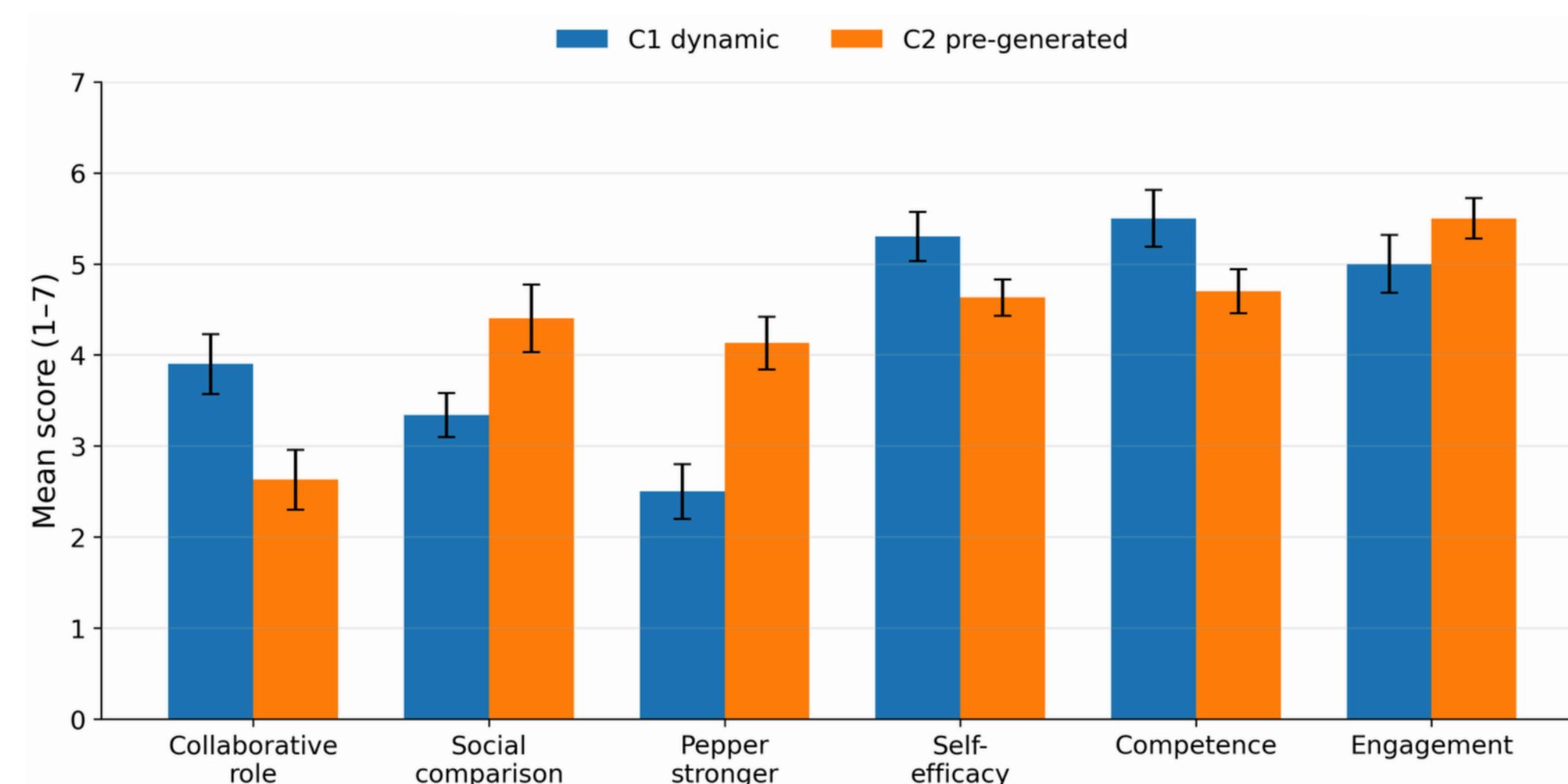
## 04 WHAT WAS MEASURED

- Role framing
- Social comparison
- Upward comparison
- Contribution sliders
- Self-efficacy
- Perceived competence
- Engagement
- Pepper usefulness
- Creative confidence
- Task evaluation

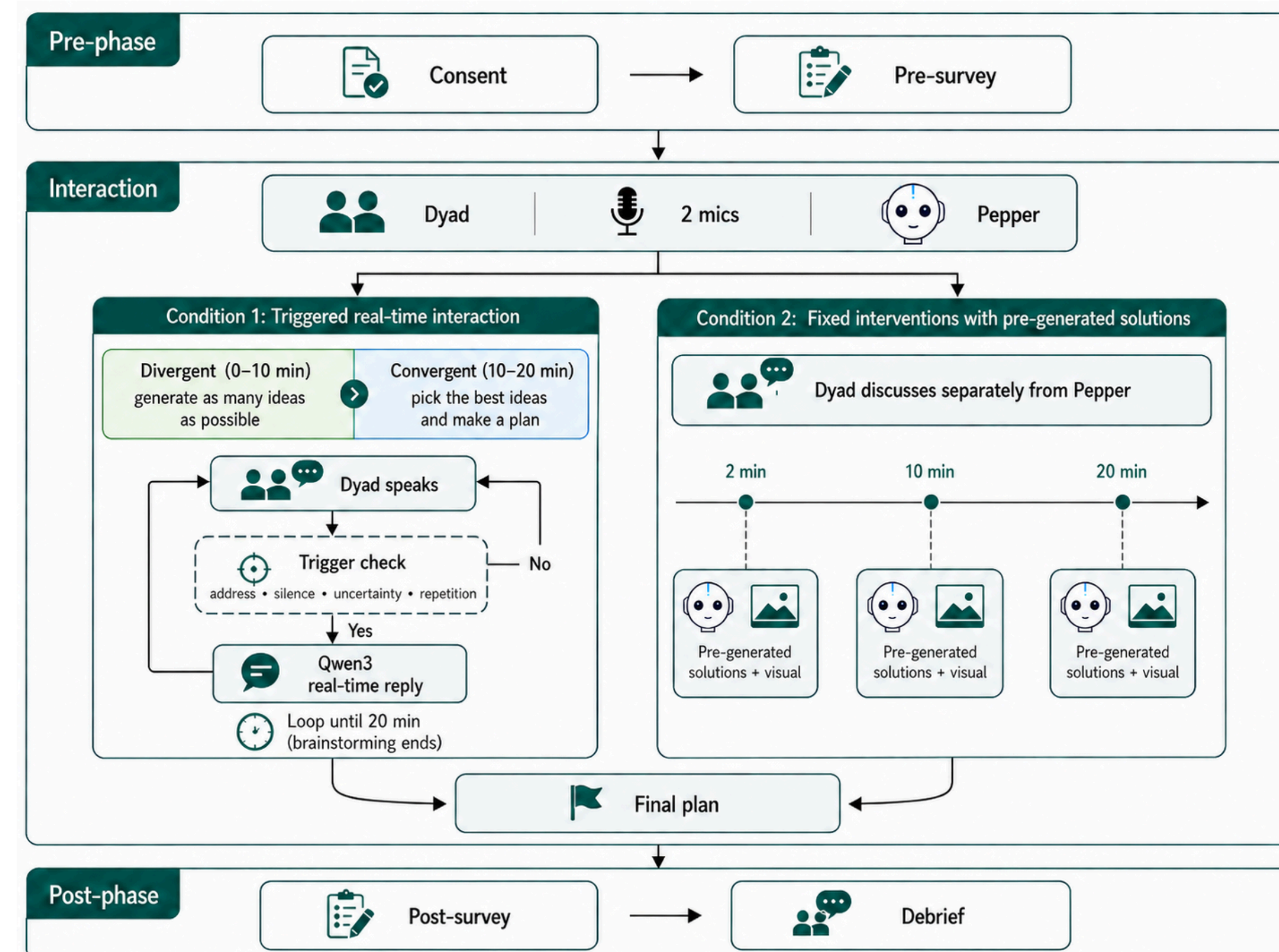
Pre: creative confidence + baseline · Post: Likert items, contribution sliders, and open responses

Main check: does Pepper feel collaborative, comparative, useful, or influential?

## 06 KEY RESULTS



## 05 EXPERIMENTAL PROCEDURE



- Shared setup: dyad, two microphones, Pepper, Deepgram Nova EU endpoint.
- C1: triggered real-time Qwen3-8B replies during divergent/convergent phases.
- C2: Pepper intervened at 2, 10, and 20 min with pre-generated solutions + laptop visuals.
- Both conditions ended with a final plan, followed by post-survey and debrief.

**Design contrast**  
Adaptive dialogue vs. fixed prepared interventions

## 07 DISCUSSION

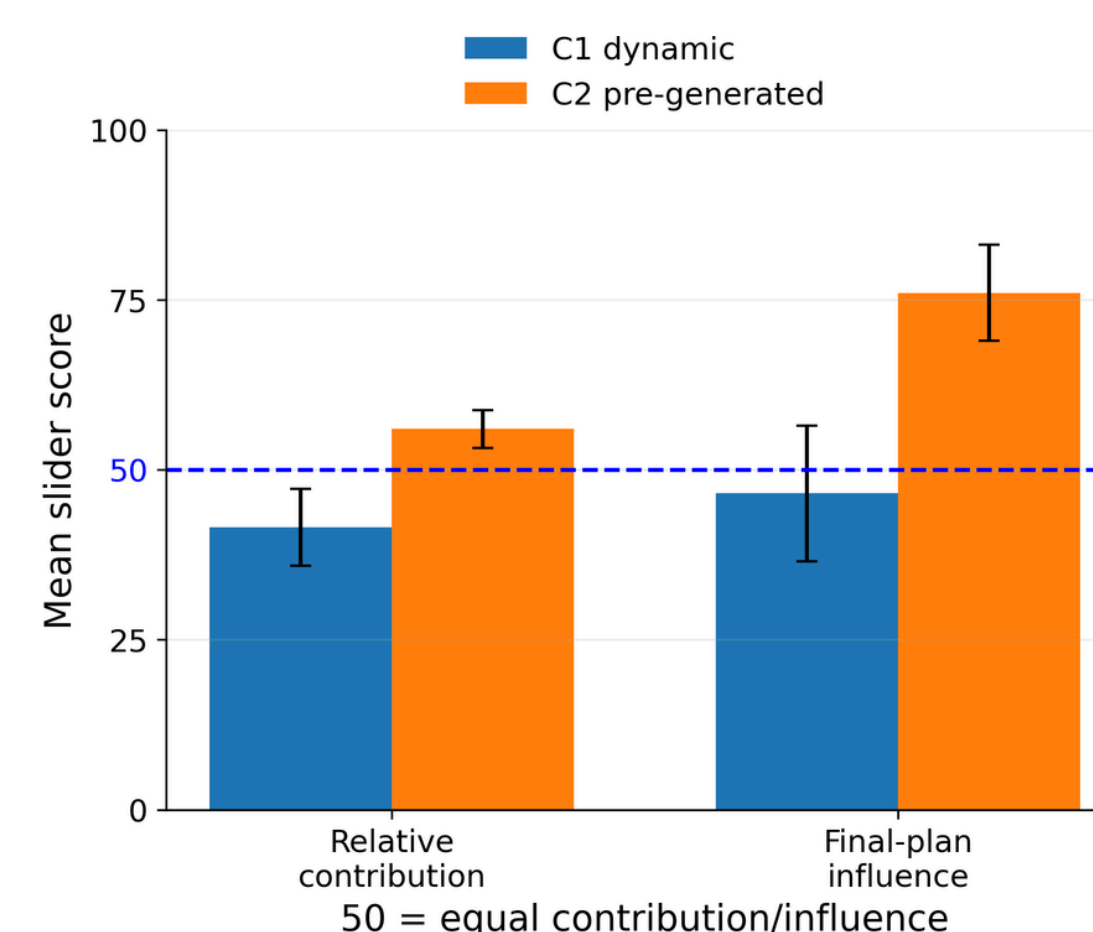
- C2 helped at idea-fatigue moments, but made Pepper more dominant and comparison-heavy.
- C1 preserved competence more, but depended on timing, turn-taking, and responsiveness (Hoffman, 2019).
- Extends robot comparison work from side-by-side tasks to dialogue-based creative work (Yaar et al., 2024; Yaar & Erel, 2025).

## 08 CONTRIBUTION

- Compares two competent robot facilitation styles in creative dialogue.
- Shows facilitation style shapes self-evaluation, not only task support.
- Design goal: useful robot contributions without reducing human agency.

## 09 LIMITATIONS

- Small descriptive study; dyad responses were not fully independent.
- Complete systems were compared: Qwen3-8B dialogue vs. GPT-5.5 + GPT-image-2.0 visuals.
- No independent expert rating of final plans or idea quality.



- C1 felt more collaborative and supported higher self-efficacy/competence.
- C2 produced stronger social comparison and much higher perceived Pepper influence.
- Engagement/usefulness were slightly higher in C2; task evaluation stayed similar.
- Error bars = standard error; results are descriptive.