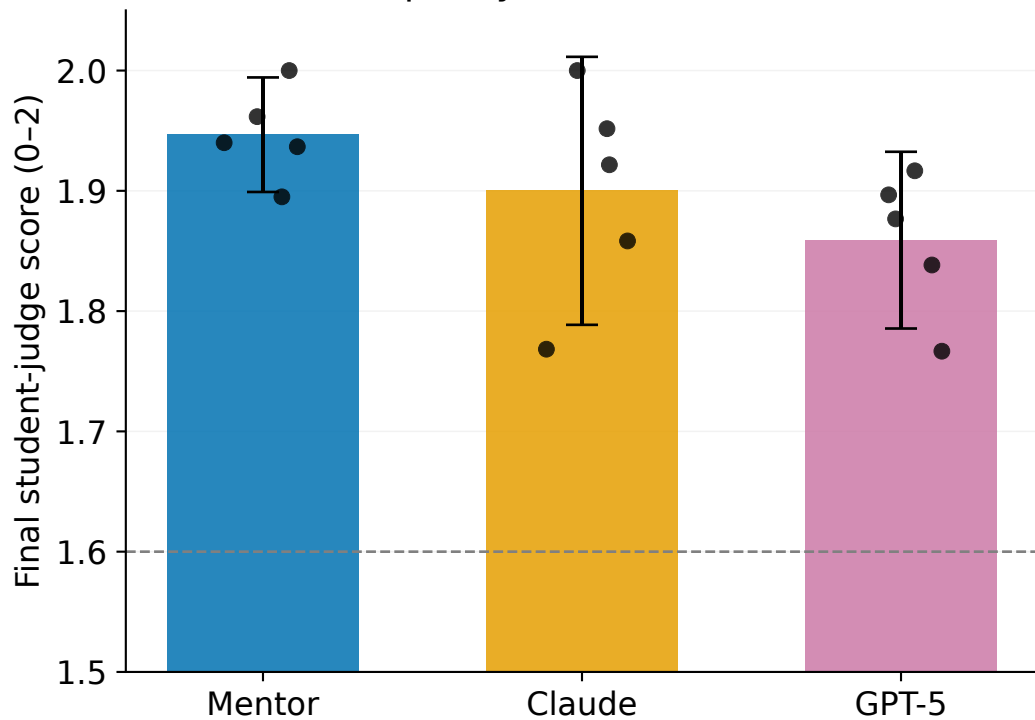


Multi-turn conversation quality (student judge only)

A

Final quality after conversation

**B**

Average score trajectory across scenarios

