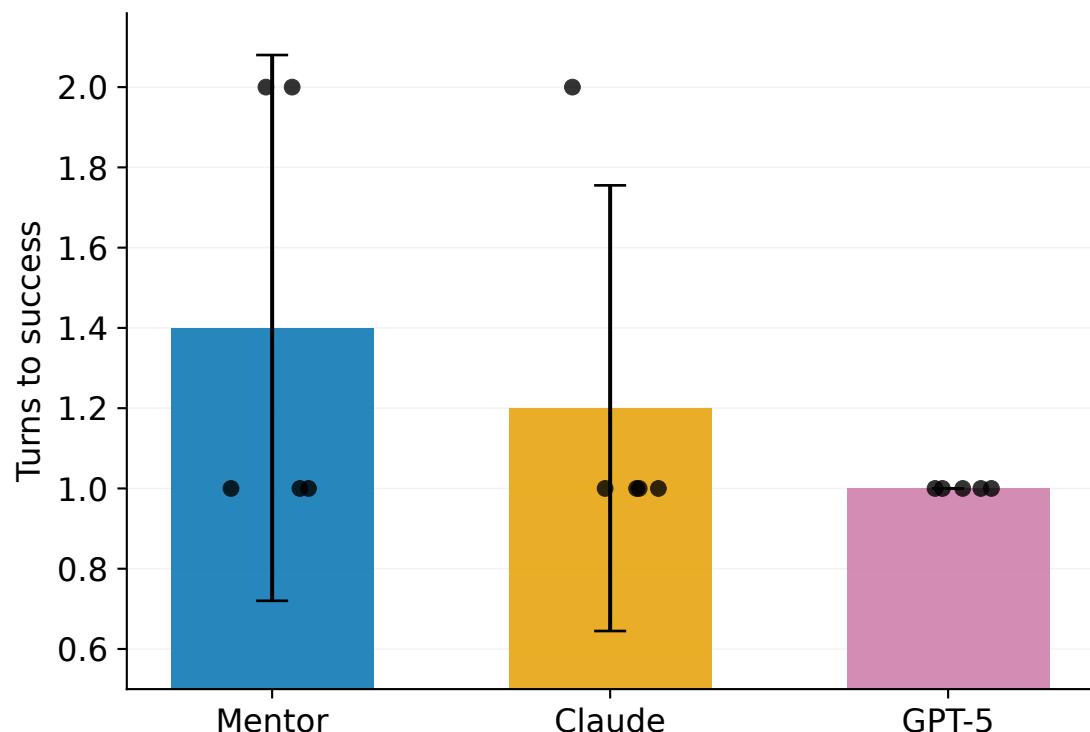


Multi-turn conversation quality and efficiency

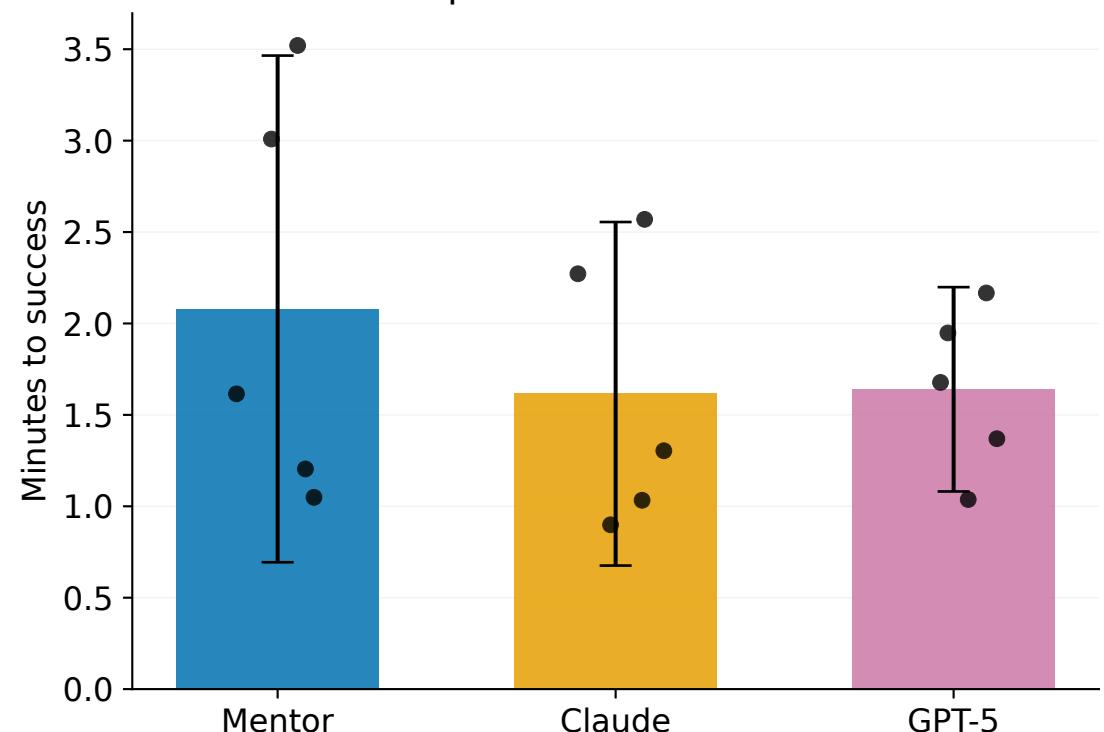
A

Conversation turns to success



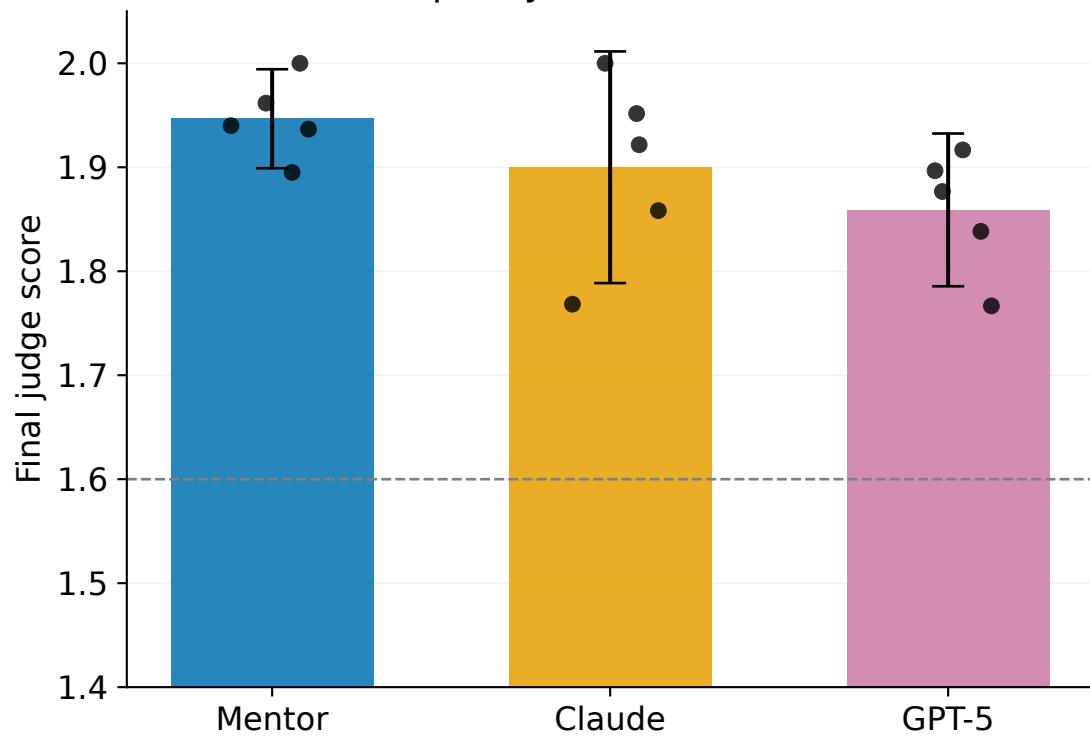
B

Elapsed time to success



C

Final quality after conversation



D

Average score trajectory across scenarios

