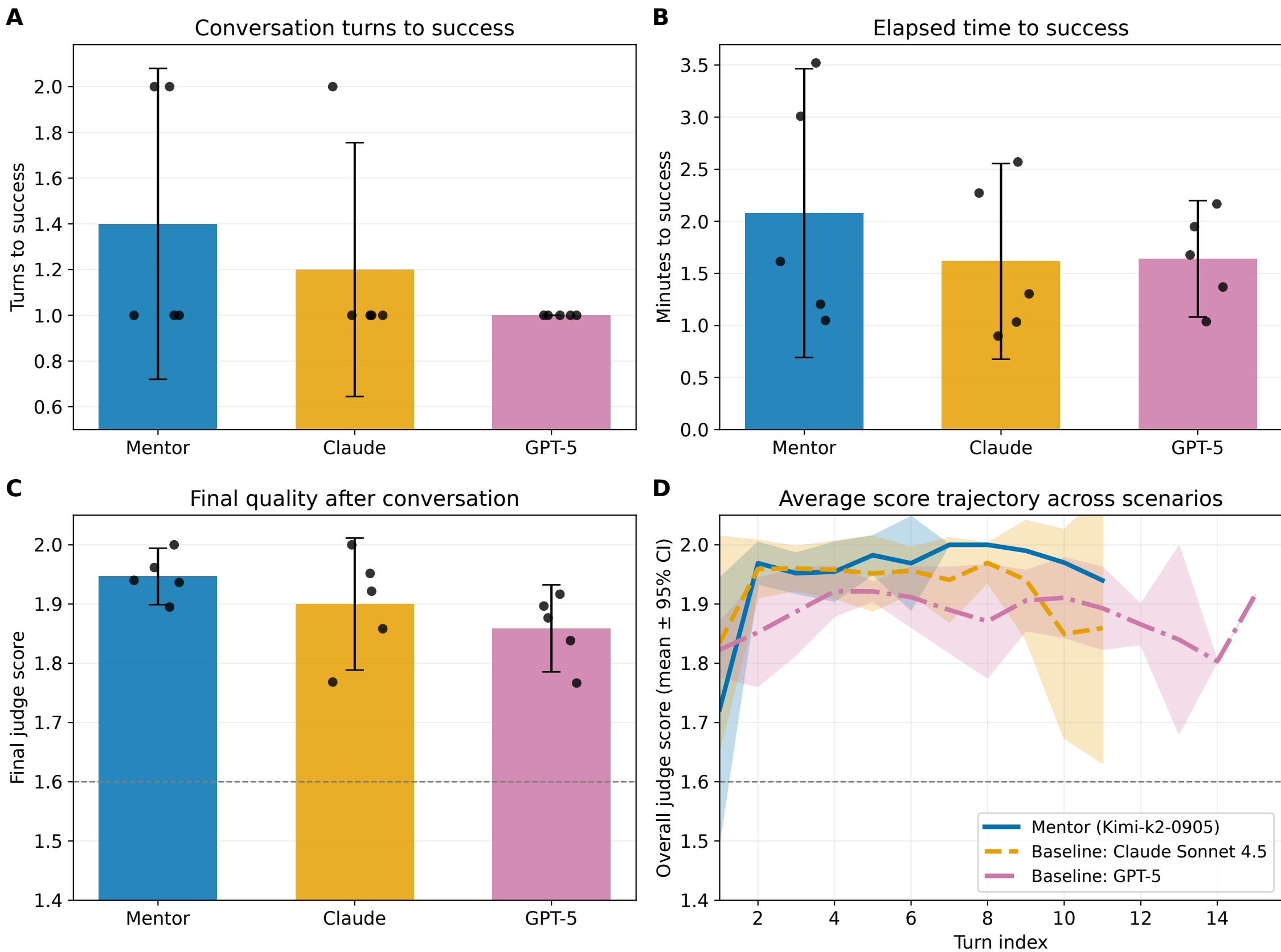


# Multi-turn conversation quality and efficiency



Error bars show 95% CI; dots mark individual scenarios (n=5 per agent). Success threshold (score  $\geq 1.6$ ) shown as dashed gray line.