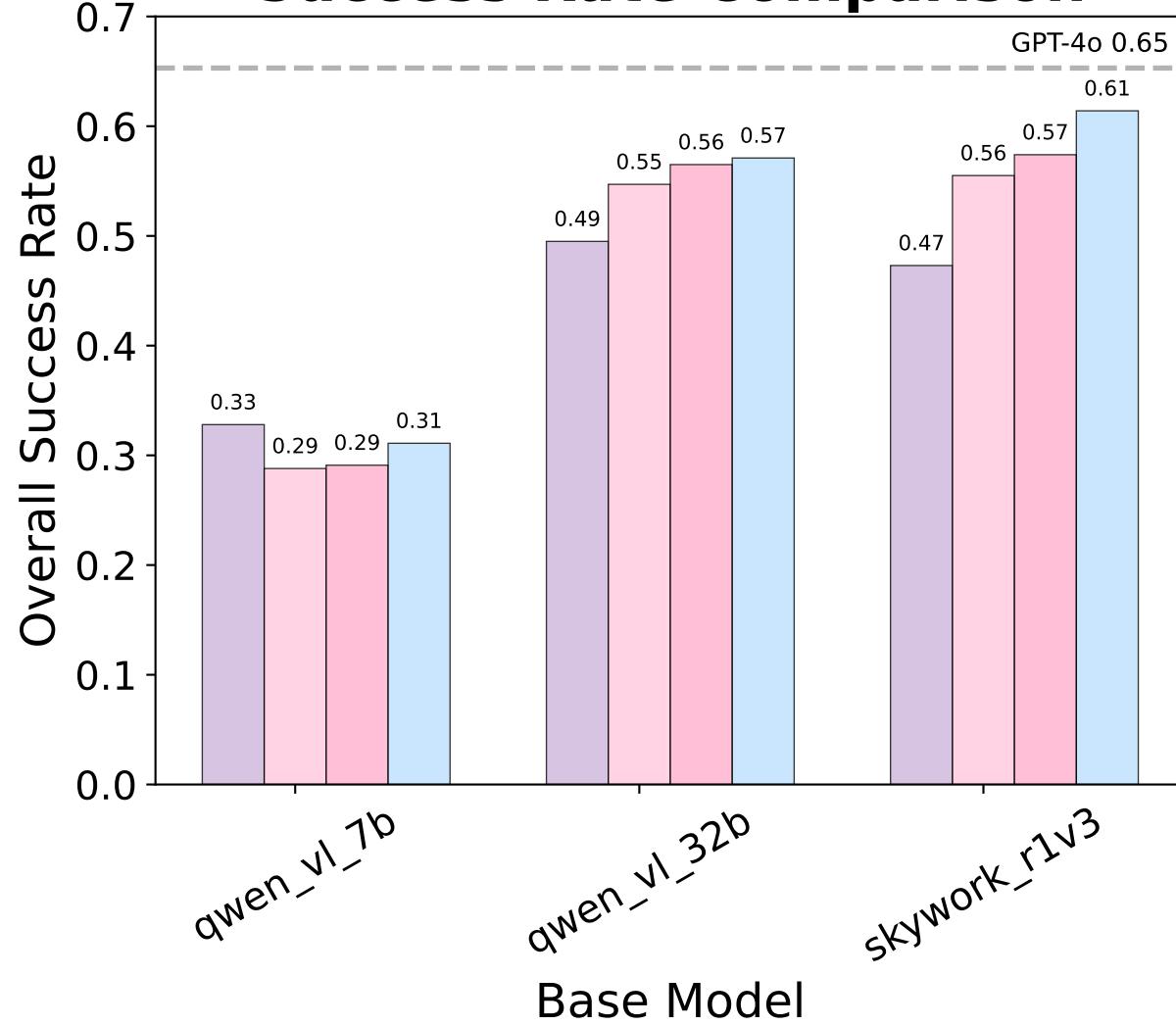
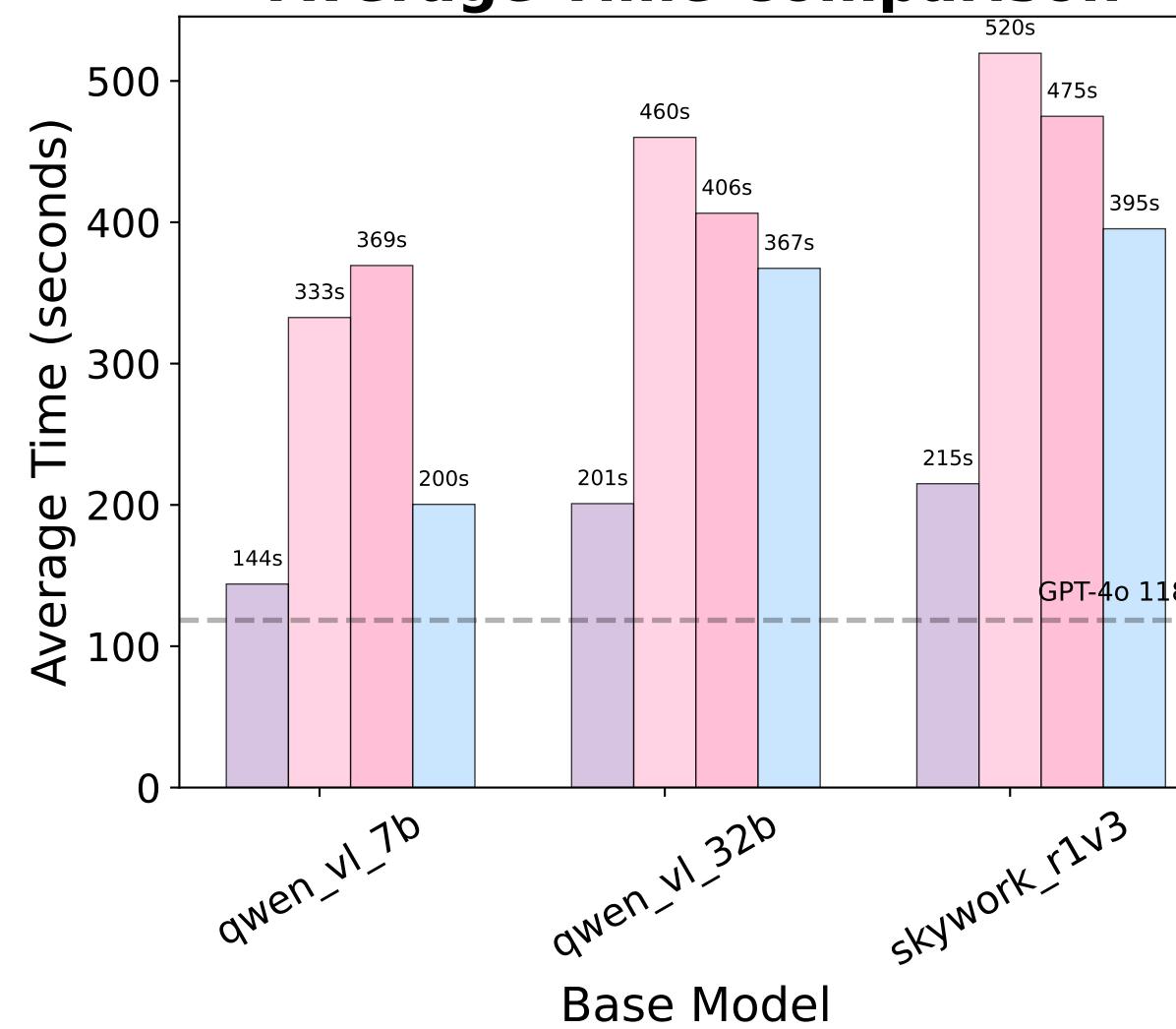


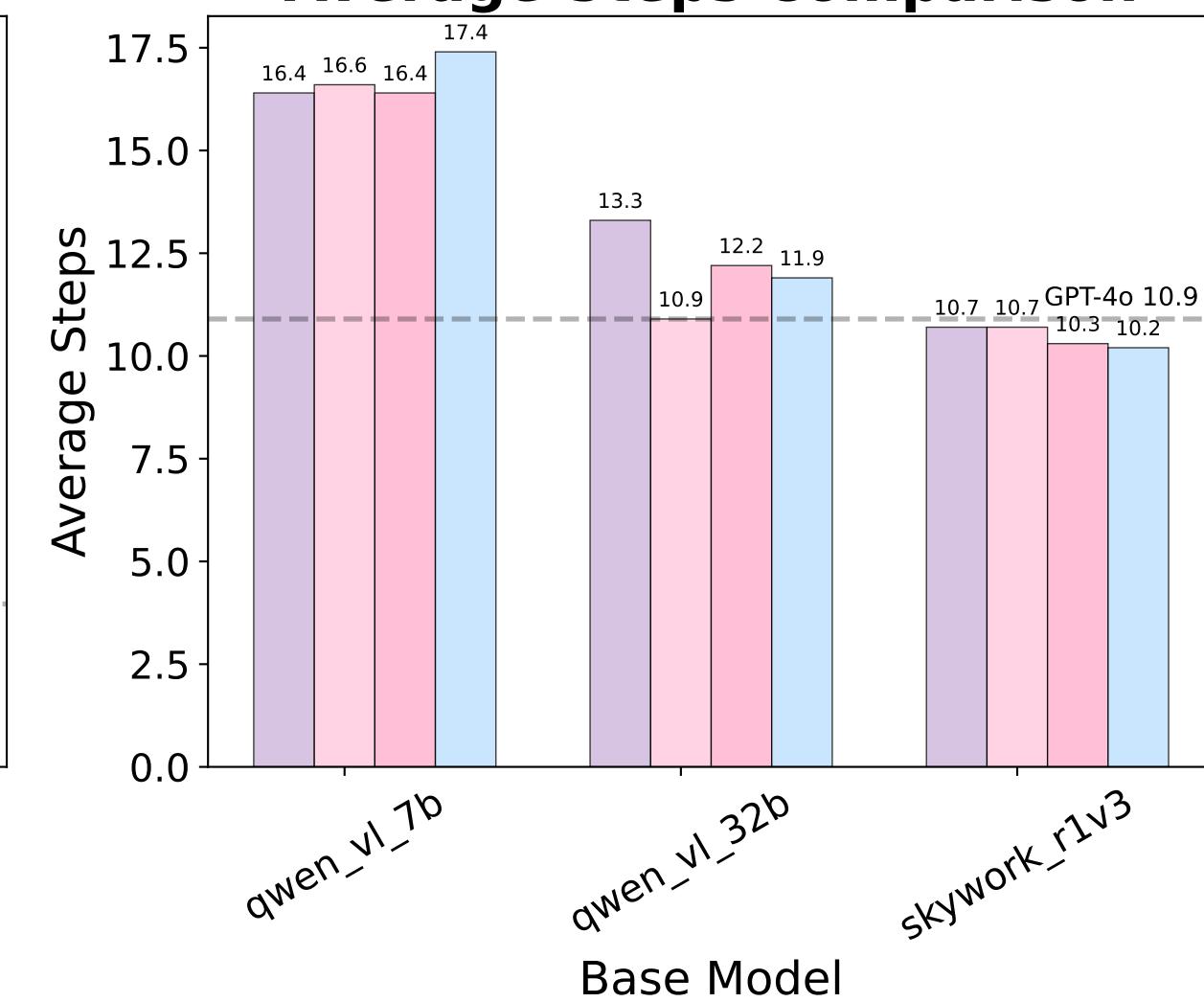
Success Rate Comparison



Average Time Comparison



Average Steps Comparison



Baseline WebCoach (Qwen3-8b) --- GPT-4o baseline
WebCoach (GPT-4o) WebCoach (Qwen3) with Self-exp