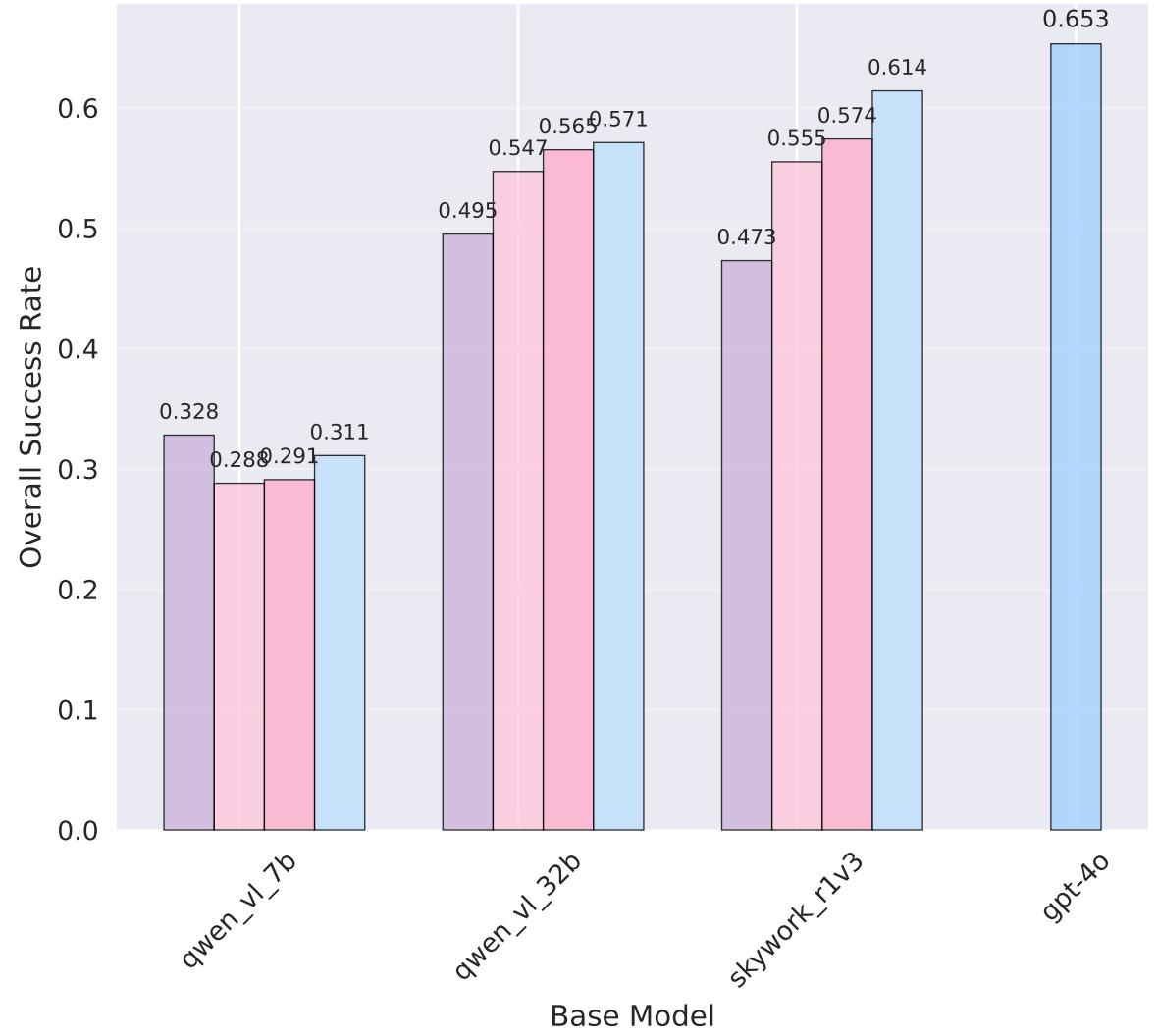
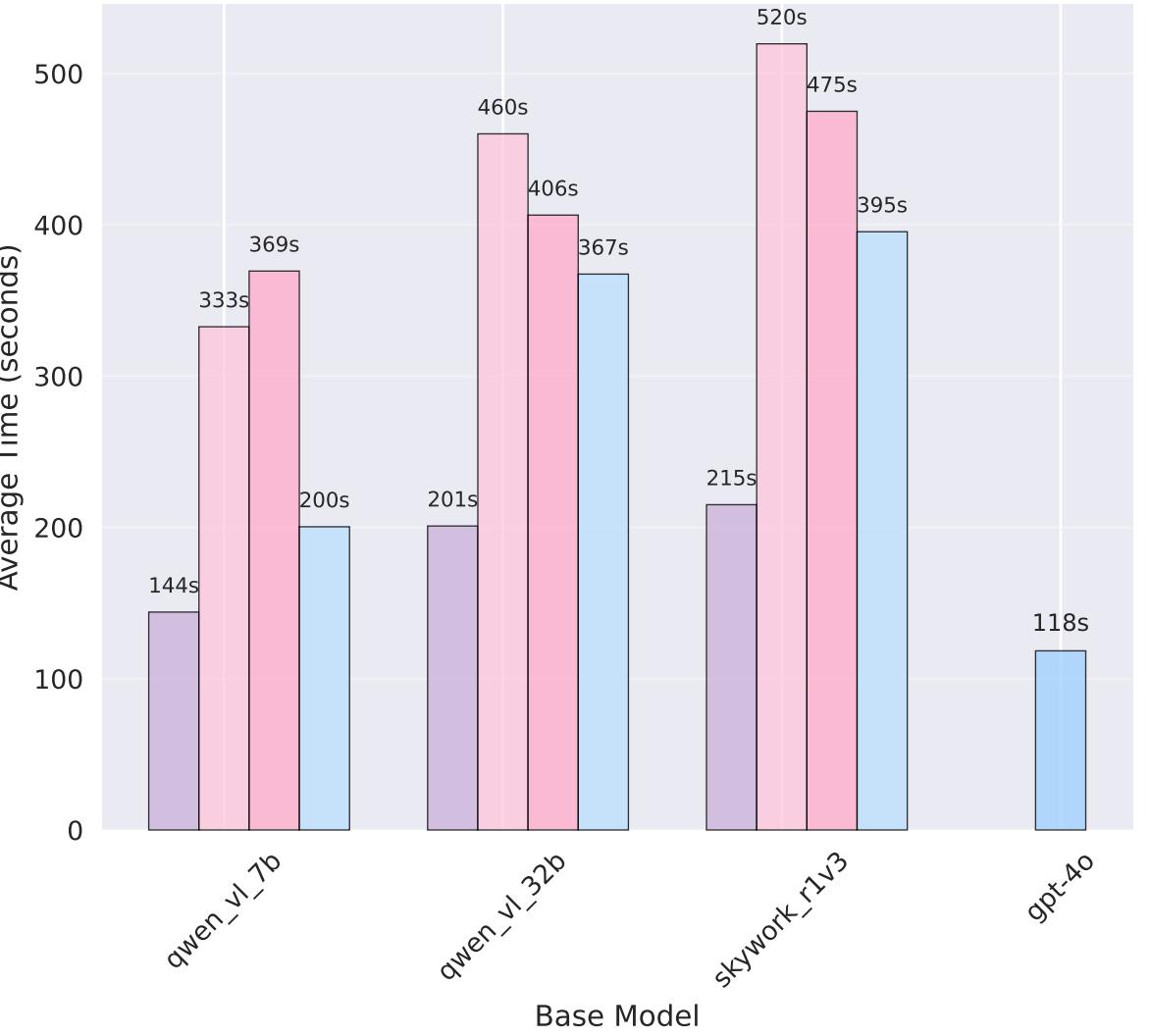


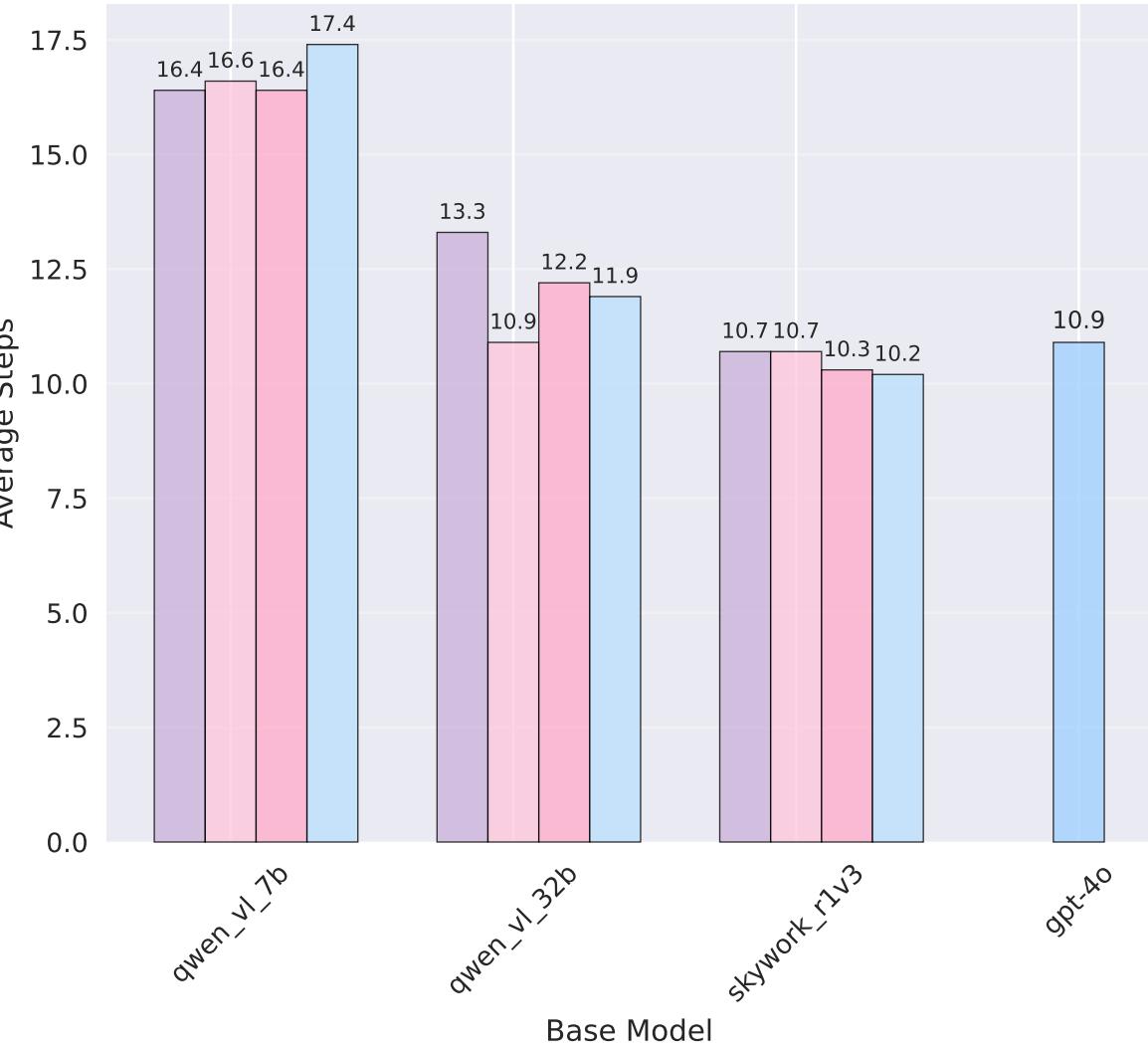
Success Rate Comparison



Average Time Comparison



Average Steps Comparison



Legend:
■ Baseline
■ WebCoach (Qwen3-8b)
■ WebCoach (GPT-4o)
■ WebCoach (Qwen3) with Self-exp
■ GPT-4o Baseline