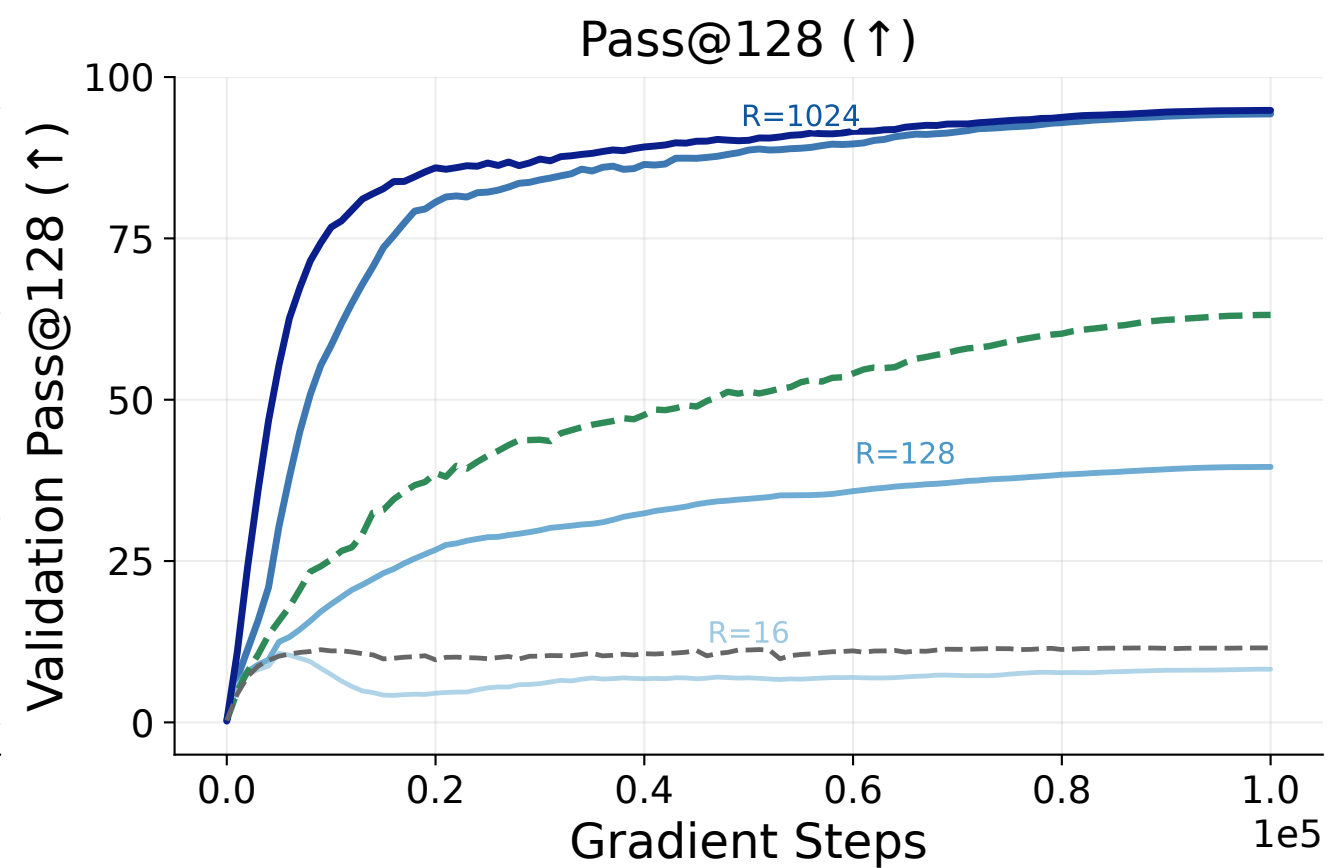
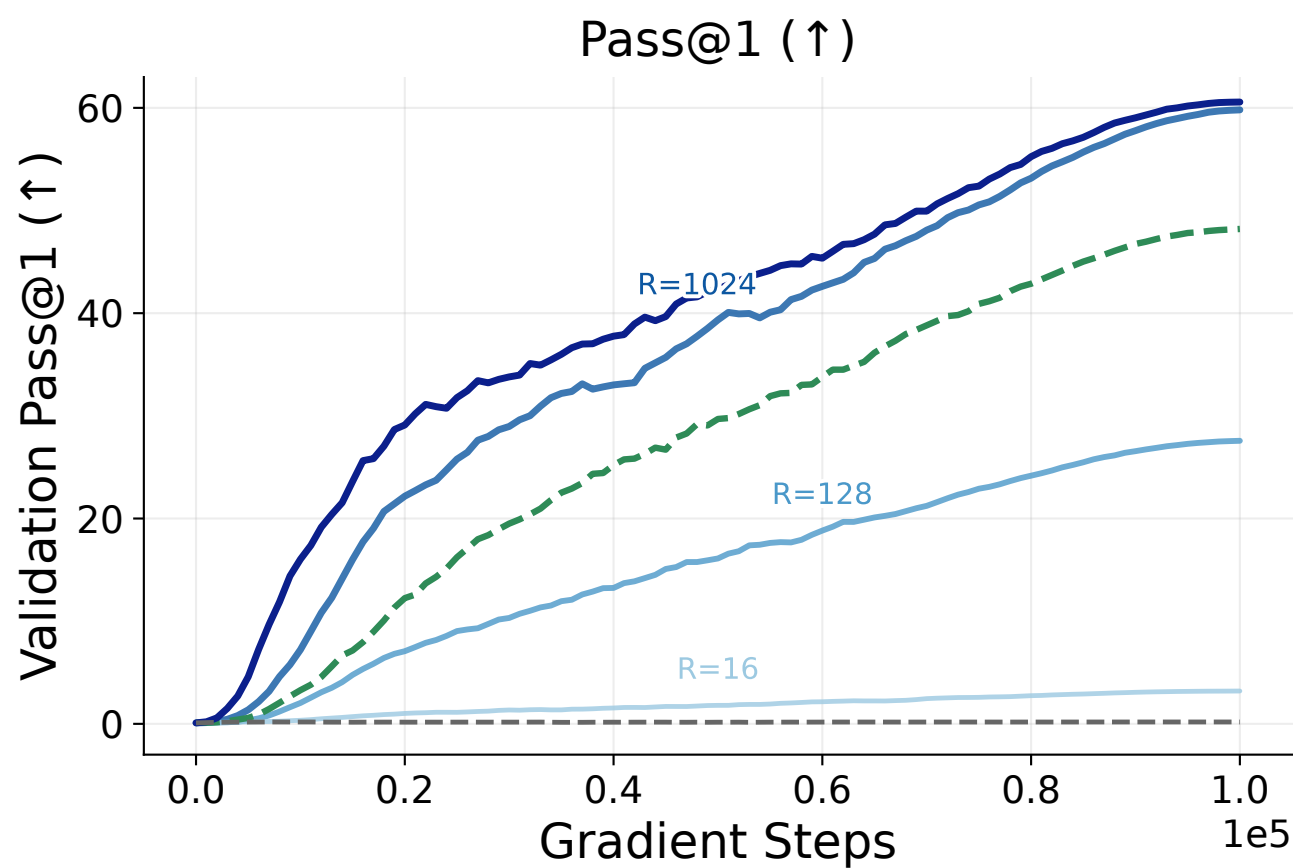




— SAGE($\tau = 0.2$) - - - GRPO (Rollout=1024) — Cross Entropy - - - REINFORCE (Rollout=1024)