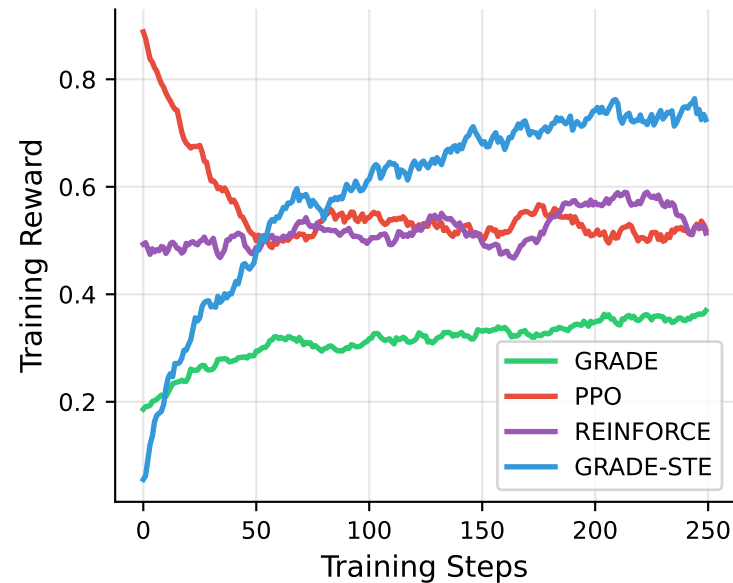
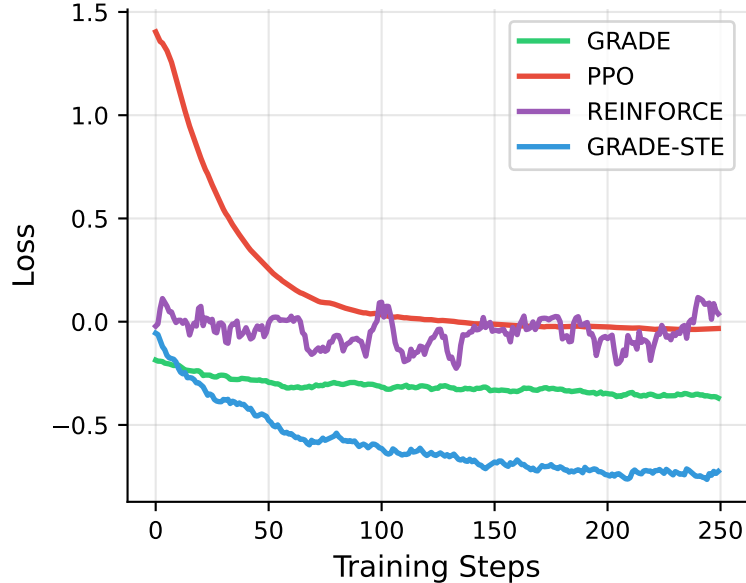


(a) Training Reward vs Steps



(b) Loss vs Training Steps



(c) KL from Reference Model

