(b) RL Training Dynamics