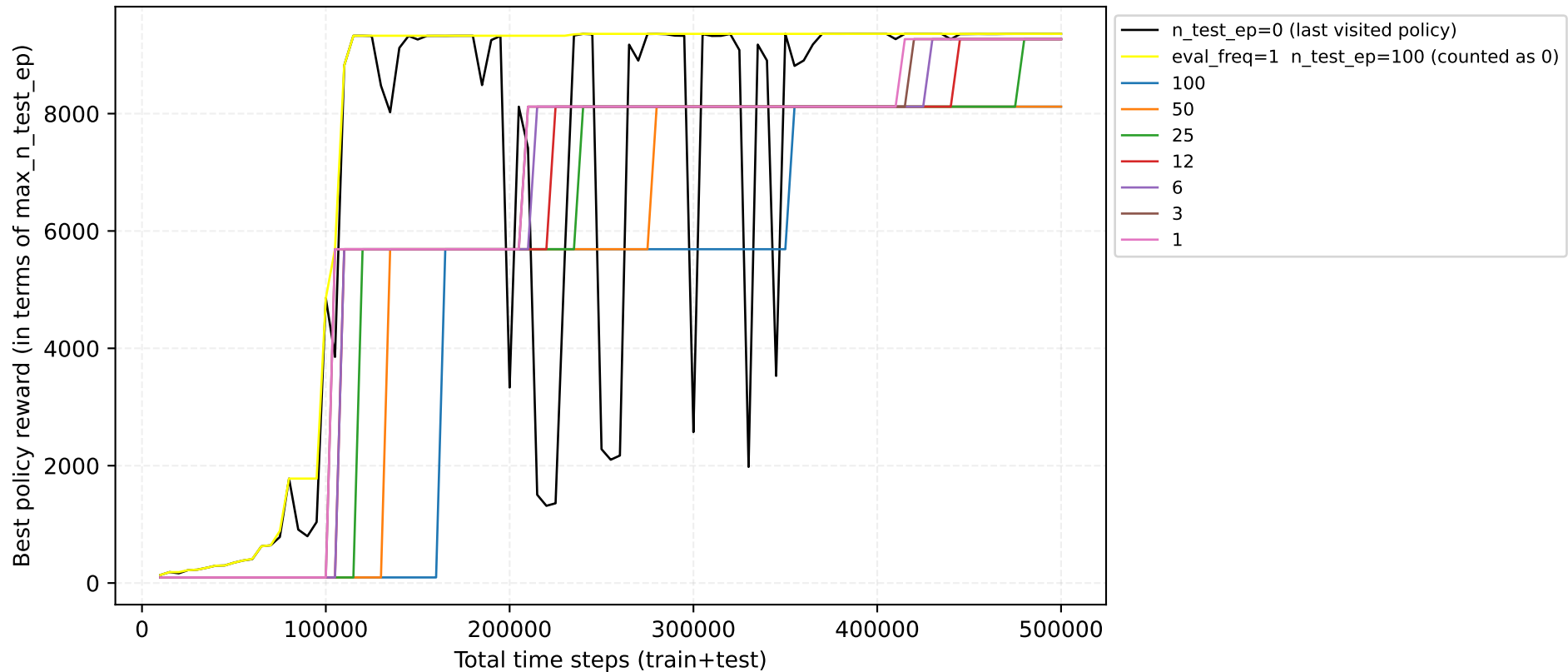


Learning-curve: best policy found  
Constant eval\_freq: 50 policy; Variable n\_test\_ep (legend)



(2048 train timesteps = training 1 policy = 1 trajectory)