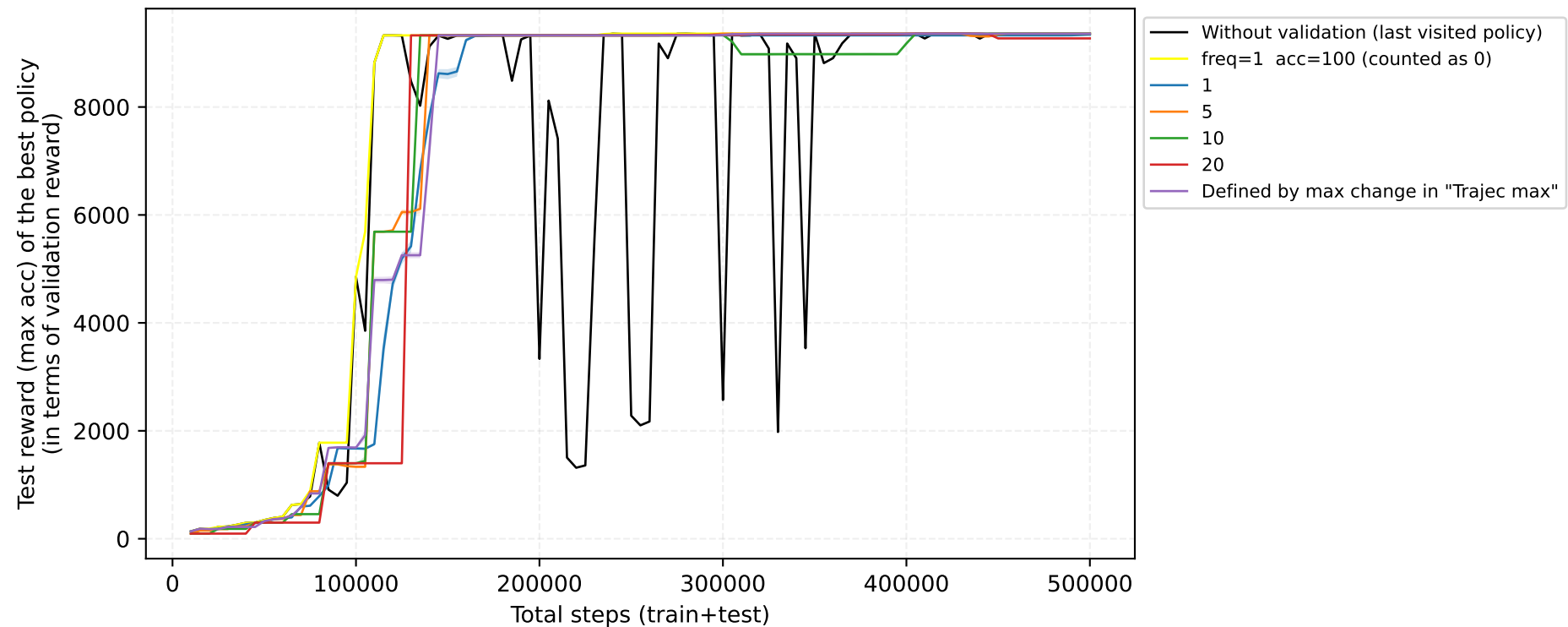


Learning-curves with

Same validation acc: 5 (test episodes)
Different validation freq (policies, see legend)



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