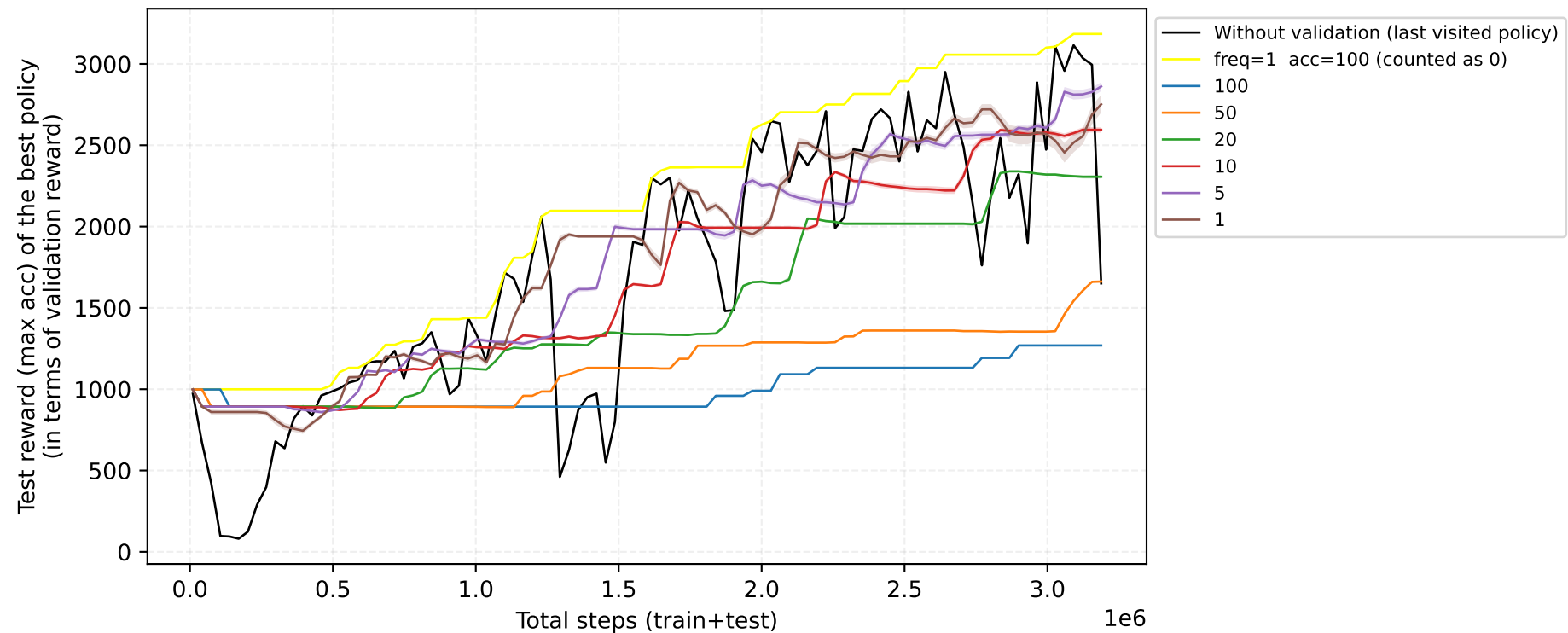


Learning-curves with

Same validation freq: 10 (policies)
Different validation acc (test episodes, see legend)



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