Learning-curves with
different metrics from train rewards to validate all policies