Testing episodes duration w.r.t. starting position smoothed with moving window average of size 10 200 DQN with heuristic reward DQN with RND reward Dyna reward 180 160 **Duration** 140 120 100 80 -0.60-0.55-0.50-0.45-0.40Starting position