RLHF Chosen vs Rejected Win Rates