Phi 3:medium - métricas por conjunto de teste

