Code Generation Performance Comparison on ParEval