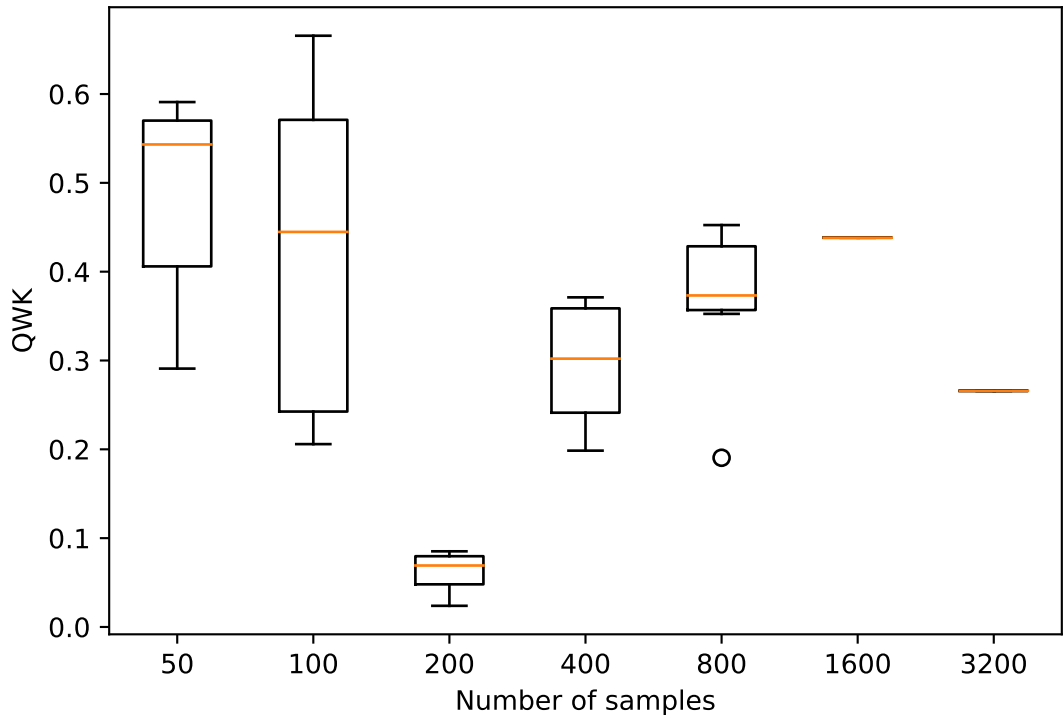


QWK of score\_1 assigned to samples created by turbo-testing by two BERT Models  
One trained with samples created by turbo one trained with samples created by gpt4.



Each box represents exactly 6 kappa values