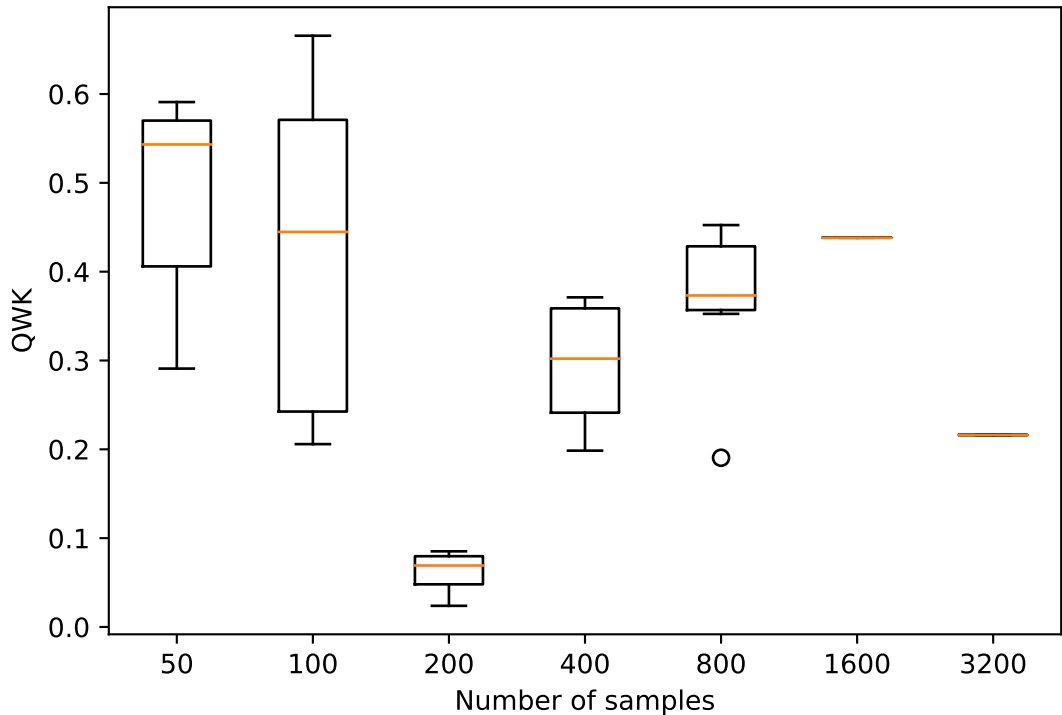


QWK of score_1 assigned to samples created by turbo-testing by two BERT Models
One trained with samples created by gpt4 one trained with samples created by turbo.



Each box represents exactly 6 kappa values