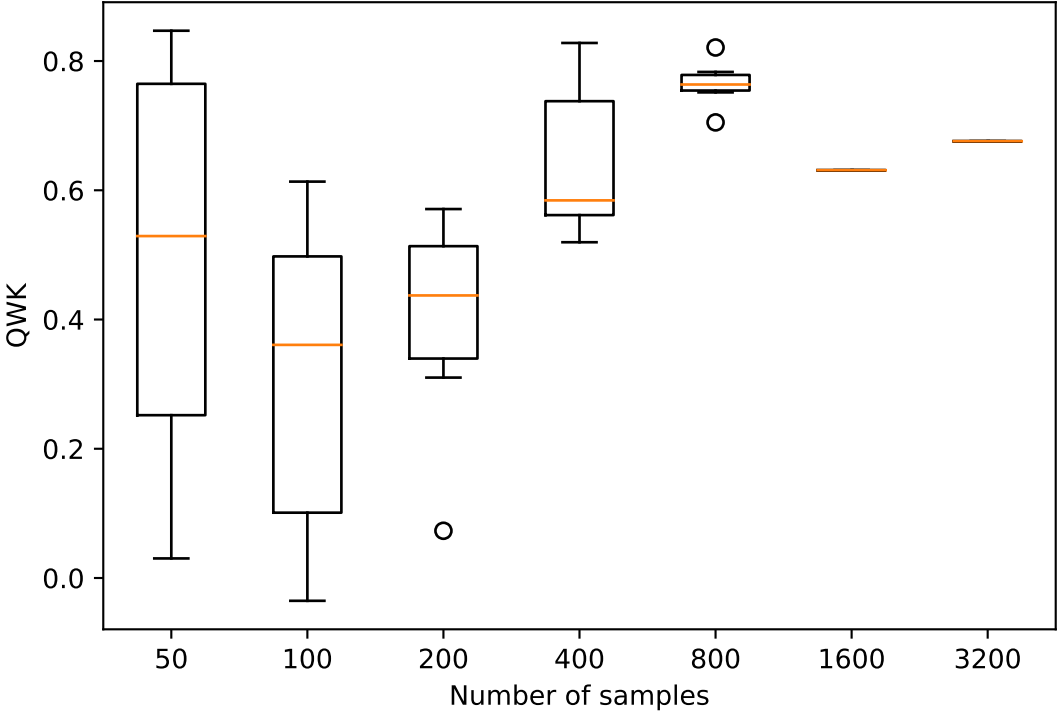


QWK of score\_1 assigned to samples created by gpt4-testing by two BERT Models  
One trained with samples created by davinci one trained with samples created by turbo.



Each box represents exactly 6 kappa values