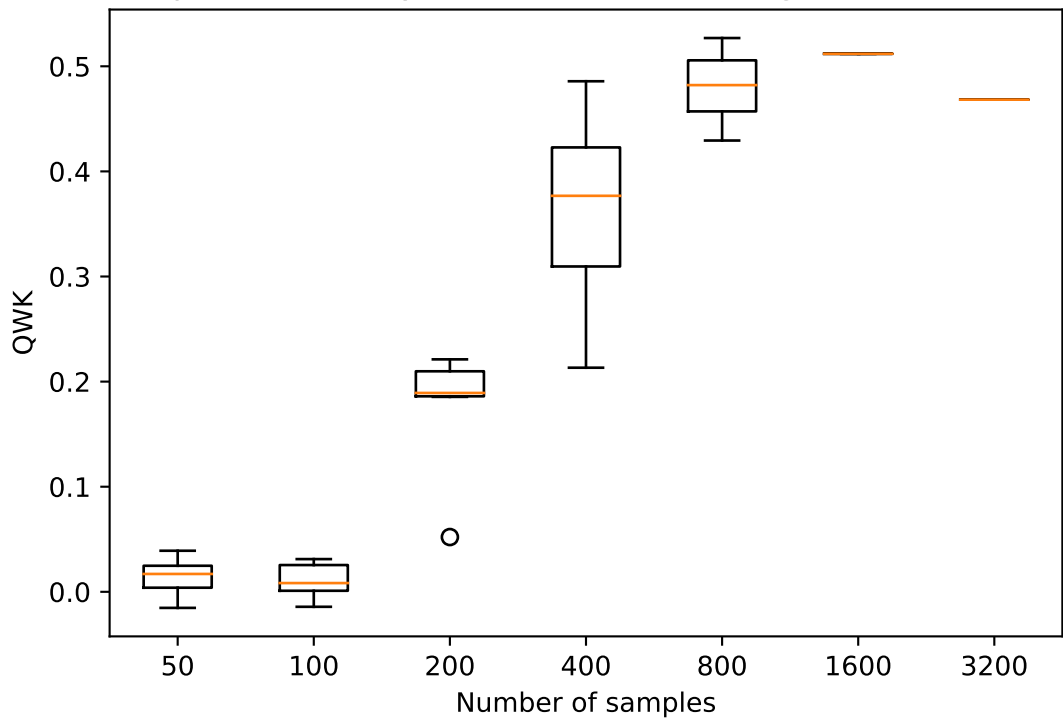


BERT Model trained with samples created by text-davinci-003 rating answers created by gpt4 used for testing



Each box represents exactly 6 kappa values