Test F1 Score on BERT Large for the QQP Data Set 0.7 0.6 Fl 0.5 0.4 Vanilla Finetuning **REGRADS** Majority Classifier 256 1024 64 Training Set Size