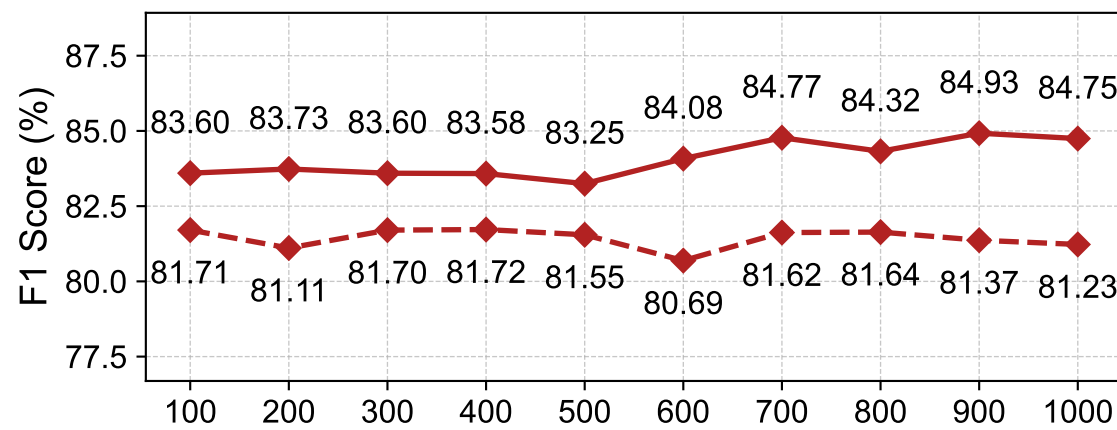


Wilcoxon, $p_{adj} = 0.006$ **

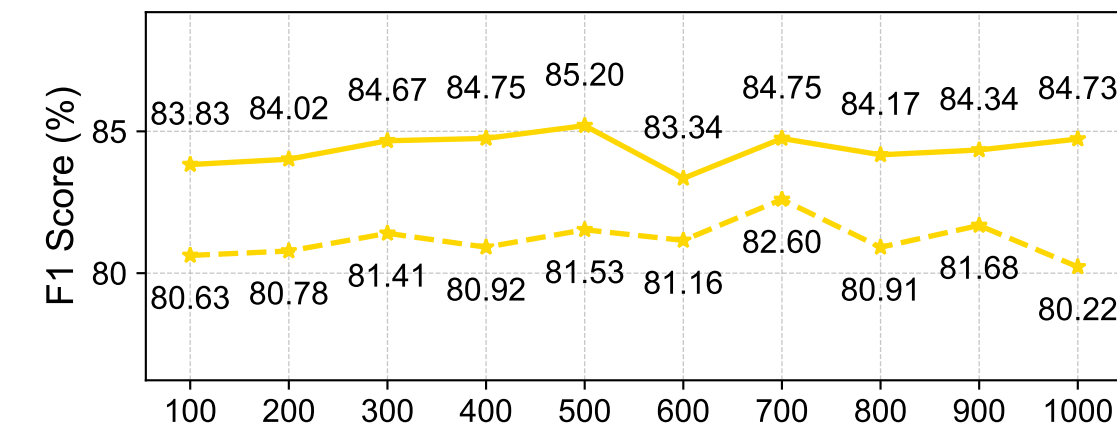
ACD

$M_{diff} = 2.63$



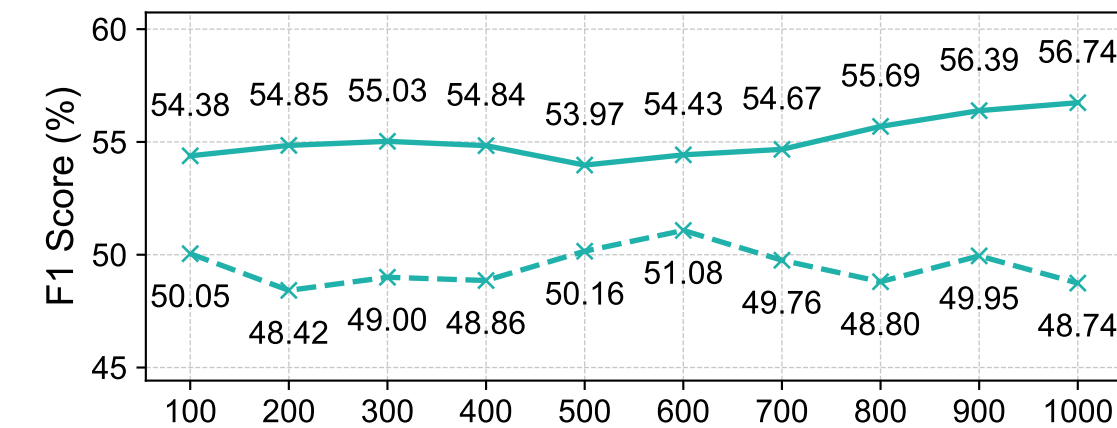
t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 3.20$



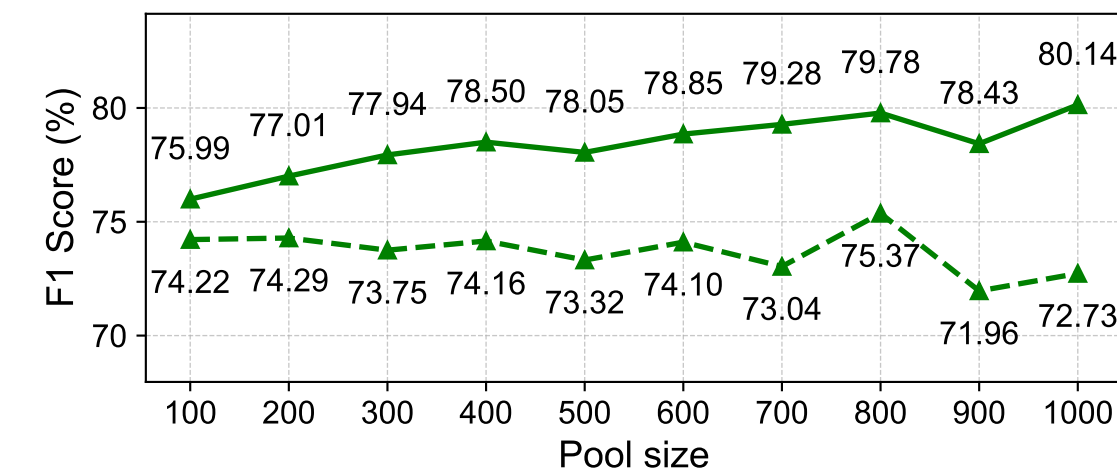
t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 5.62$



t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 4.70$



Rest16

FlightABSA

Coursera

Hotels

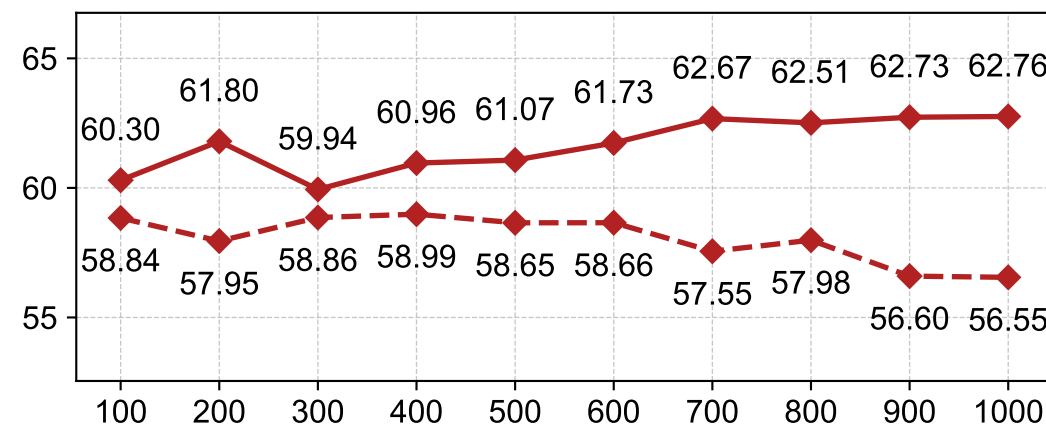
RAG

Random

t-test, $p_{adj} < 0.001$ ***

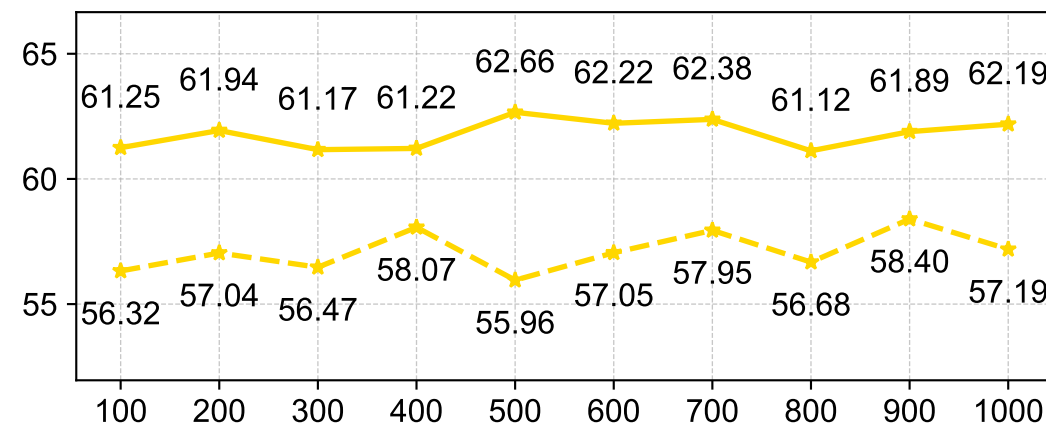
TASD

$M_{diff} = 3.59$



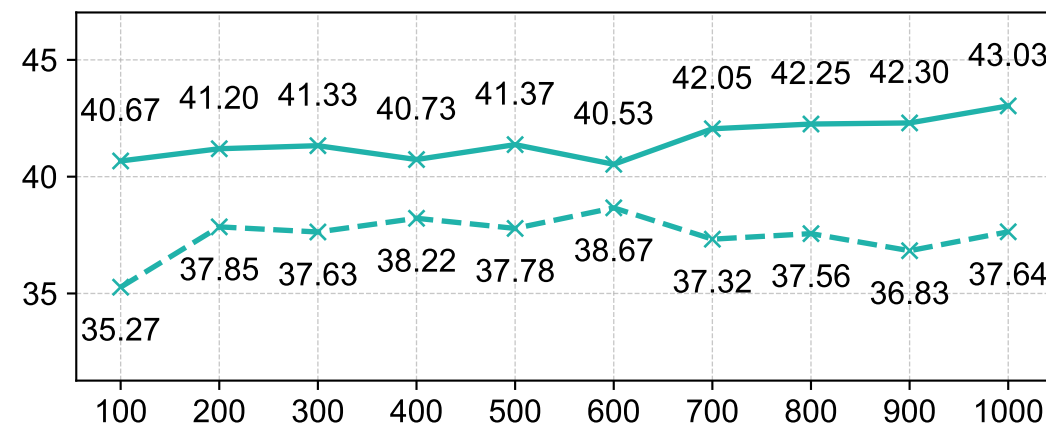
t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 4.69$



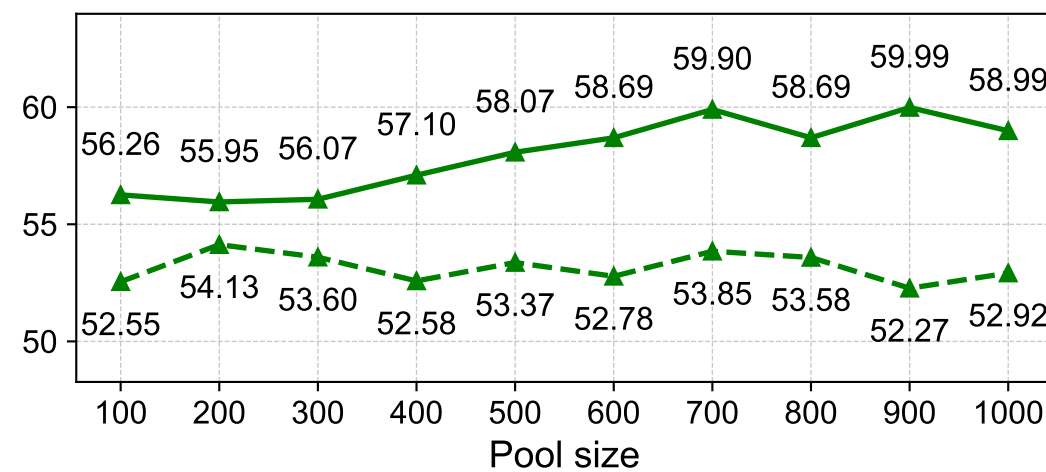
t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 4.07$



t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 4.81$



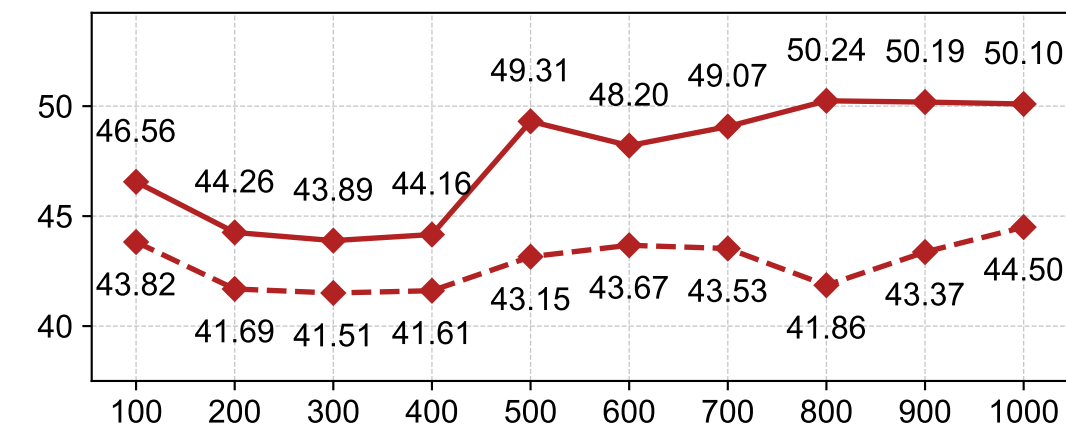
RAG

Random

Wilcoxon, $p_{adj} = 0.006$ **

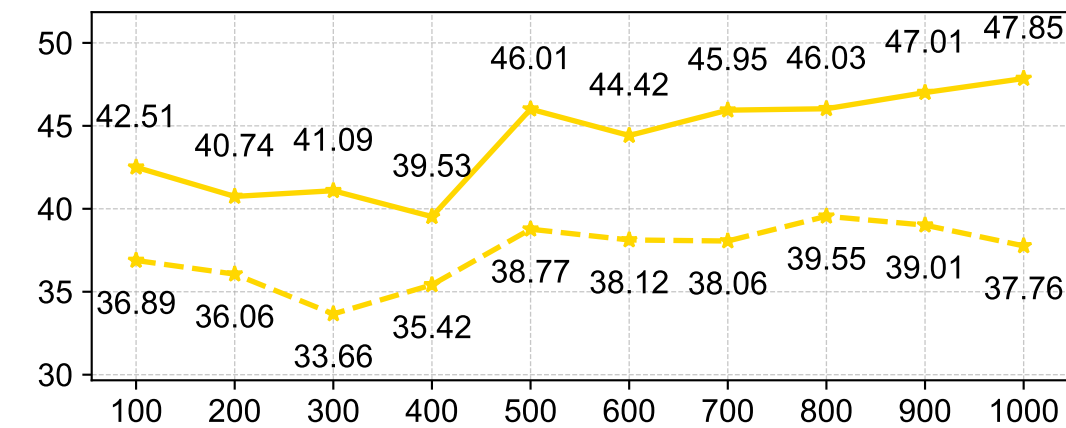
ASQP

$M_{diff} = 4.73$



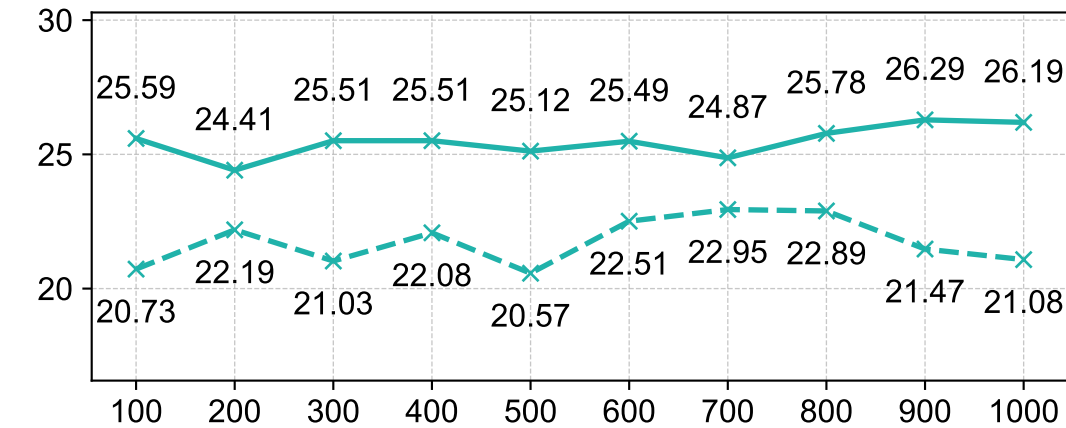
t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 6.78$



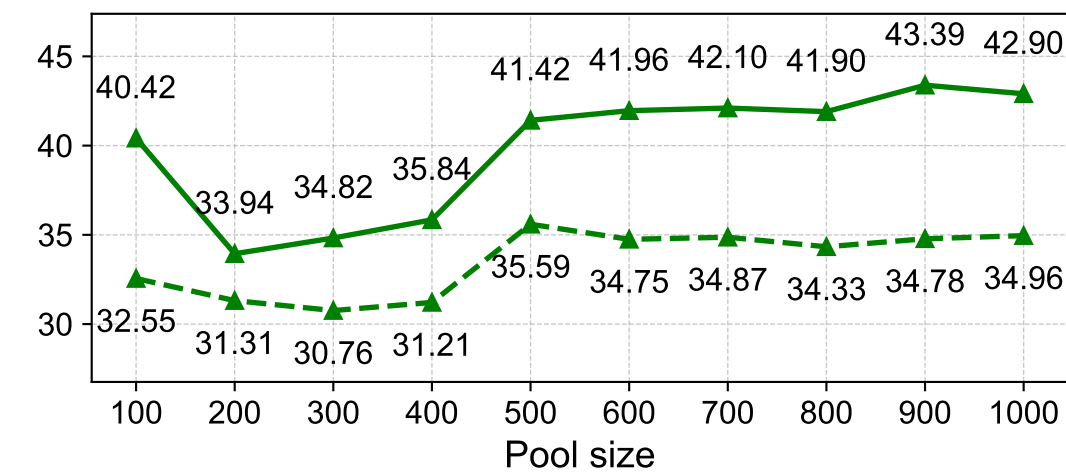
t-test, $p_{adj} < 0.001$ ***

$M_{diff} = 3.73$



Wilcoxon, $p_{adj} = 0.006$ **

$M_{diff} = 6.36$



RAG

Random