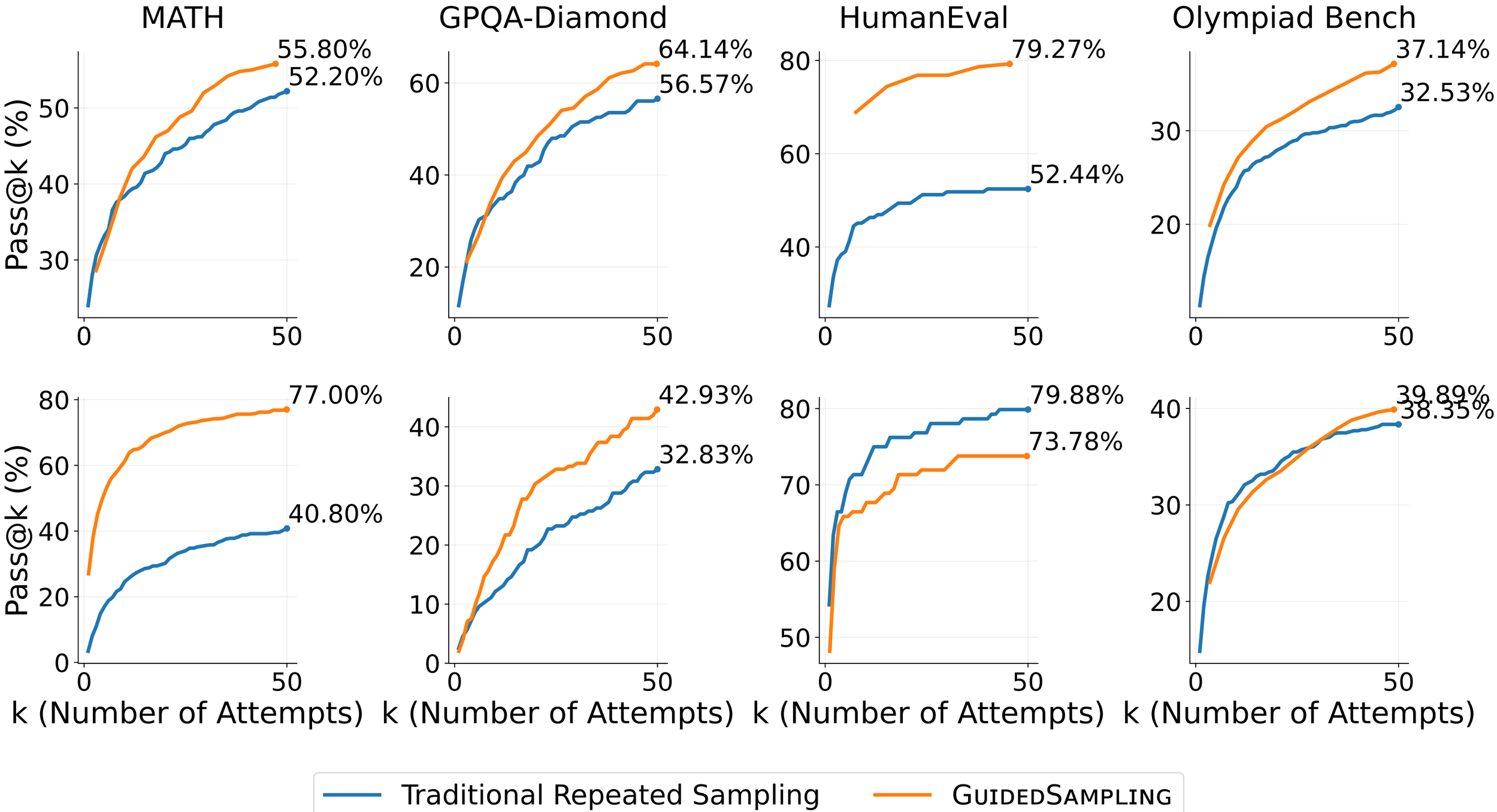
MATH
GPQA-Diamond
HumanEval
Olympiad Bench
Pass@k (%)
55.80%
52.20%
64.14%
56.57%
79.27%
52.44%
37.14%
32.53%
77.00%
40.80%
42.93%
32.83%
79.88%
73.78%
39.89%
38.35%
k (Number of Attempts)
k (Number of Attempts)
k (Number of Attempts)
k (Number of Attempts)
Traditional Repeated Sampling
GuidedSampling