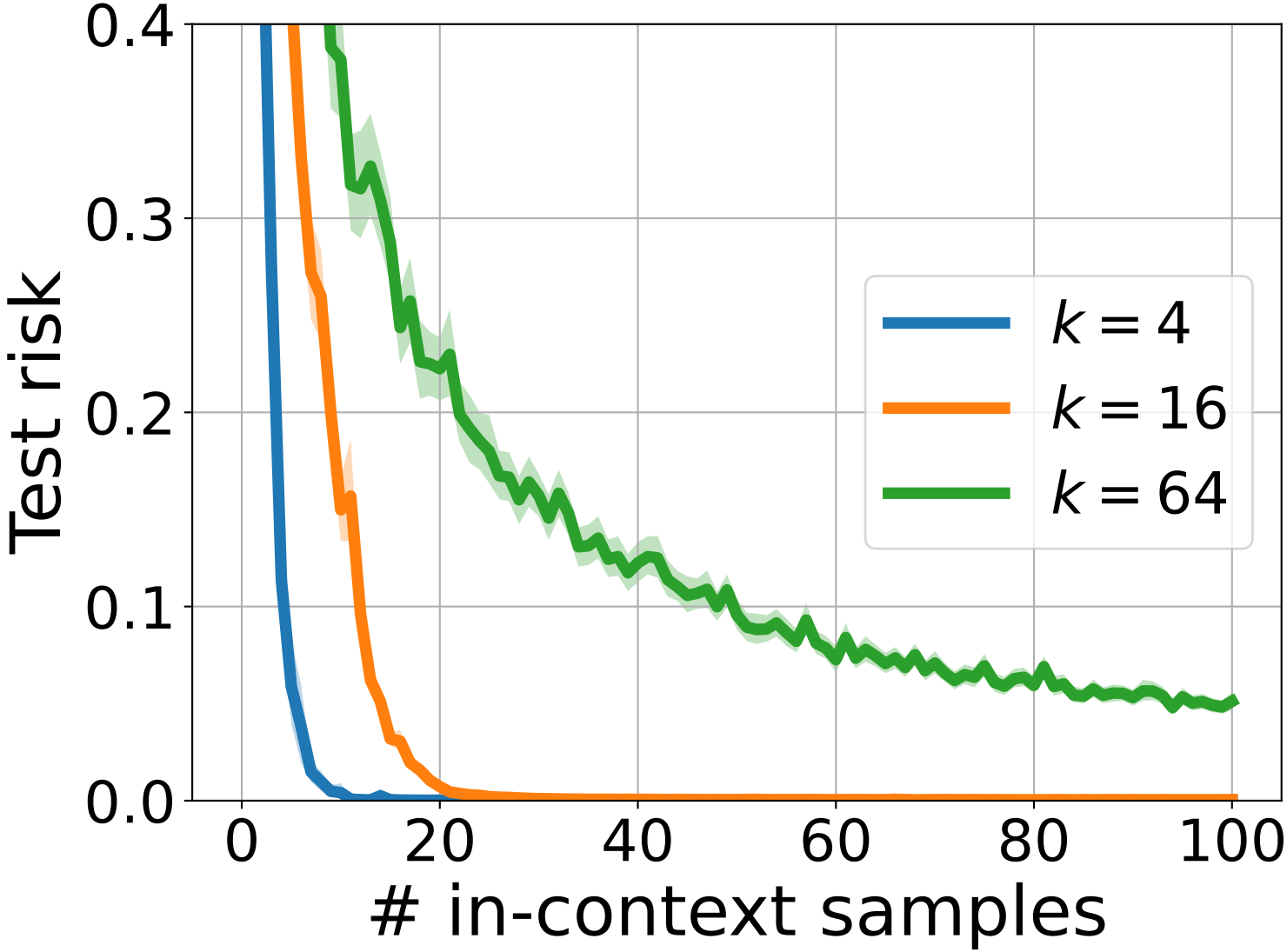
0.4
0.3
0.2
0.1
0.0
Test risk
0
20
40
60
80
100
in-context samples
$k = 4$
$k = 16$
$k = 64$