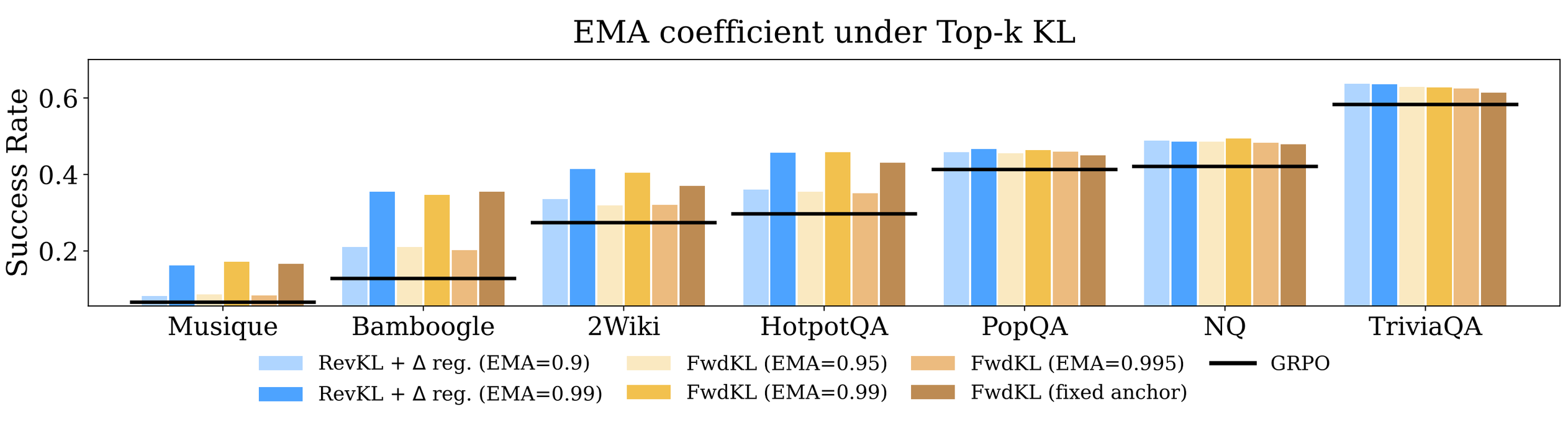

EMA coefficient under Top-k KL
Success Rate
0.2
0.4
0.6
Musique
Bamboogle
2Wiki
HotpotQA
PopQA
NQ
TriviaQA
RevKL + Δ reg. (EMA=0.9)
RevKL + Δ reg. (EMA=0.99)
FwdKL (EMA=0.95)
FwdKL (EMA=0.99)
FwdKL (EMA=0.995)
FwdKL (fixed anchor)
GRPO