Table 2 Test Cohort II: eight methods on 500 HotpotQA and 500 TriviaQA questions. EM and F1 are percentages.

Method	H EM	H F1	T EM	T F1	All EM	All F1
Direct	17.40	24.71	31.60	38.60	24.50	31.66
Always-RAG	28.20	38.27	30.00	36.22	29.10	37.25
Always-Multihop	29.20	40.12	29.60	36.18	29.40	38.15
Matched Binary	25.80	35.17	33.60	40.22	29.70	37.69
TARG-style	25.80	35.10	33.60	40.36	29.70	37.73
SUGAR-L-style	28.40	38.35	36.40	42.43	32.40	40.39
CAGR-R	24.40	33.84	35.00	42.31	29.70	38.07
CAGR-F	26.00	35.48	33.40	40.04	29.70	37.76

Method	Latency (ms)	Retr./q
Direct	190.71	0.000
Always-RAG	563.13	1.000
Always-Multihop	1810.69	2.000
Matched Binary	476.04	0.543
TARG-style	486.09	0.561
SUGAR-L-style	2354.80	1.255
CAGR-R	540.90	0.516
CAGR-F	491.23	0.553