Table 2 HotpotQA full-wiki retrieval, serialized evidence, reader outcomes, and structural cost on the same 500 questions. "Full para." requires every complete annotated supporting paragraph to remain after per-document truncation; dashes mark a paragraph-retention audit not repeated for the later learned-reranker row.

A. Retrieval and evidence interface

Condition	All title@10	Full para. 450	Full para. 900	Ret. calls	Query gen.
OneR/BM25	0.298	0.112	0.288	1.000	0.000
HyDE-BM25	0.348	0.102	0.332	1.000	1.000
Official IRCoT + Qwen	0.354	0.140	0.340	2.104	2.104
BM25 + BGE-base rescore	0.460	0.176	0.434	1.000	0.000
BM25 + BGE reranker v2-m3	0.468	–	–	1.000	0.000

B. Frozen evidence under two readers

Condition	Qwen EM	Qwen F1	Mistral EM	Mistral F1
OneR/BM25	0.220	0.2815	0.148	0.2674
HyDE-BM25	0.236	0.3017	0.170	0.3032
Official IRCoT + Qwen	0.262	0.3366	0.160	0.2920
BM25 + BGE-base rescore	0.270	0.3469	0.186	0.3372
BM25 + BGE reranker v2-m3	0.264	0.3480	0.208	0.3476

Exact title and exact title–paragraph coverage coincide in the four interface-audited rows. Full-paragraph retention is conservative because the aligned artifact lacks official supporting-sentence IDs. Each method has one final reader call. BGE-base is retained as a separately labeled bi-encoder sensitivity comparator; the learned cross-encoder was evaluated with both readers.