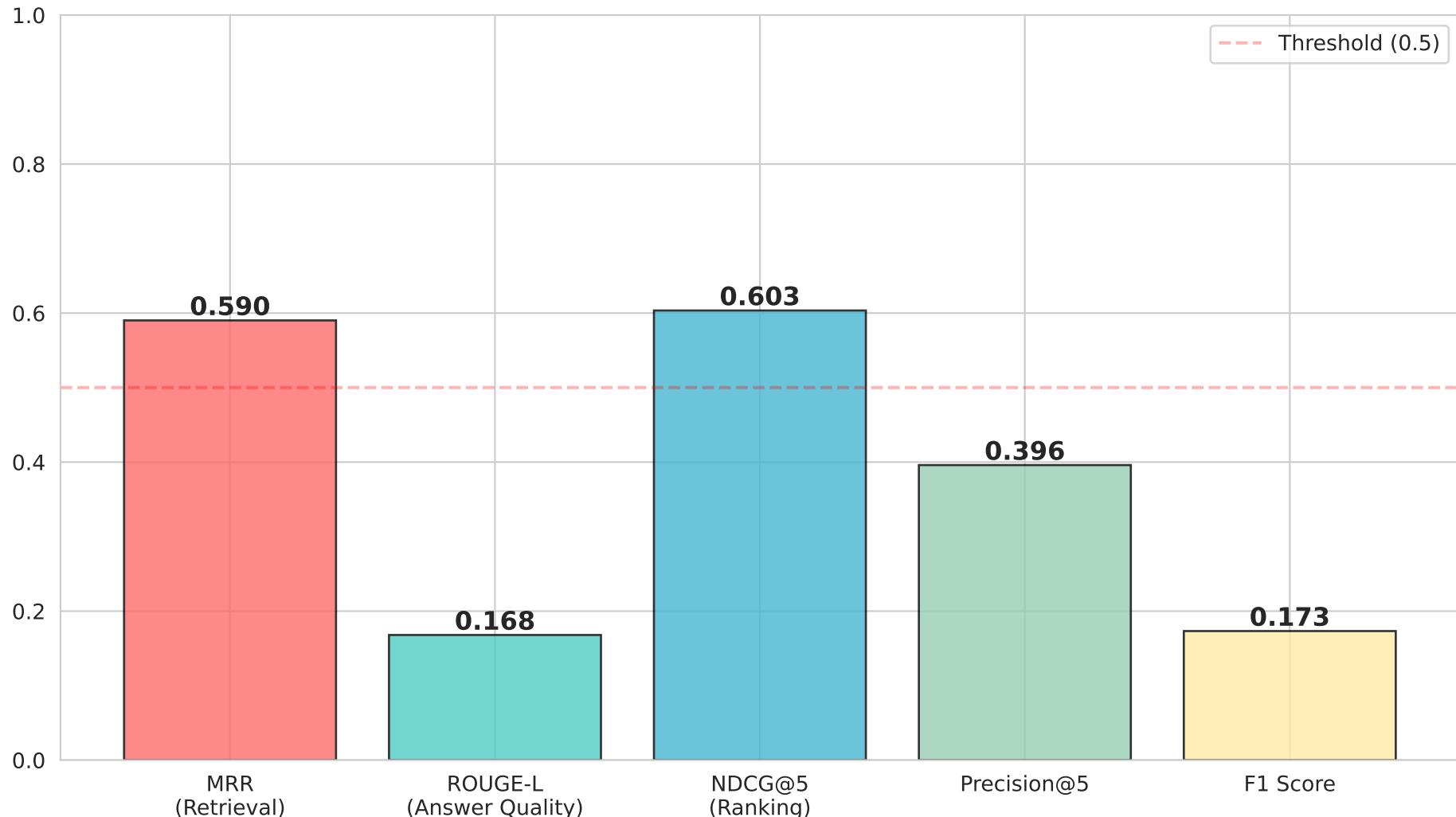
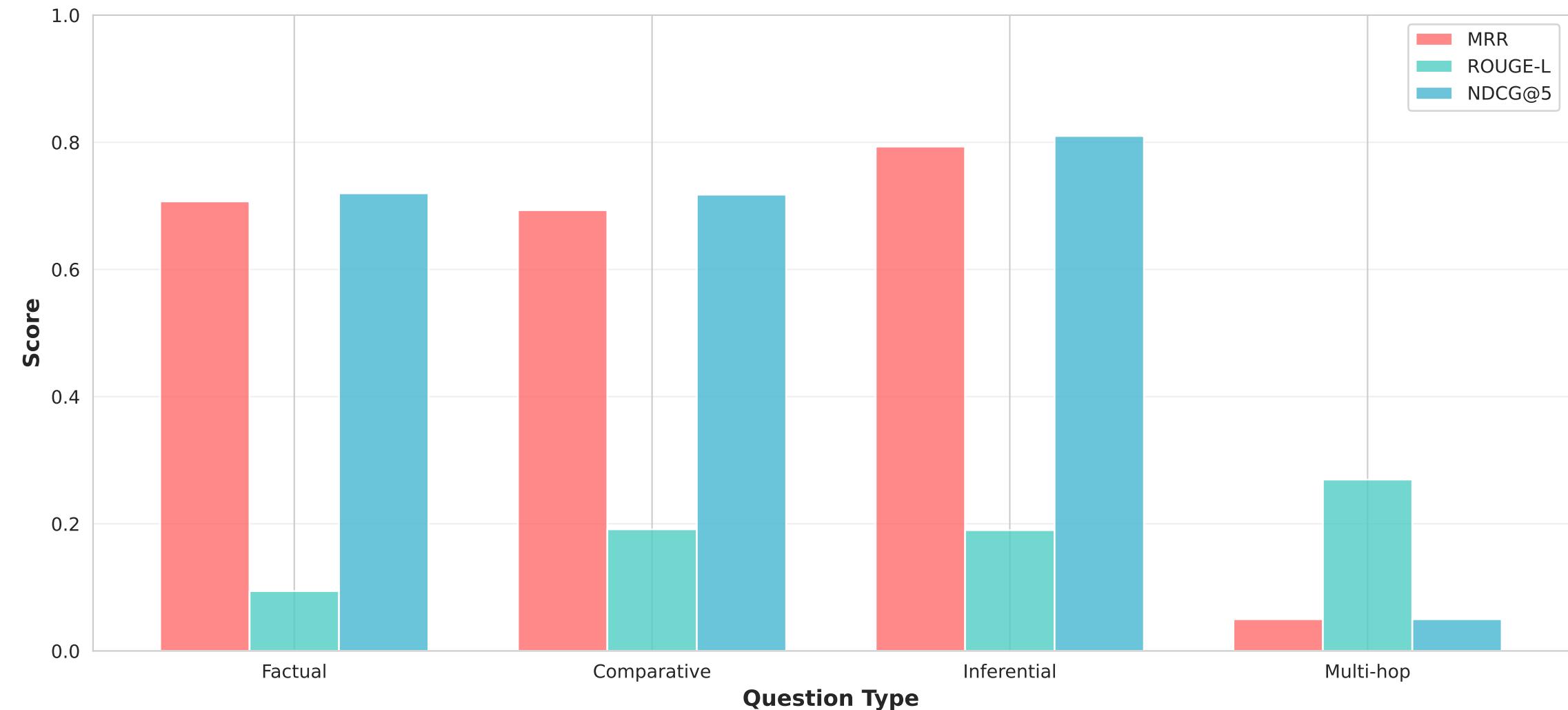


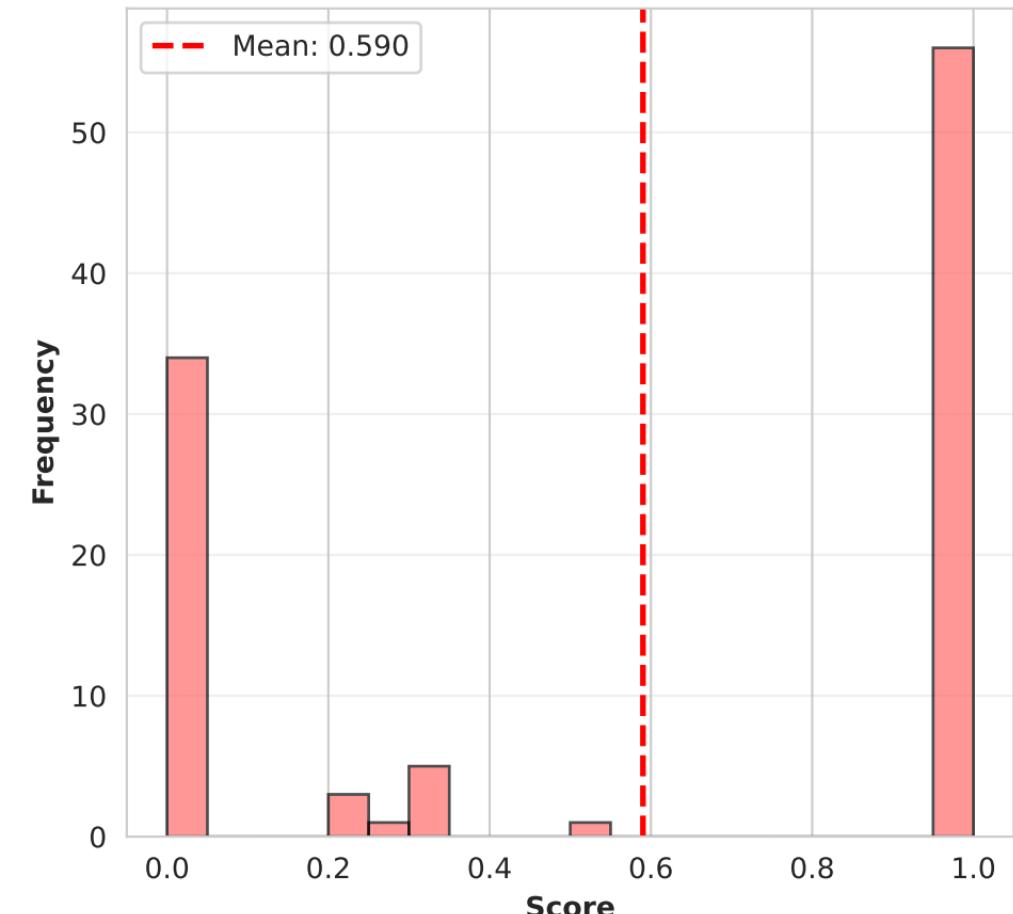
Evaluation Metrics Comparison



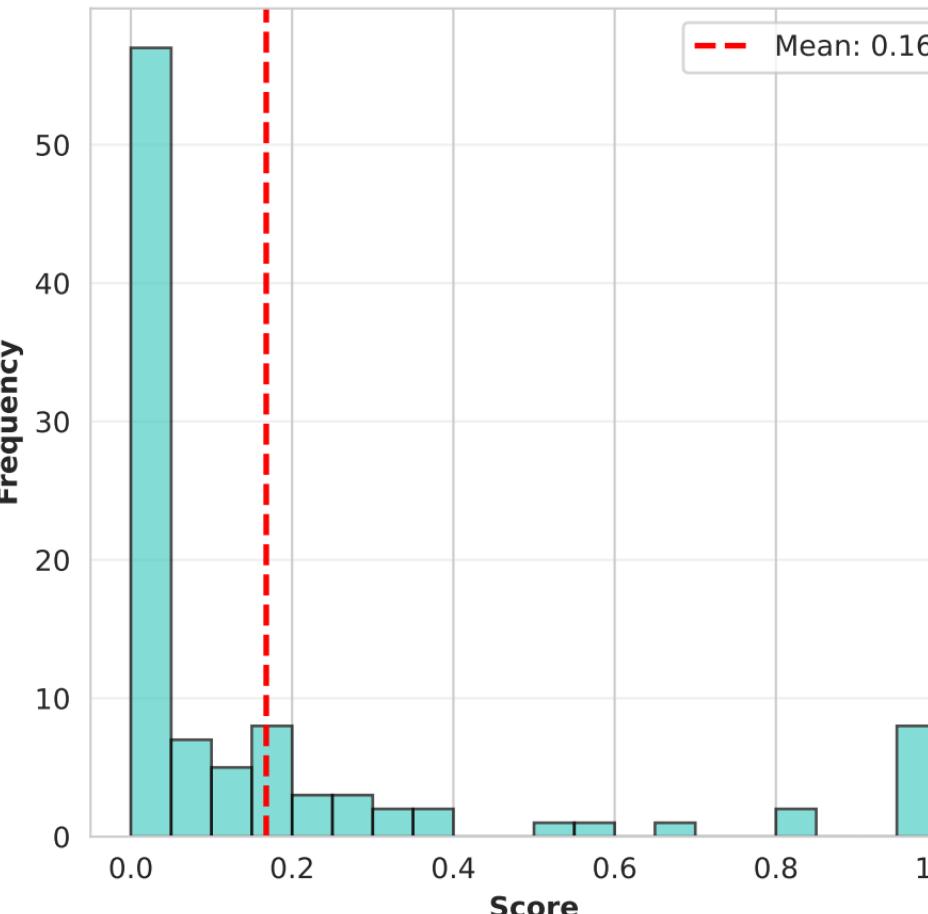
Performance by Question Type



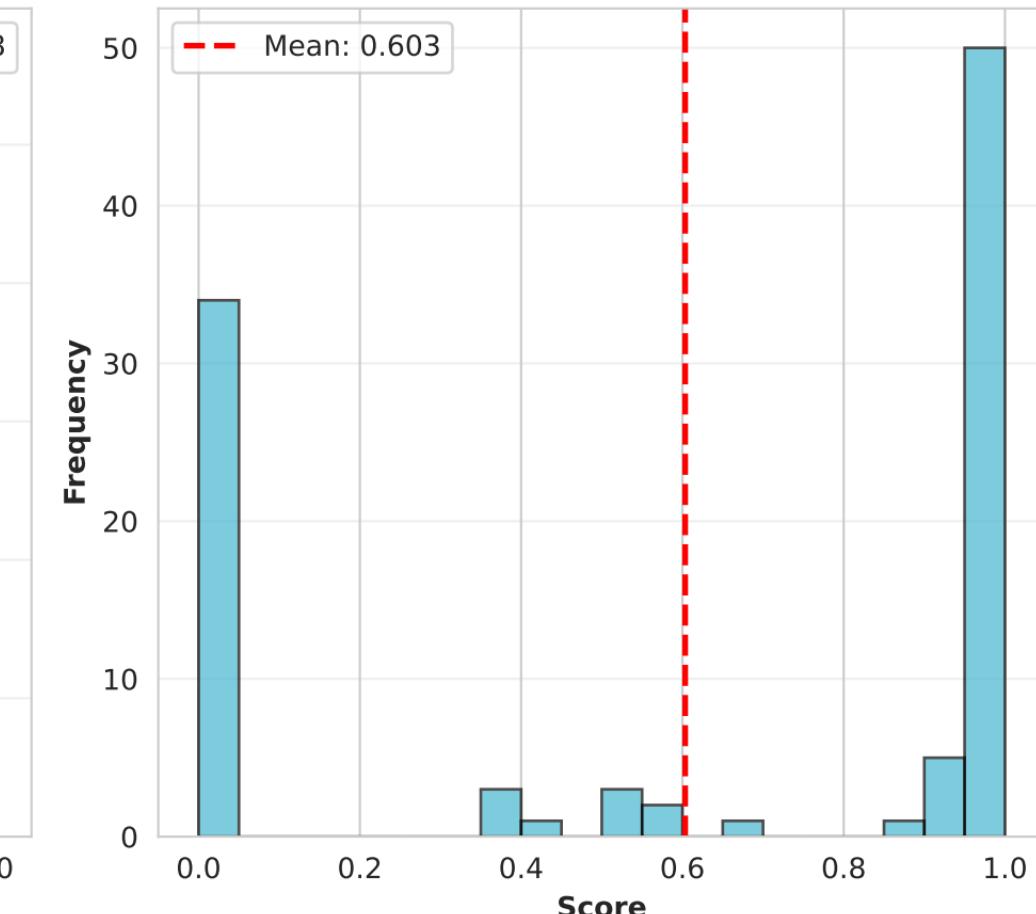
MRR Distribution



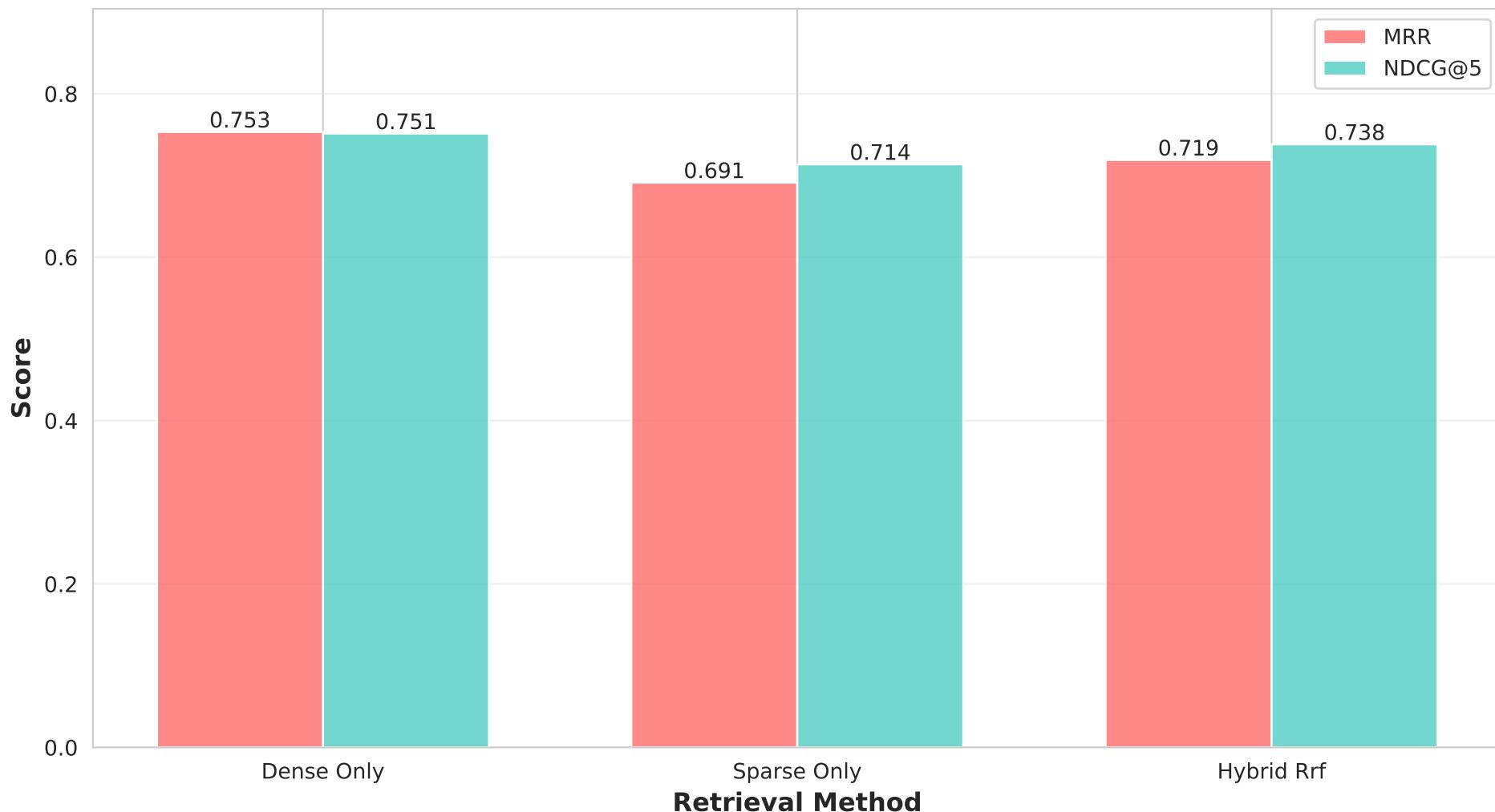
ROUGE-L Distribution



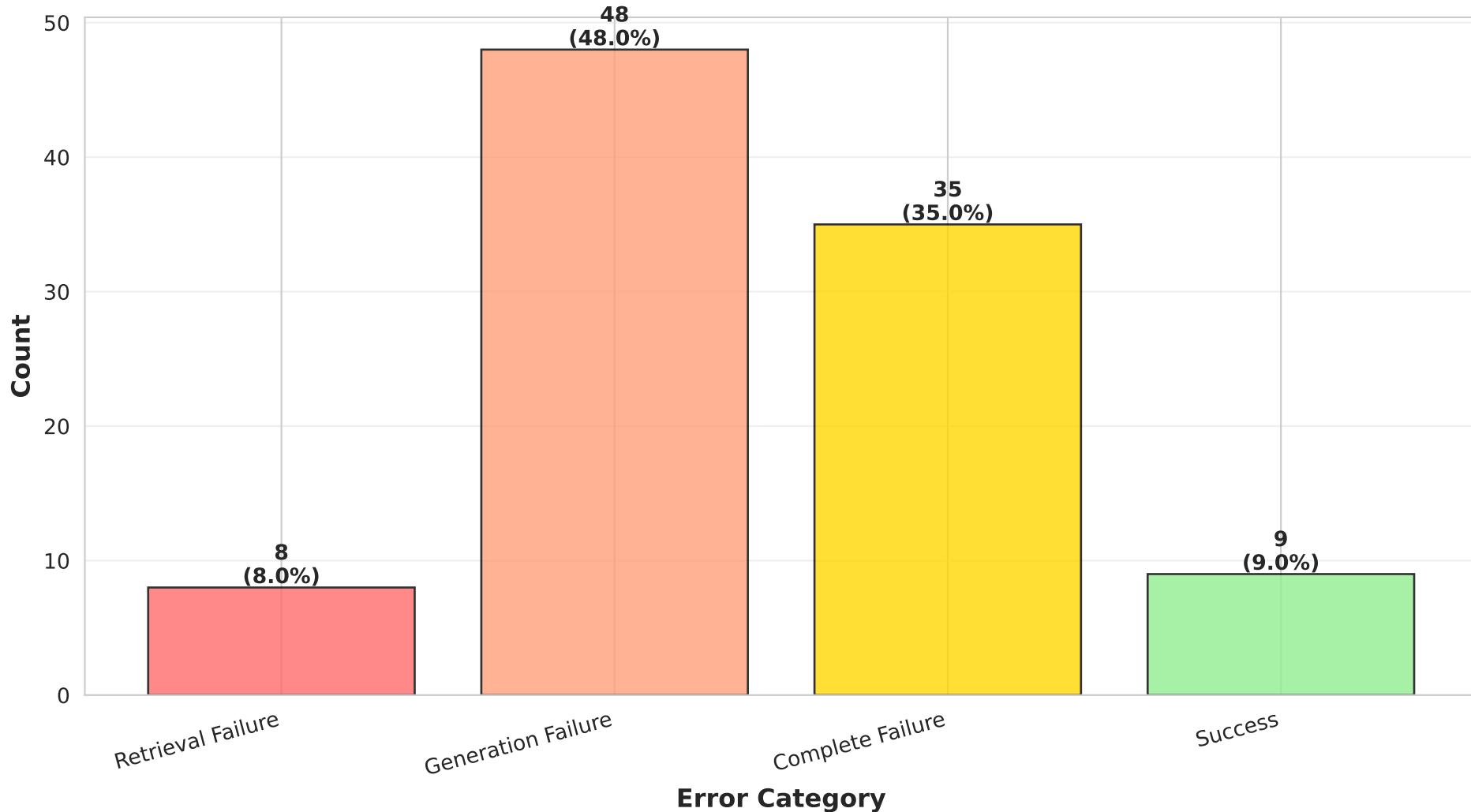
NDCG@5 Distribution



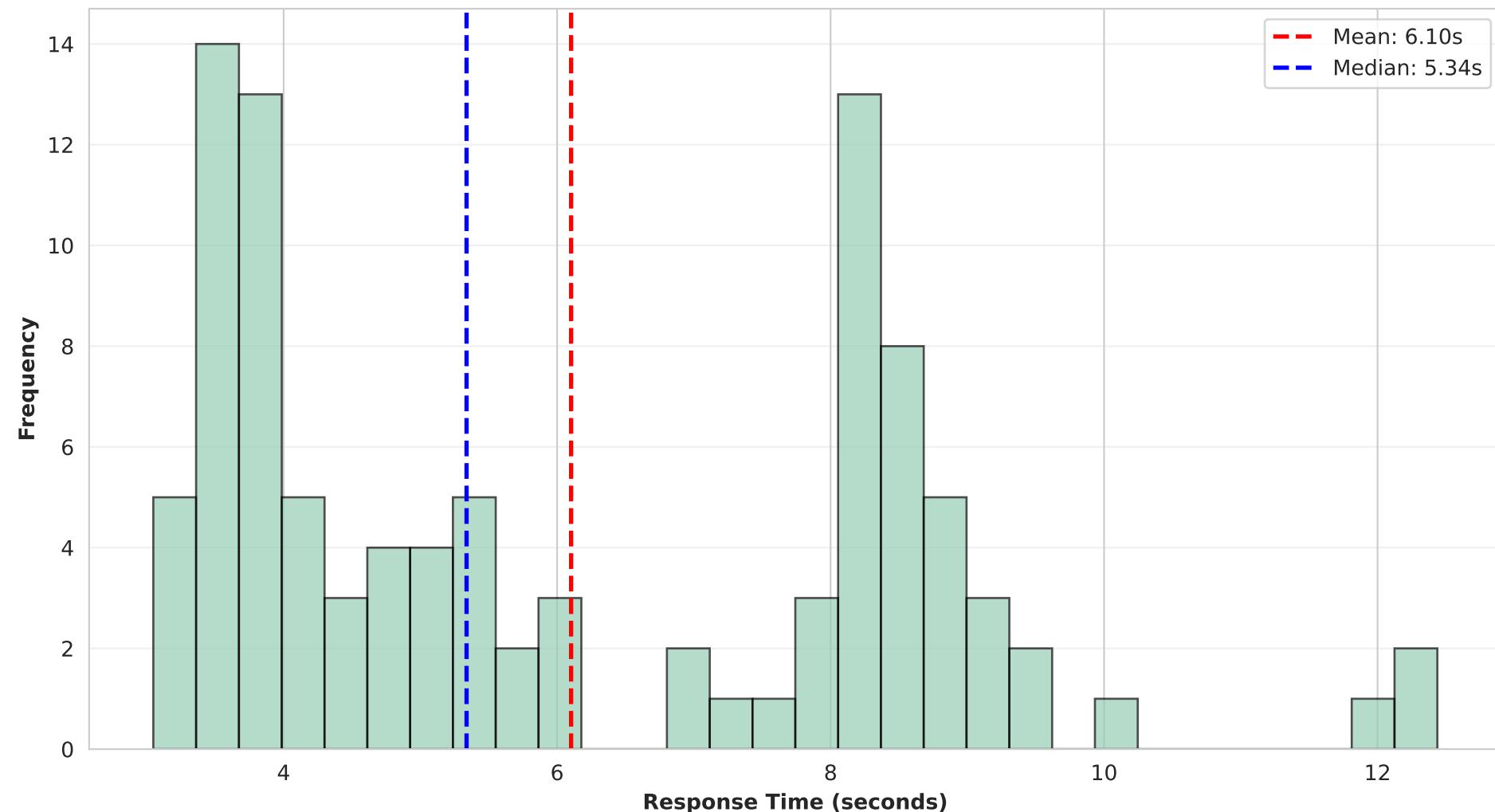
Ablation Study: Retrieval Method Comparison



Error Analysis: Failure Categories



Response Time Distribution



Metric Correlation Matrix

