Table A: **Ablation studies on Router designs of WISE.** Compared with Table 1, a significant decrease in Loc. proves the effectiveness of the Router. "-" indicates the decrease compared with $\text{WISE}_{\text{w. } L_a}$. ZsRE. `LlaMA-2-7B`

$\text{WISE}_{\text{w.o. } L_a}$	Rel.	Gen.	Loc.	Avg.
$T = 1$	1.00	0.96	0.93 -0.07	0.96 -0.01
$T = 10$	0.93	0.90	0.88 -0.12	0.90 -0.04
$T = 100$	0.92	0.85	0.81 -0.19	0.86 -0.04
$T = 1000$	0.84	0.79	0.72 -0.22	0.78 -0.05

Table B: **WISE demonstrates scalability to larger LMs.** Observed surpassing editing performance on the 7B model at $T = 1000$. ZsRE. `LlaMA-2-13B-chat`

	Rel.	Gen.	Loc.	Avg.
$T = 1$	1.00	0.96	1.00	0.99
$T = 10$	0.94	0.89	1.00	0.94
$T = 100$	0.95	0.85	1.00	0.93
$T = 1000$	0.80	0.75	1.00	0.85

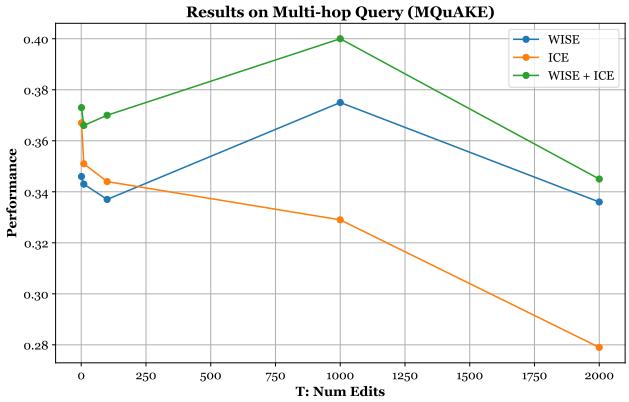

Figure A: Results show that **WISE + ICE (i.e., in-context editing)** (combining both) outperforms using either alone. WISE **demonstrates enhanced multi-hop reasoning at** $T = 1000, 2000$. ZsRE. `LlaMA-2-7B`

Table C: **Results of applying Random Prefix Token to GRACAE.** ZsRE. `LlaMA-2-7B`

	ZsRE				Hallucination	
	Rel.	Gen.	Loc.	Avg.	Rel. (*ppl* ↓)	Loc.
$T = 1$	1.00	0.36 +0.28	1.00	0.79	2.21	1.00
$T = 10$	0.99	0.15 +0.15	1.00	0.71	8.67	1.00
$T = 100$	0.98	0.15 +0.15	1.00	0.71	9.34	1.00
$T = 1000$	0.93	0.08 +0.00	1.00	0.67	9.74	1.00

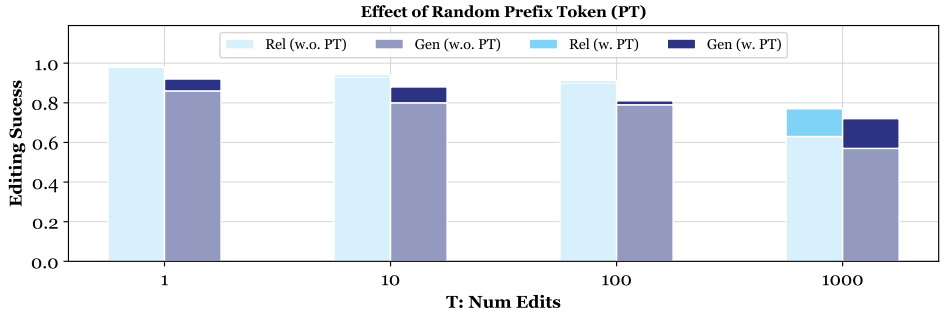

Figure B: **Ablation studies on Random Prefix Token (PT) of WISE.** Light/Dark colors indicate the Editing Sucess w.o./w. PT addition, demonstrating enhanced paraphrase comprehension and leveraging editing facts across different contexts. ZsRE. `LlaMA-2-7B`