Accuracy over Questions - comparative-1-gpt4o-gsm8k