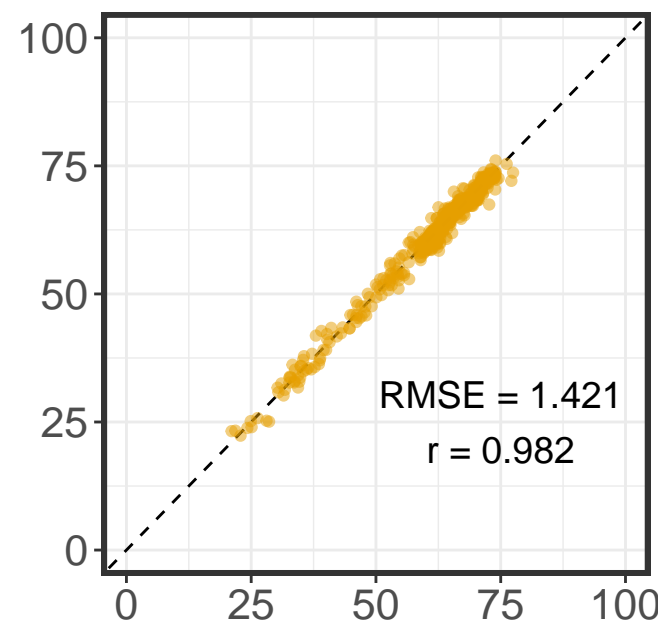
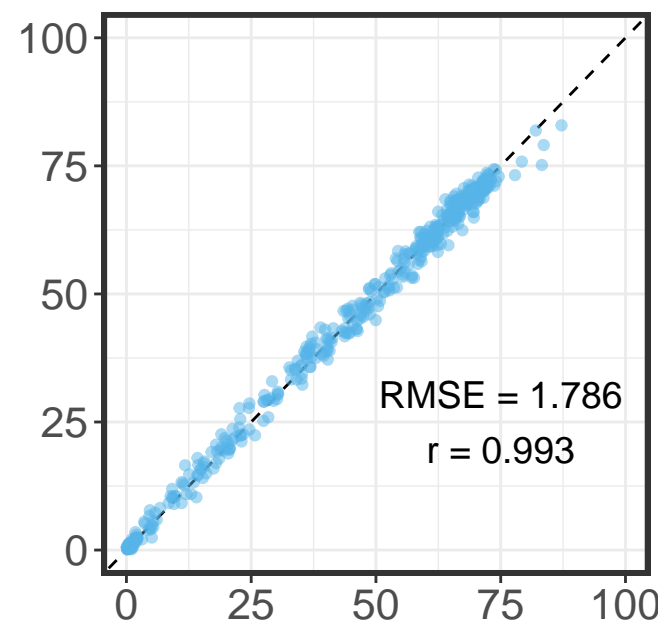


**A**

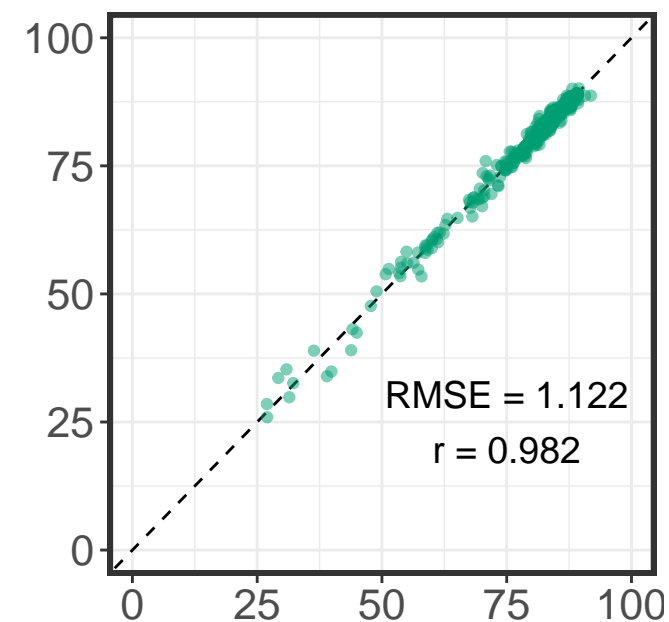
ARC



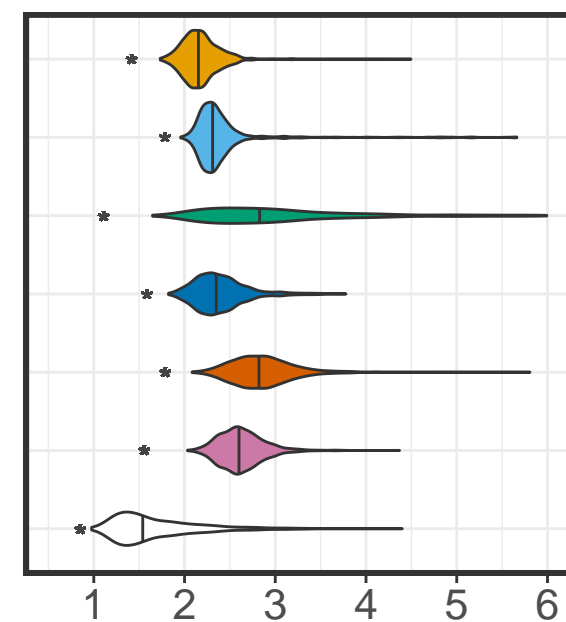
GSM8K



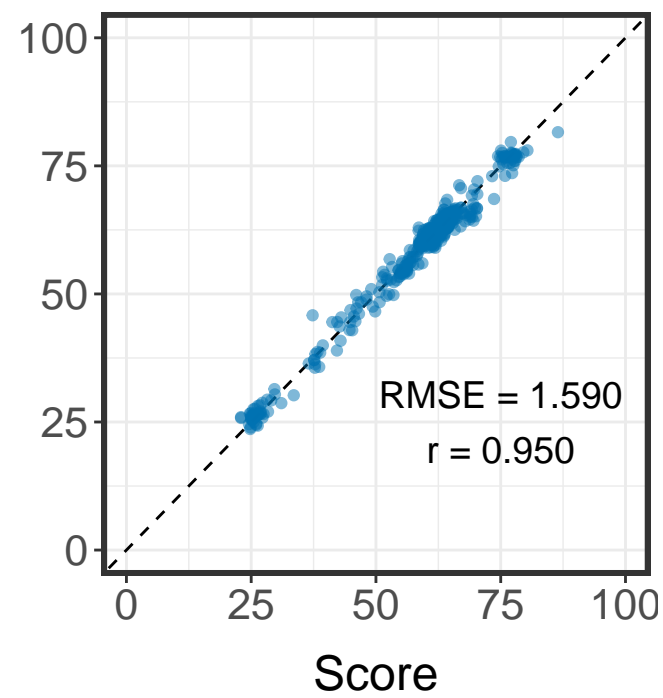
HellaSwag

**B**

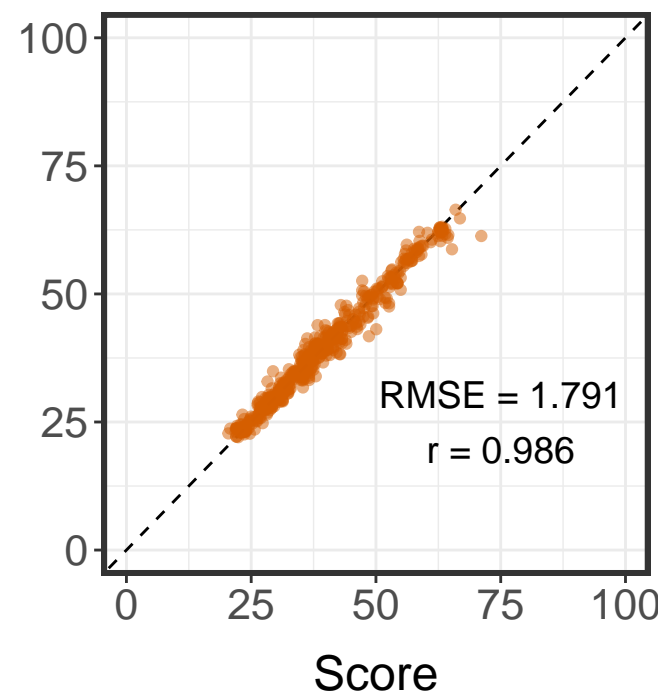
Random



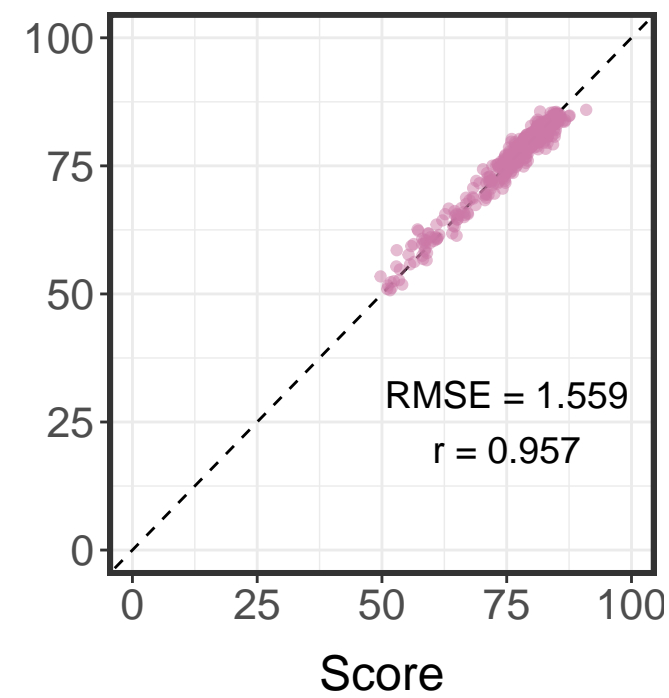
MMLU



TruthfulQA



Winogrande

**c**

metabench (d = 845)

