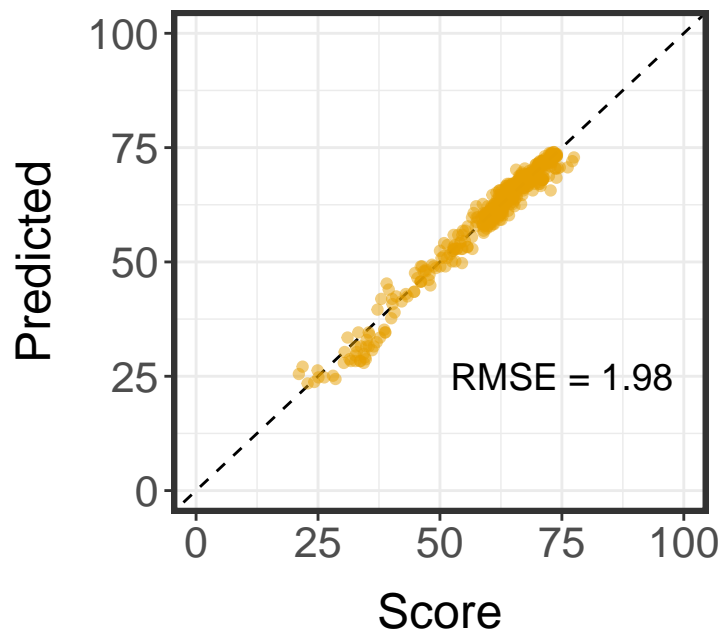
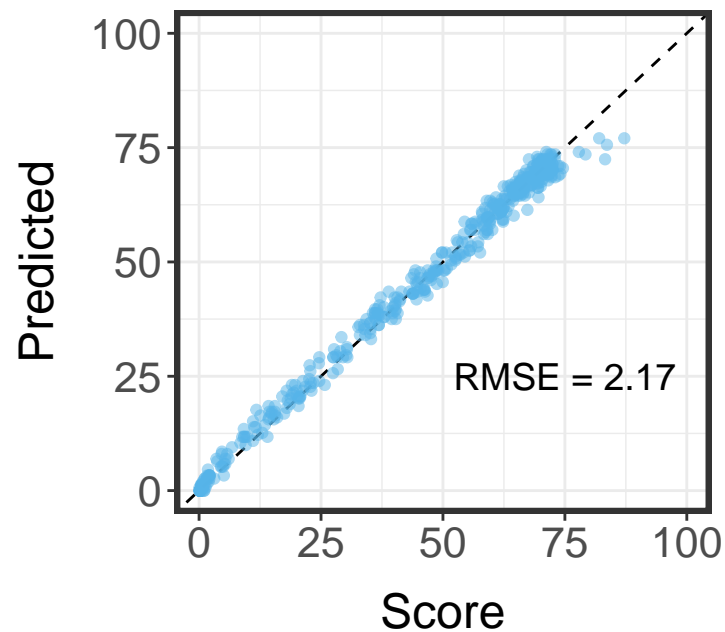


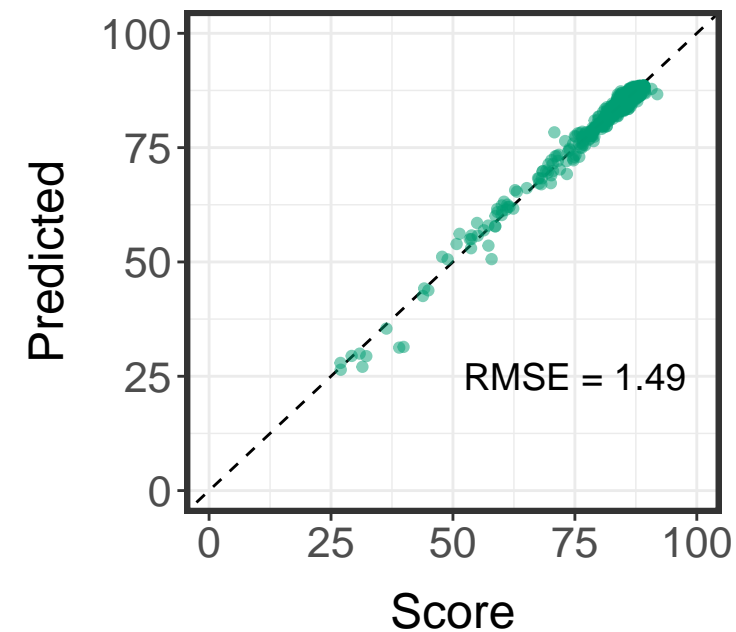
ARC (n = 150)



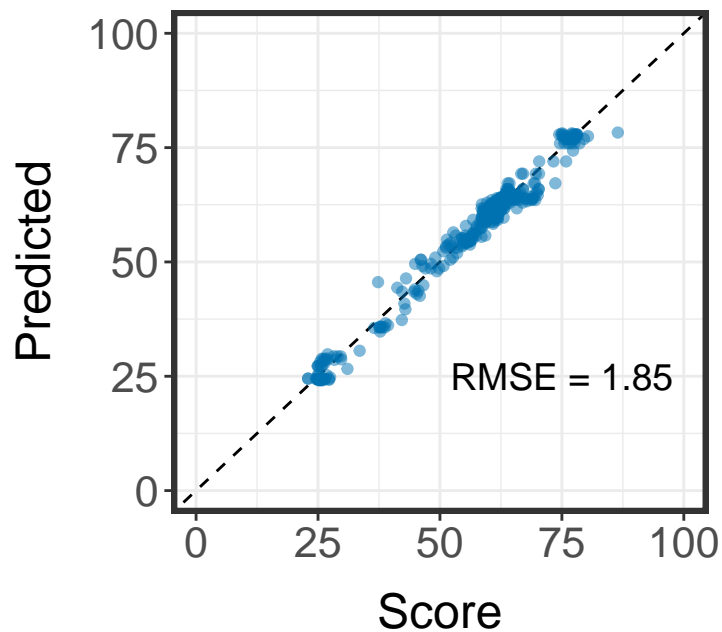
GSM8K (n = 189)



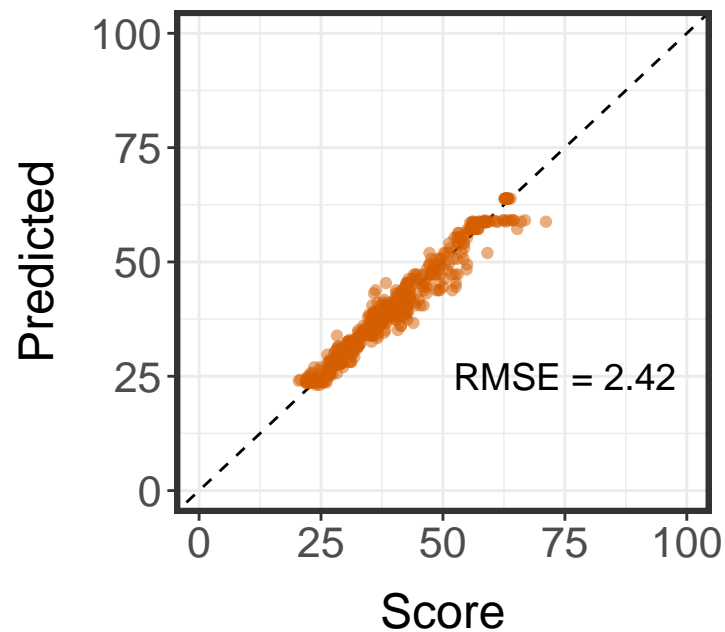
HellaSwag (n = 200)



MMLU (n = 141)



TruthfulQA (n = 65)



Winogrande (n = 100)

