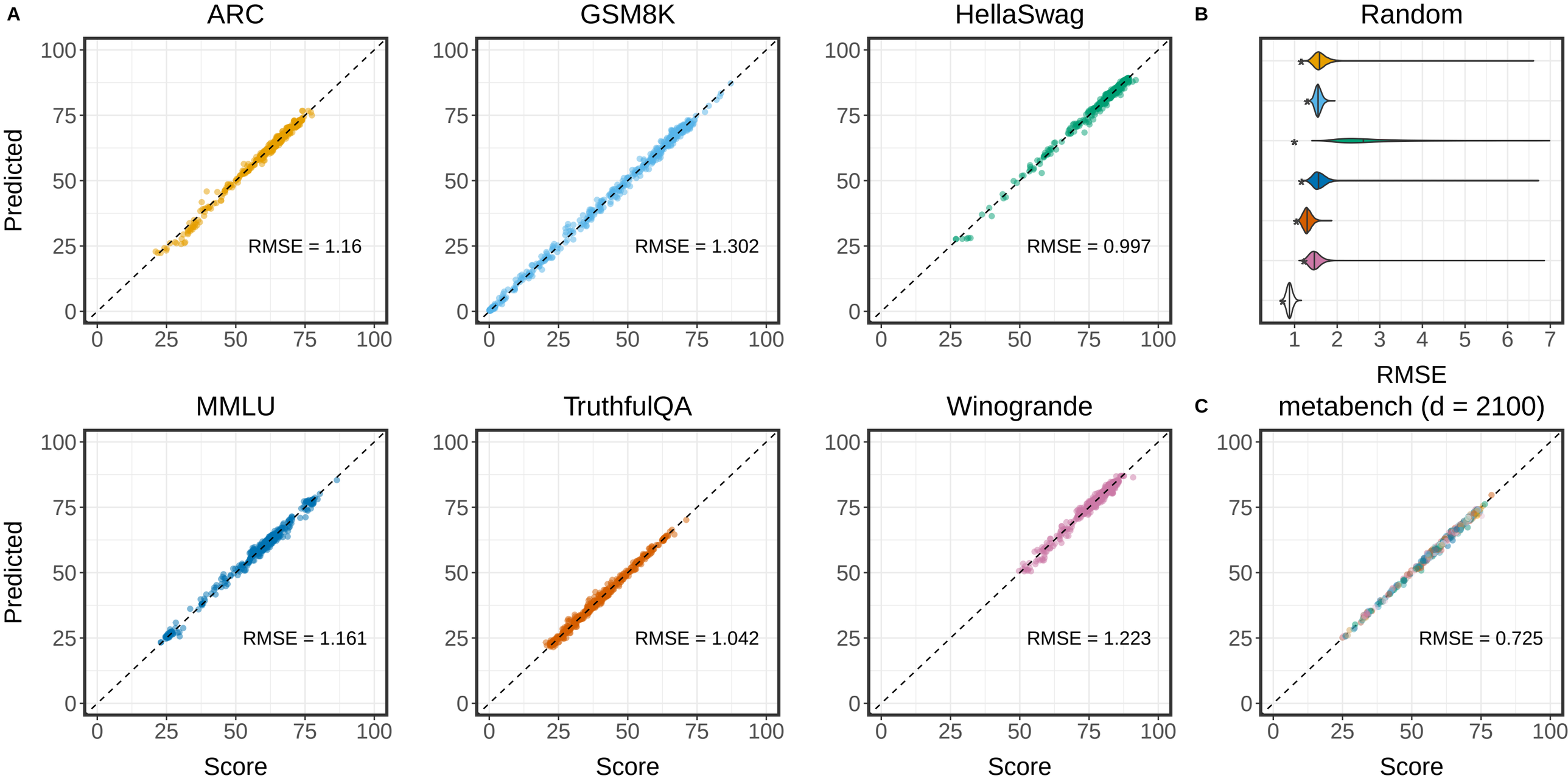
A
ARC
GSM8K
HellaSwag
B
Random
MMLU
TruthfulQA
Winogrande
C
metabench (d = 2100)
Predicted
Score
RMSE
RMSE = 1.16
RMSE = 1.302
RMSE = 0.997
RMSE = 1.161
RMSE = 1.042
RMSE = 1.223
RMSE = 0.725
0
25
50
75
100
1
2
3
4
5
6
7