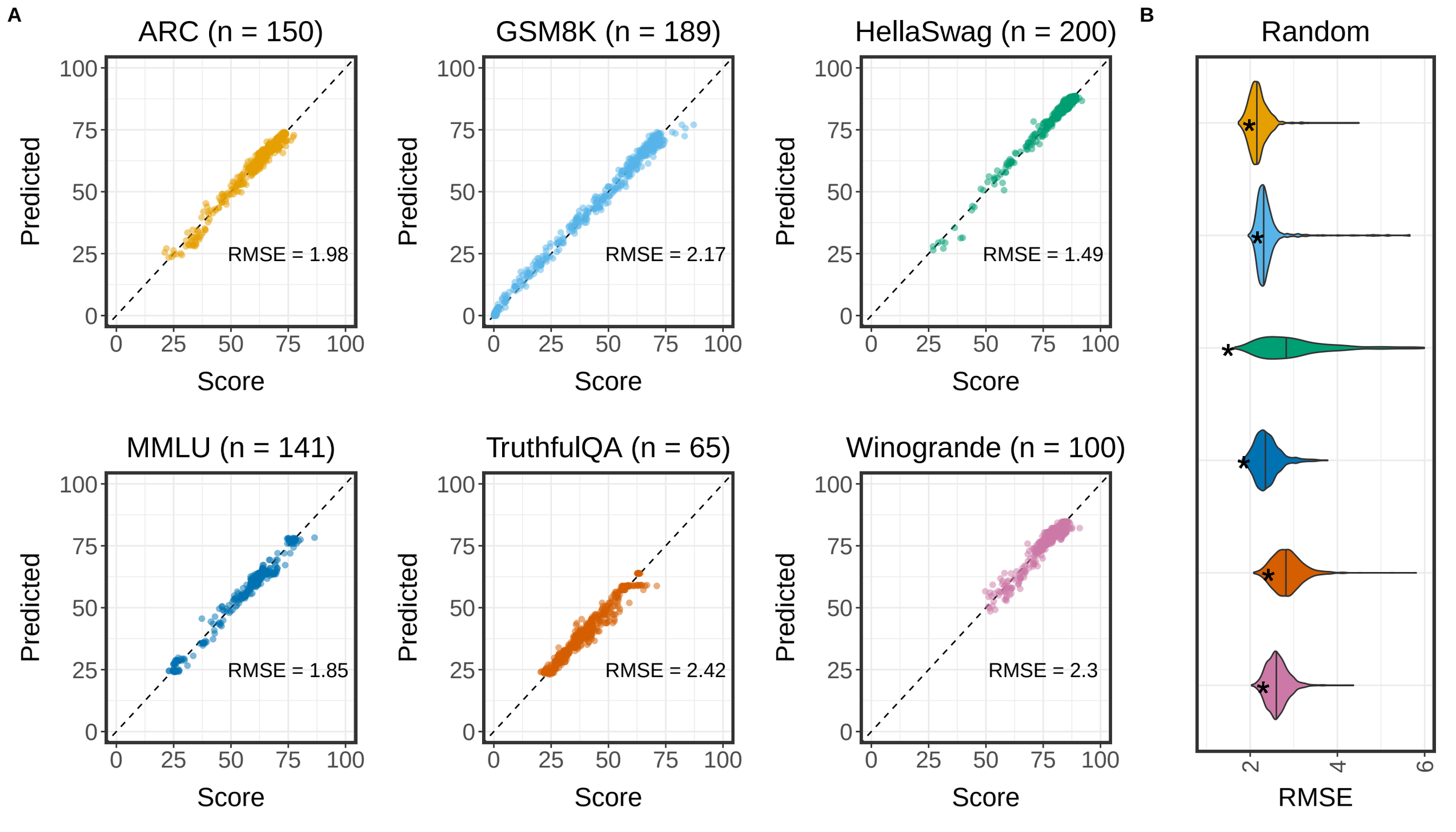
A
ARC (n = 150)
RMSE = 1.98
GSM8K (n = 189)
RMSE = 2.17
HellaSwag (n = 200)
RMSE = 1.49
MMLU (n = 141)
RMSE = 1.85
TruthfulQA (n = 65)
RMSE = 2.42
Winogrande (n = 100)
RMSE = 2.3
Predicted
Score
0
25
50
75
100
B
Random
2
4
6
RMSE