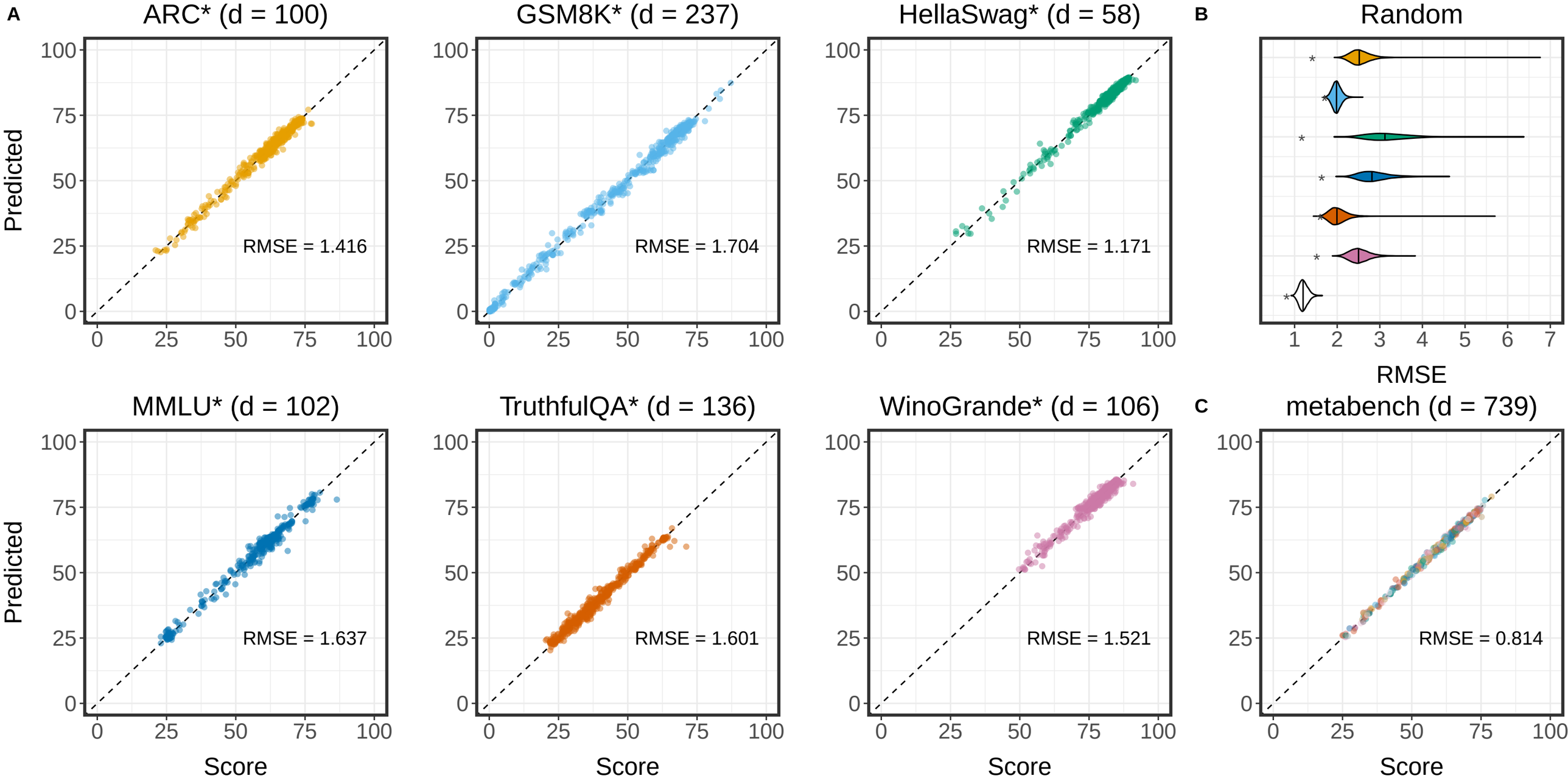
A
ARC* (d = 100)
GSM8K* (d = 237)
HellaSwag* (d = 58)
B
Random
MMLU* (d = 102)
TruthfulQA* (d = 136)
WinoGrande* (d = 106)
C
metabench (d = 739)
RMSE = 1.416
RMSE = 1.704
RMSE = 1.171
RMSE = 1.637
RMSE = 1.601
RMSE = 1.521
RMSE = 0.814
Predicted
Score
RMSE