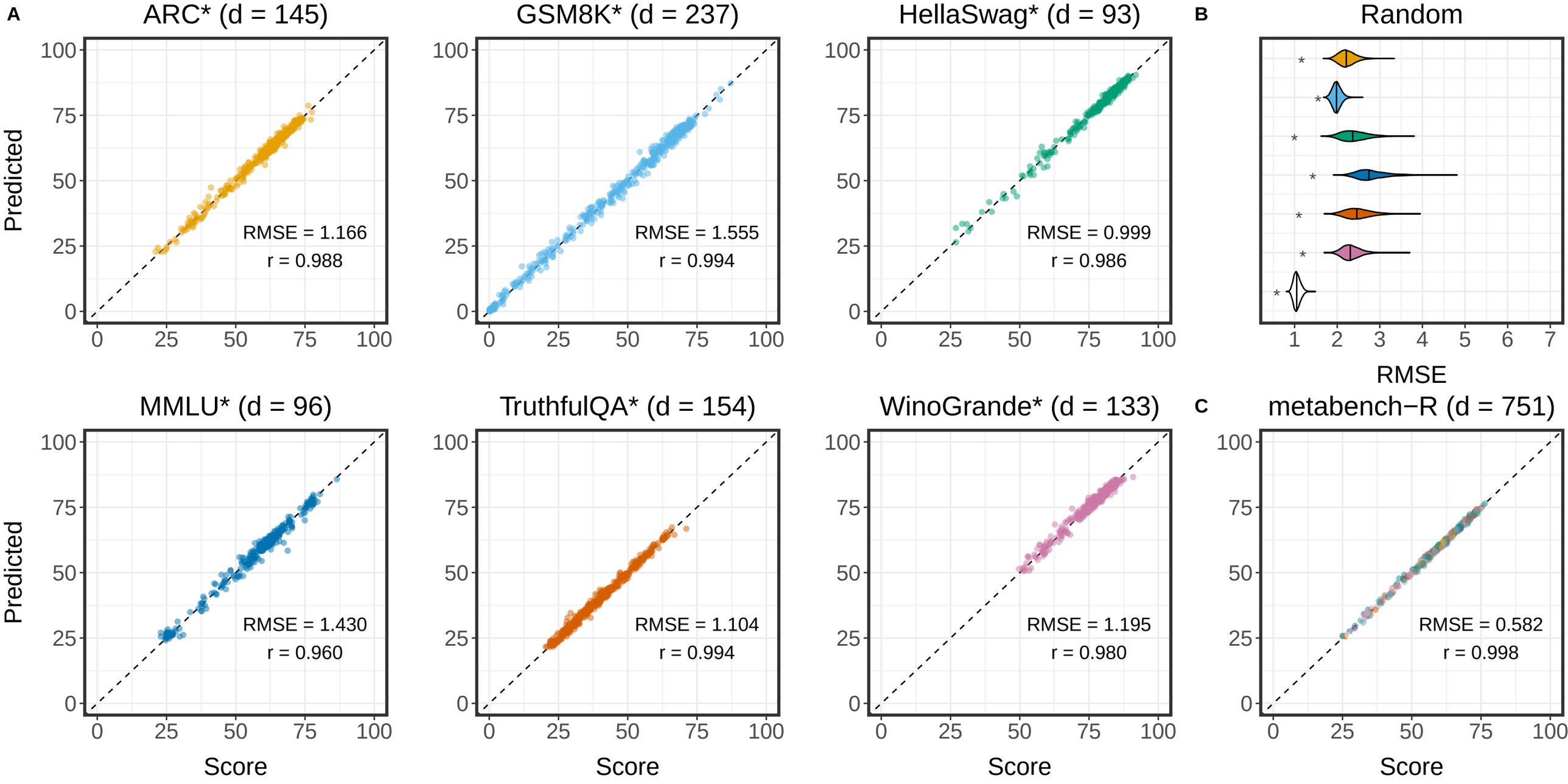
A
ARC* (d = 145)
RMSE = 1.166
r = 0.988
GSM8K* (d = 237)
RMSE = 1.555
r = 0.994
HellaSwag* (d = 93)
RMSE = 0.999
r = 0.986
B
Random
RMSE
MMLU* (d = 96)
RMSE = 1.430
r = 0.960
TruthfulQA* (d = 154)
RMSE = 1.104
r = 0.994
WinoGrande* (d = 133)
RMSE = 1.195
r = 0.980
C
metabench−R (d = 751)
RMSE = 0.582
r = 0.998
Predicted
Score