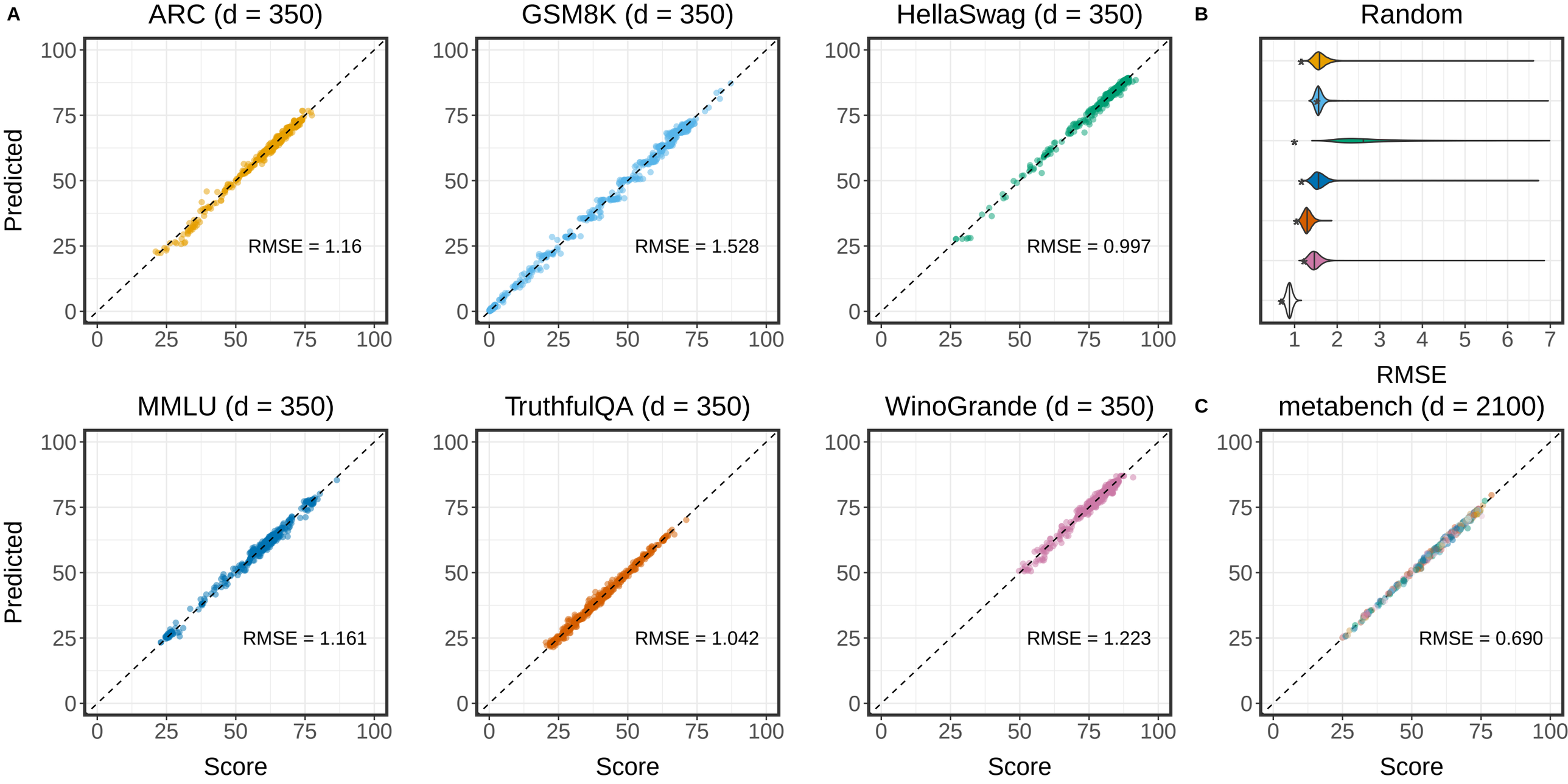
A
ARC (d = 350)
GSM8K (d = 350)
HellaSwag (d = 350)
B
Random
RMSE
MMLU (d = 350)
TruthfulQA (d = 350)
WinoGrande (d = 350)
C
metabench (d = 2100)
Predicted
Score
RMSE = 1.16
RMSE = 1.528
RMSE = 0.997
RMSE = 1.161
RMSE = 1.042
RMSE = 1.223
RMSE = 0.690