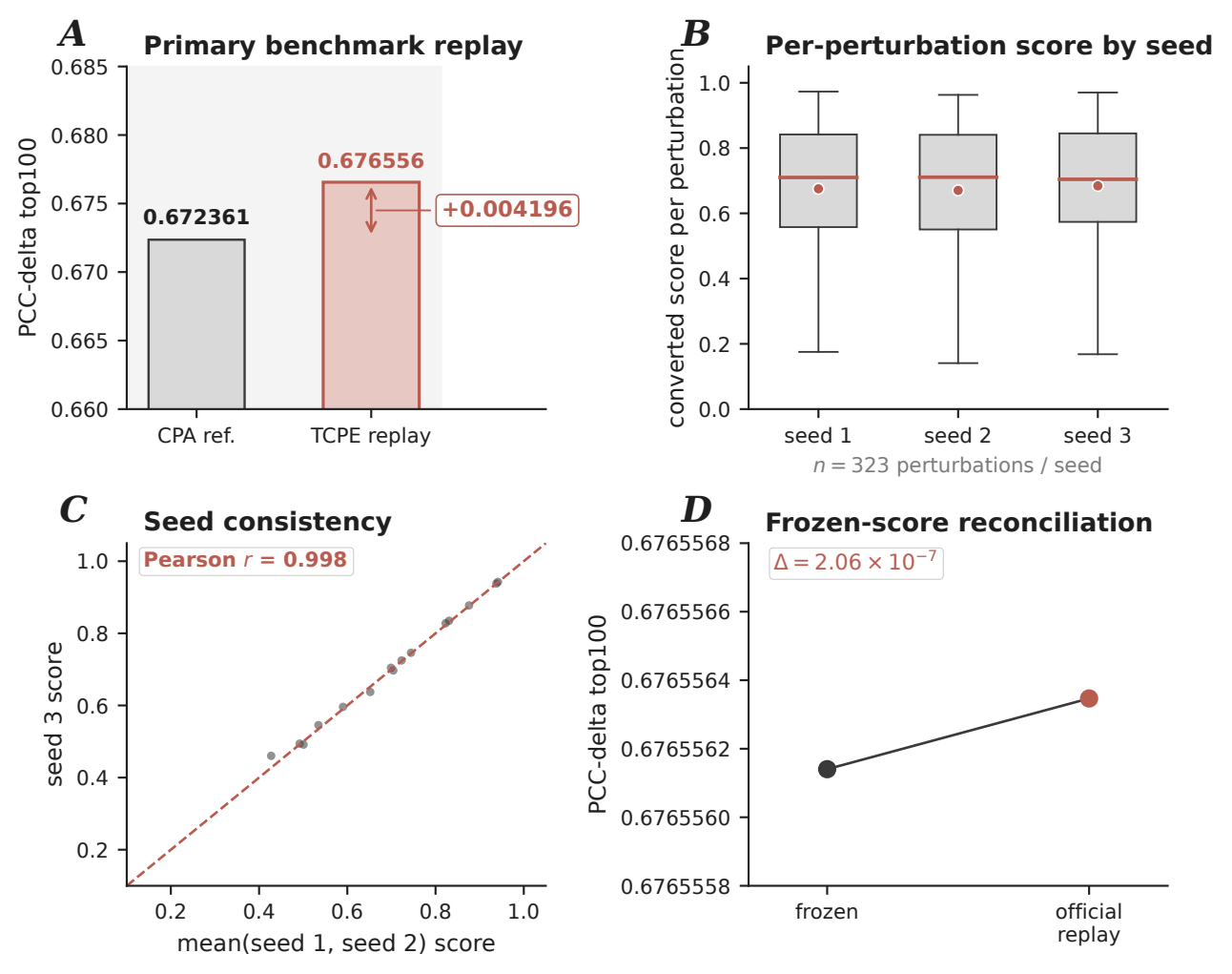
A
Primary benchmark replay
PCC-delta top100
0.685
0.680
0.675
0.670
0.665
0.660
0.672361
0.676556
+0.004196
CPA ref.
TCPE replay
B
Per-perturbation score by seed
converted score per perturbation
1.0
0.8
0.6
0.4
0.2
0.0
seed 1
seed 2
seed 3
n = 323 perturbations / seed
C
Seed consistency
Pearson r = 0.998
seed 3 score
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
mean(seed 1, seed 2) score
D
Frozen-score reconciliation
Δ = 2.06 × 10−7
PCC-delta top100
0.6765568
0.6765566
0.6765564
0.6765562
0.6765560
0.6765558
frozen
official replay