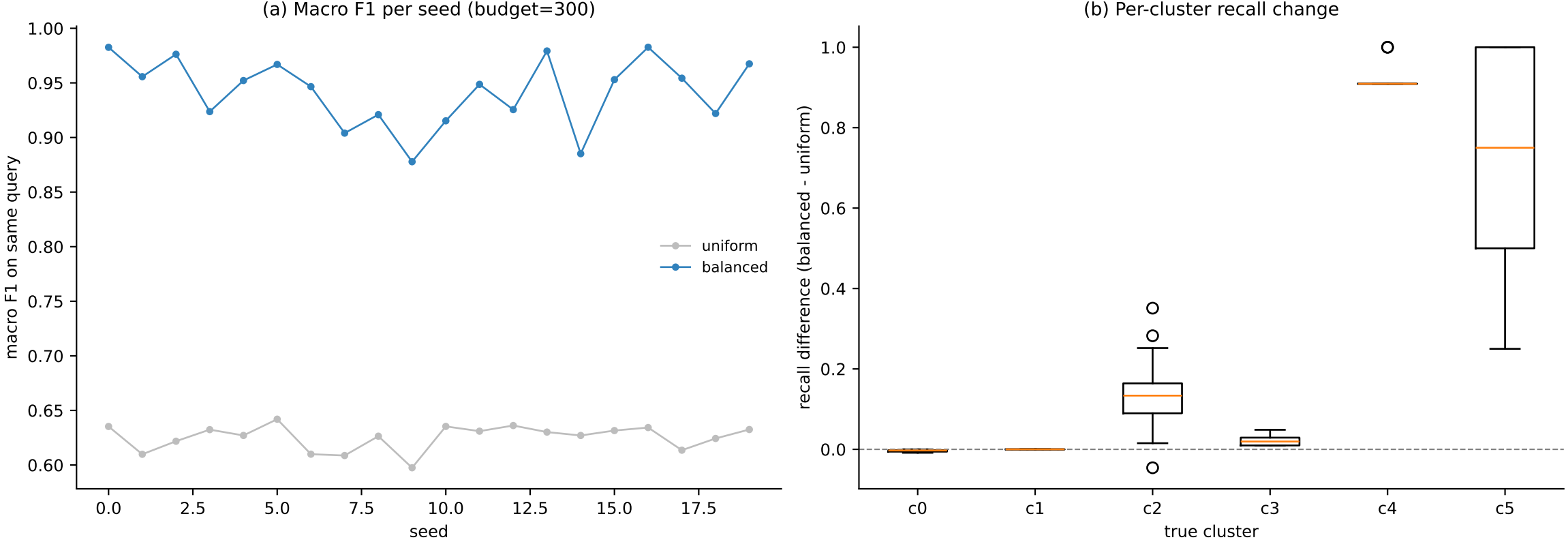
(a) Macro F1 per seed (budget=300)
macro F1 on same query
seed
uniform
balanced
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
(b) Per-cluster recall change
recall difference (balanced - uniform)
true cluster
1.0
0.8
0.6
0.4
0.2
0.0
c0
c1
c2
c3
c4
c5