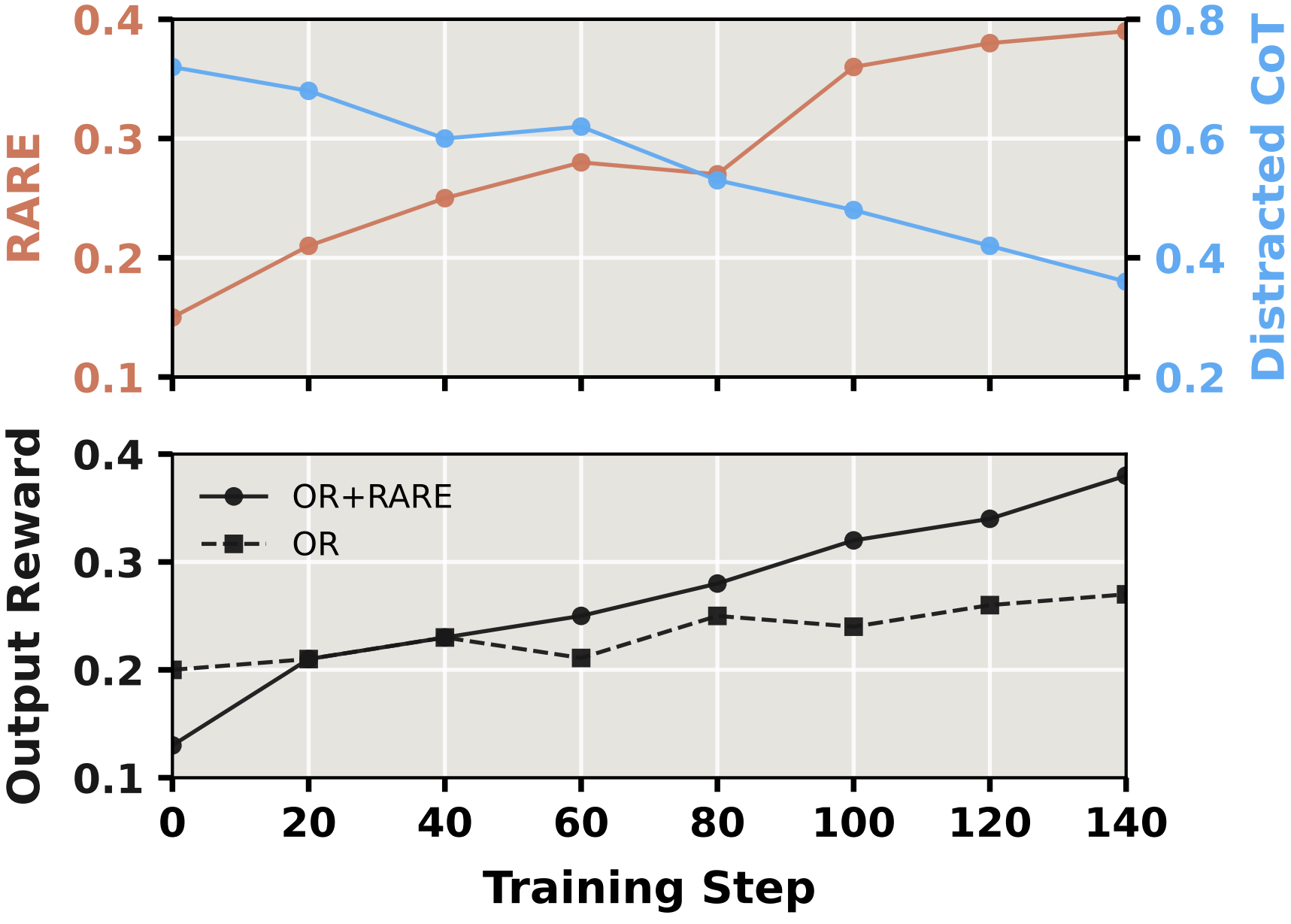

RARE
0.4
0.3
0.2
0.1
Distracted CoT
0.8
0.6
0.4
0.2
Output Reward
0.4
0.3
0.2
0.1
OR+RARE
OR
0
20
40
60
80
100
120
140
Training Step