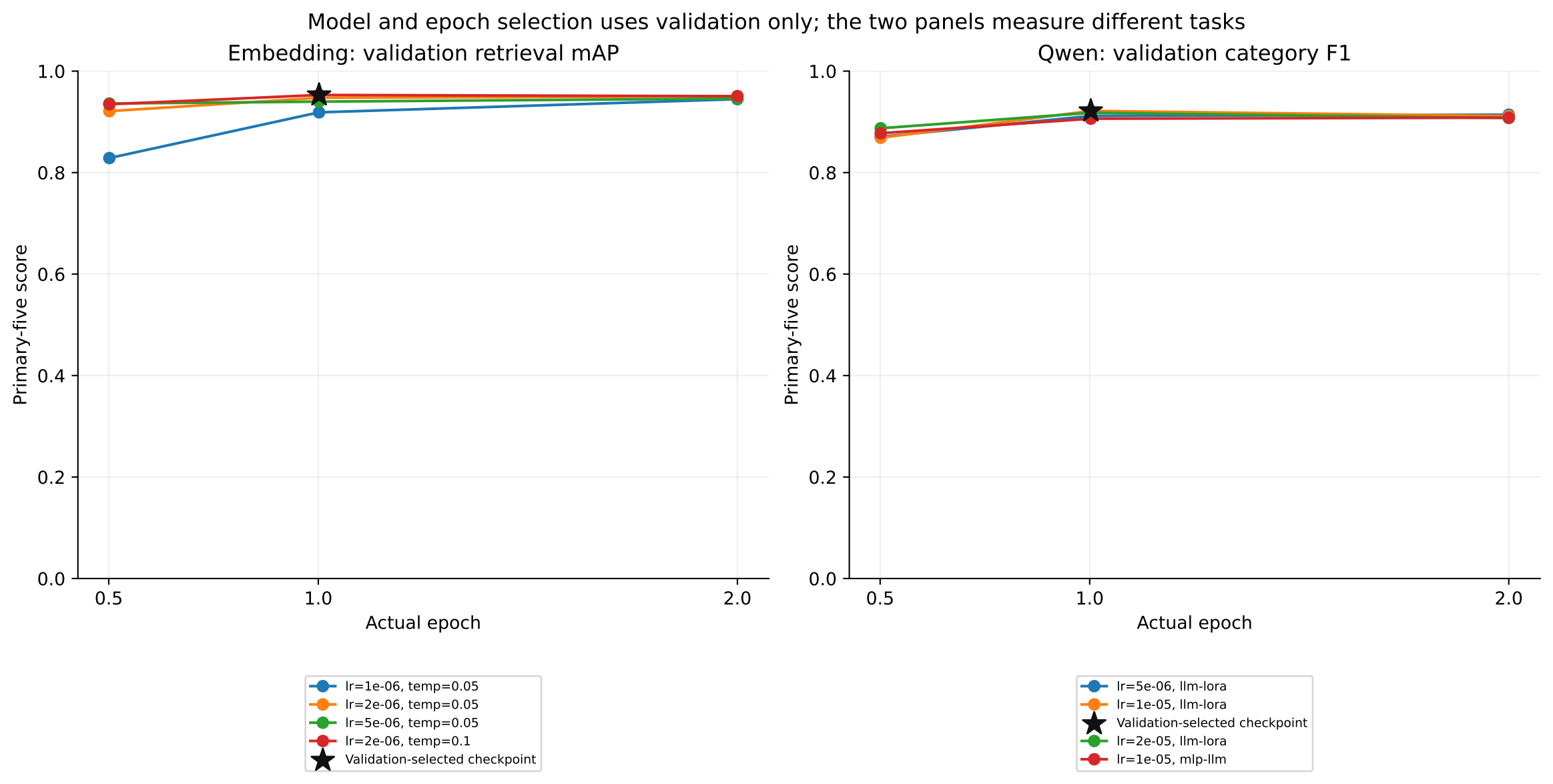

Model and epoch selection uses validation only; the two panels measure different tasks
Embedding: validation retrieval mAP
Qwen: validation category F1
Primary-five score
Actual epoch
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
2.0
lr=1e-06, temp=0.05
lr=2e-06, temp=0.05
lr=5e-06, temp=0.05
lr=2e-06, temp=0.1
Validation-selected checkpoint
lr=5e-06, llm-lora
lr=1e-05, llm-lora
Validation-selected checkpoint
lr=2e-05, llm-lora
lr=1e-05, mlp-llm