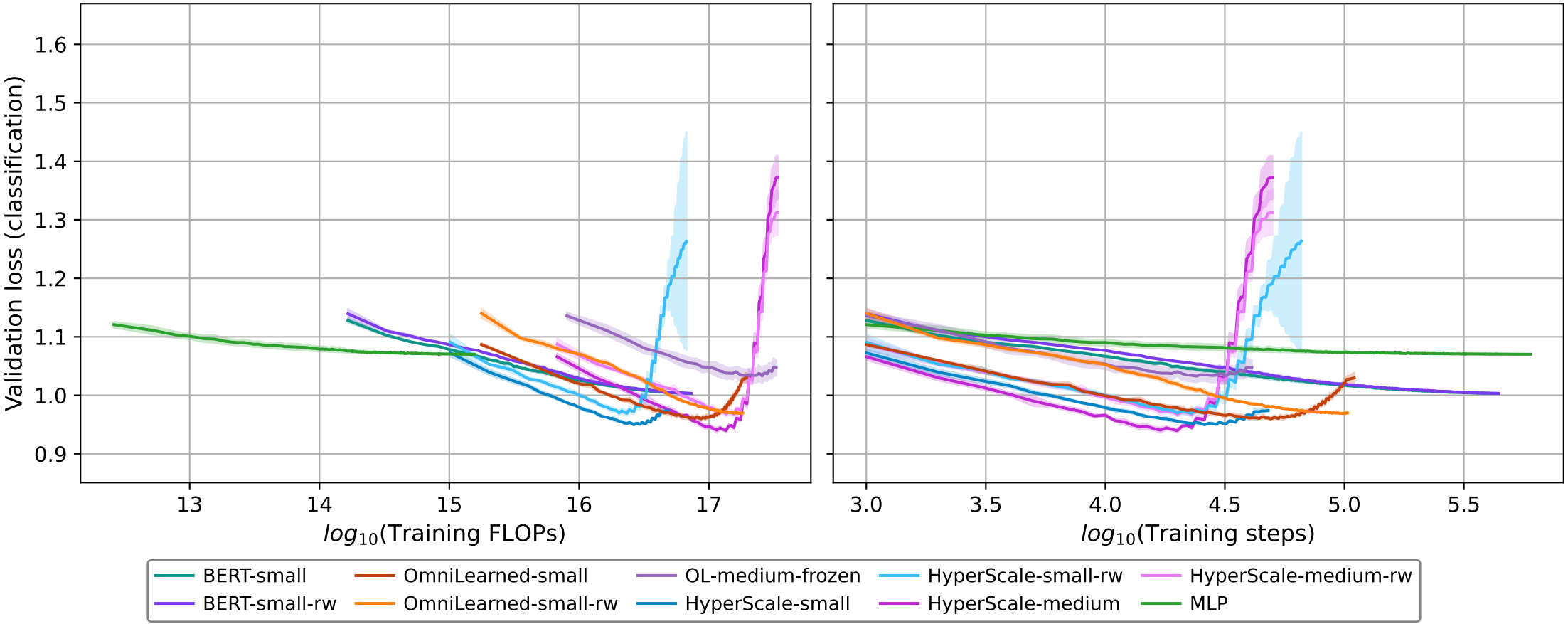

Validation loss (classification)
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
13
14
15
16
17
log10(Training FLOPs)
3.0
3.5
4.0
4.5
5.0
5.5
log10(Training steps)
BERT-small
BERT-small-rw
OmniLearned-small
OmniLearned-small-rw
OL-medium-frozen
HyperScale-small
HyperScale-small-rw
HyperScale-medium
HyperScale-medium-rw
MLP