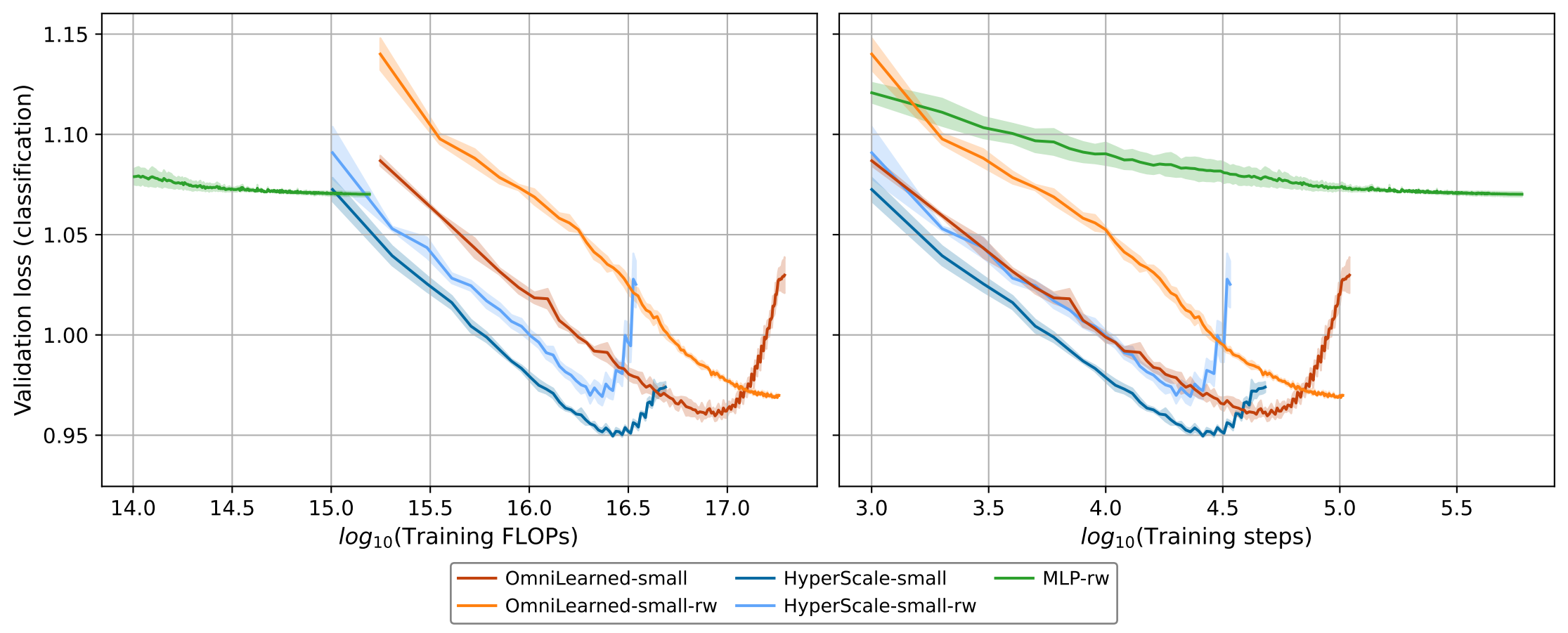

Validation loss (classification)
1.15
1.10
1.05
1.00
0.95
14.0
14.5
15.0
15.5
16.0
16.5
17.0
log10(Training FLOPs)
3.0
3.5
4.0
4.5
5.0
5.5
log10(Training steps)
OmniLearned-small
OmniLearned-small-rw
HyperScale-small
HyperScale-small-rw
MLP-rw