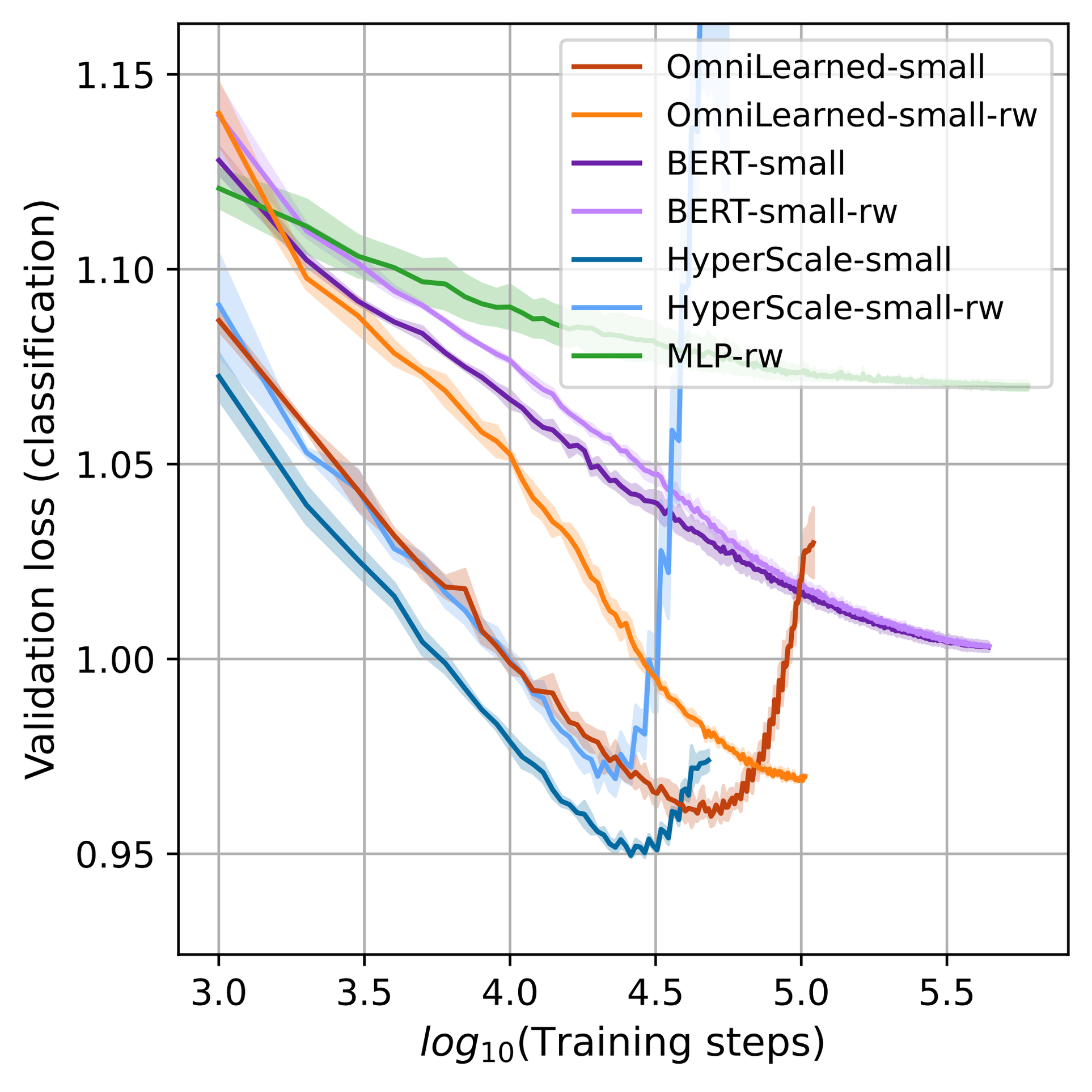

OmniLearned-small
OmniLearned-small-rw
BERT-small
BERT-small-rw
HyperScale-small
HyperScale-small-rw
MLP-rw
Validation loss (classification)
1.15
1.10
1.05
1.00
0.95
3.0
3.5
4.0
4.5
5.0
5.5
log10(Training steps)