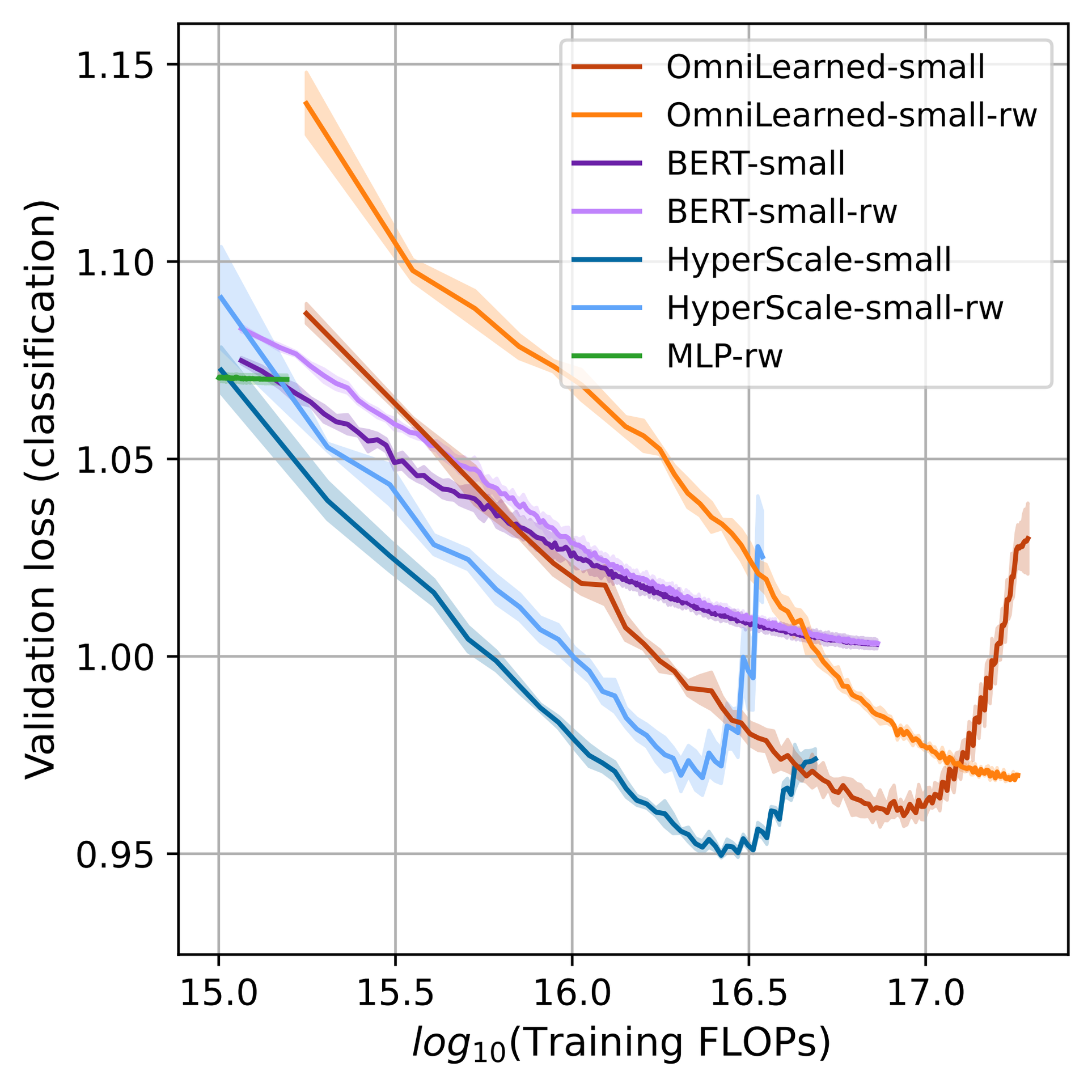

OmniLearned-small
OmniLearned-small-rw
BERT-small
BERT-small-rw
HyperScale-small
HyperScale-small-rw
MLP-rw
Validation loss (classification)
1.15
1.10
1.05
1.00
0.95
15.0
15.5
16.0
16.5
17.0
log10(Training FLOPs)