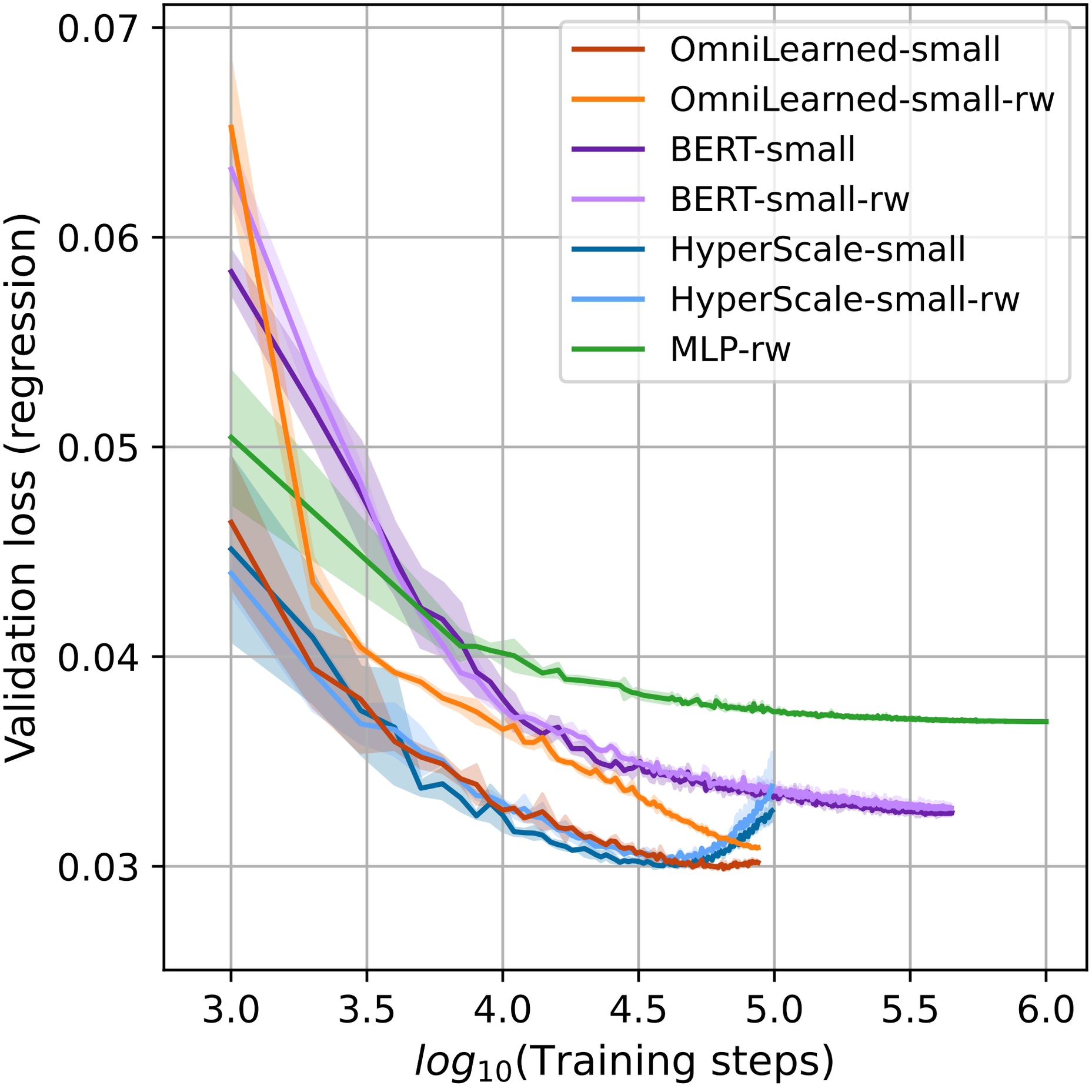

0.07
0.06
0.05
0.04
0.03
Validation loss (regression)
3.0
3.5
4.0
4.5
5.0
5.5
6.0
log_{10}(Training steps)
OmniLearned-small
OmniLearned-small-rw
BERT-small
BERT-small-rw
HyperScale-small
HyperScale-small-rw
MLP-rw