

Transformer test loss vs. learning rate and batch size, at n = 512

