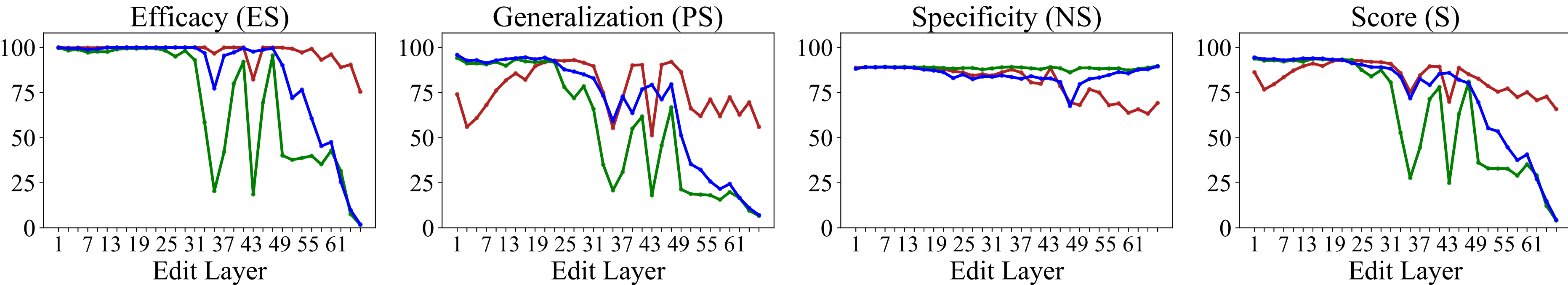


(a) ROME performance on Mamba-2.8b

$W_a^{(l)}$   $W_g^{(l)}$   $W_o^{(l)}$



(b) ROME performance on Pythia-2.8b  $W_{down}^{(l)}$

