

Full Test Set: Masked Mean loss of each category

