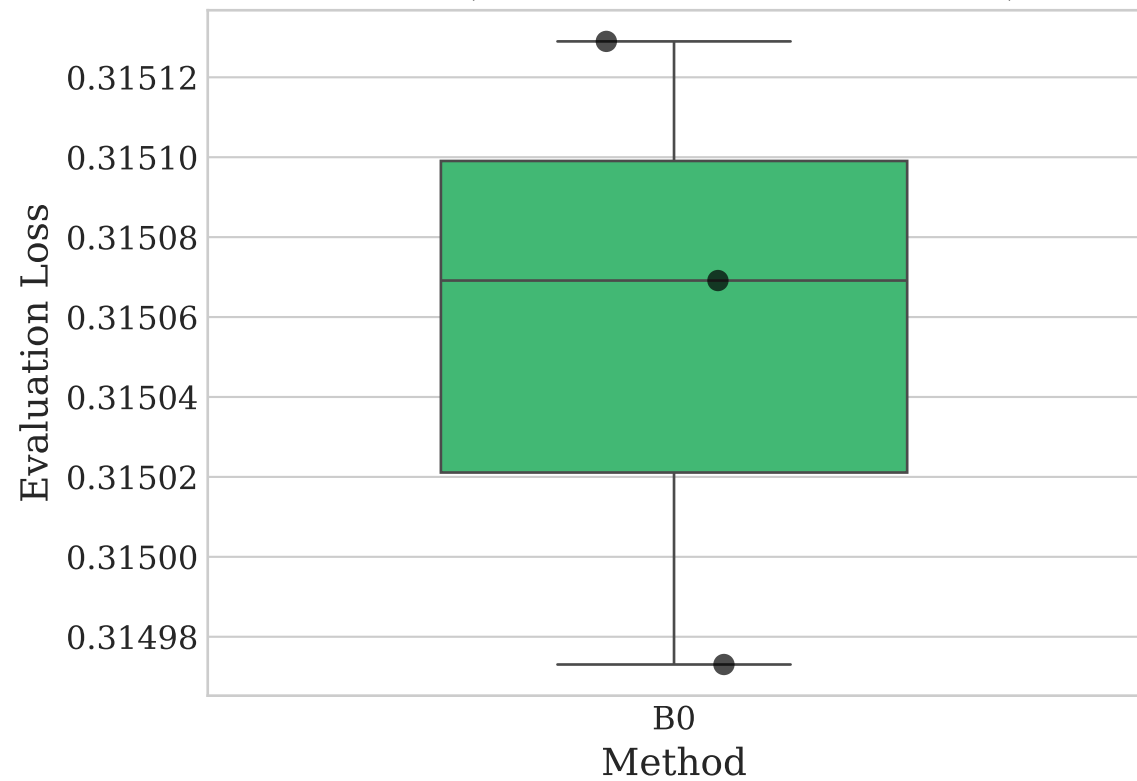Seed Variability Across Training Runs (n=3 per condition)

SST-2 (Sentiment Classification)

SQuAD (Question Answering)