Error Comparison Across Different Heads and Pretraining Sizes