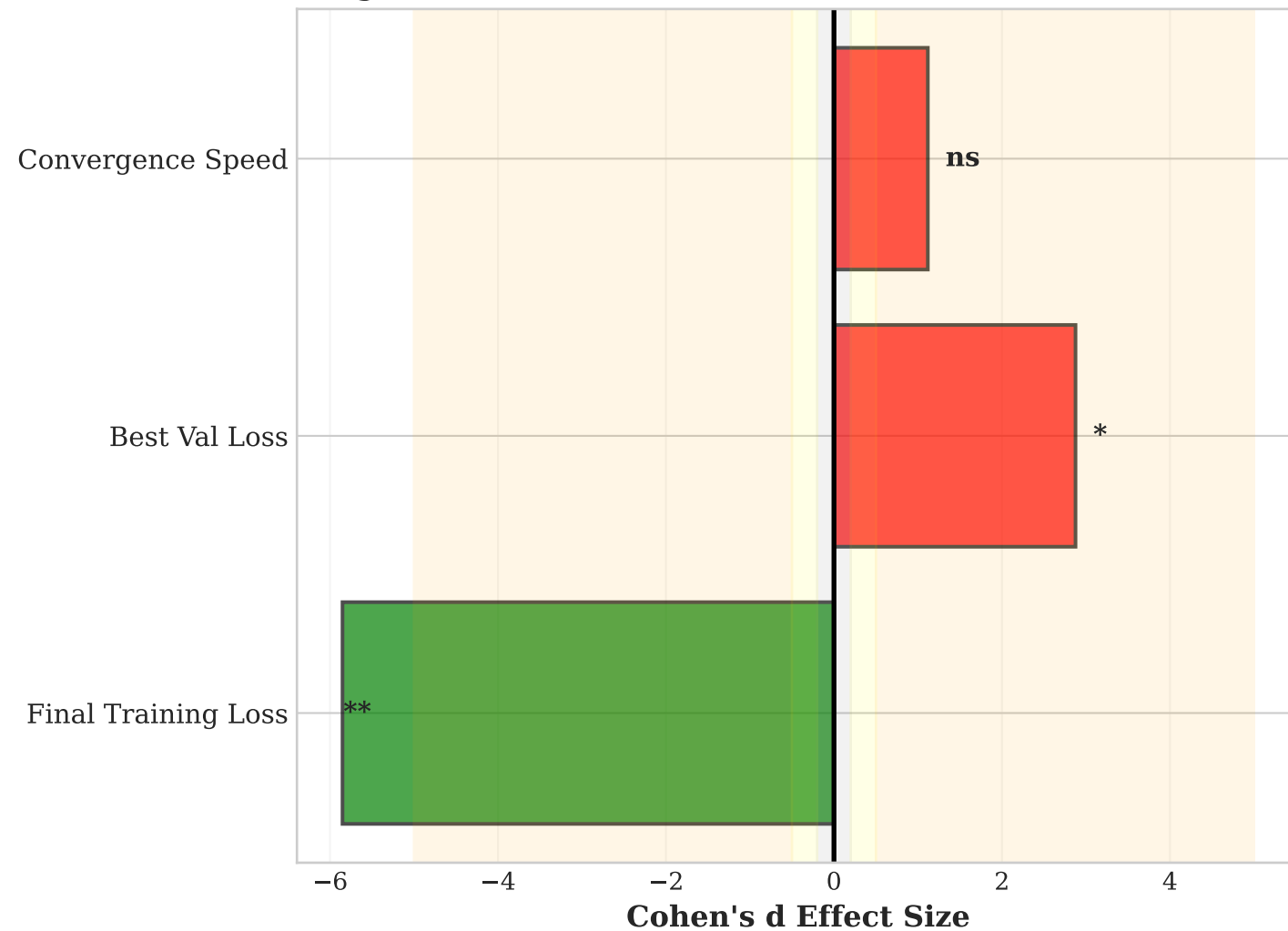


**(a) Training Metrics Effect Sizes**  
(Negative favors RoPE)



**(b) Length Generalization Improvements**

