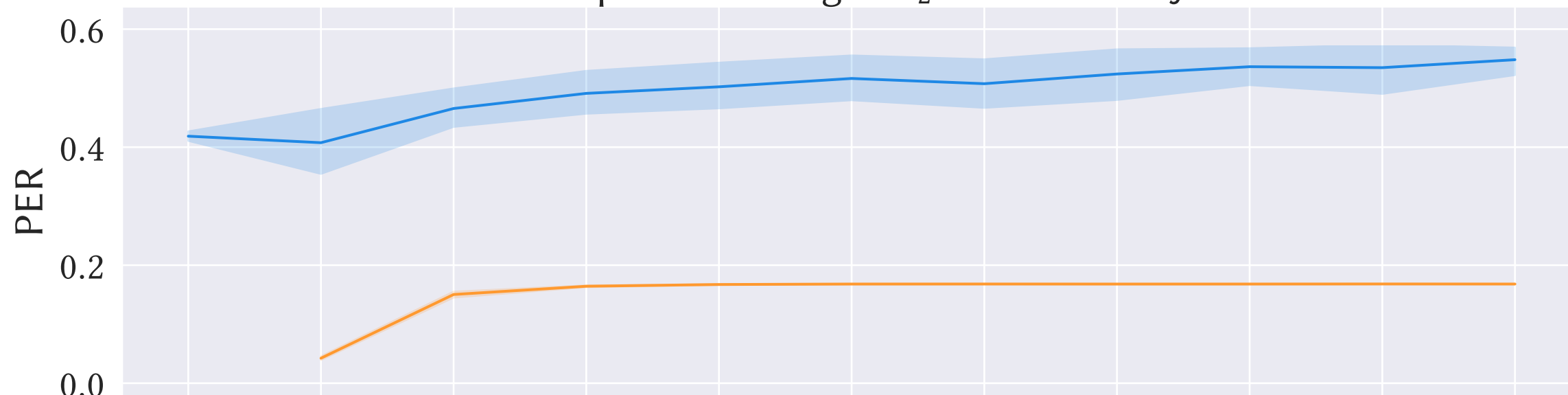
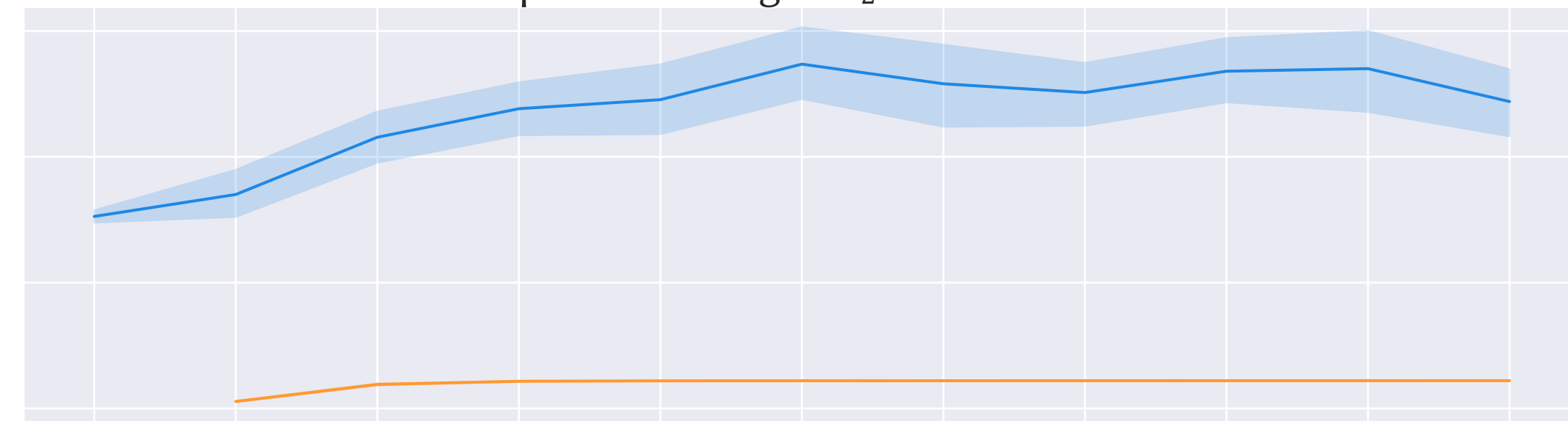
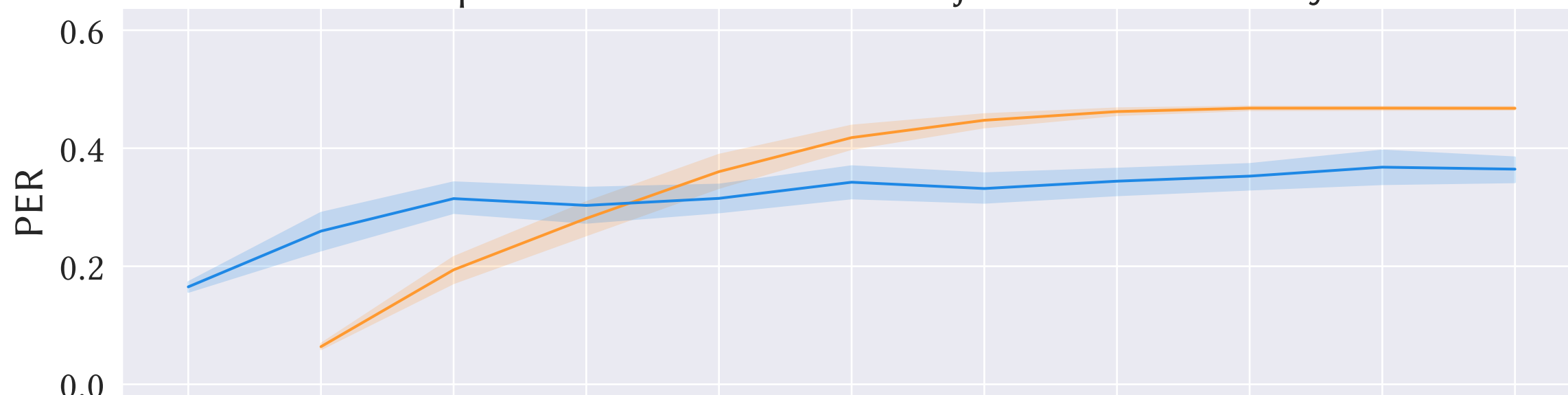
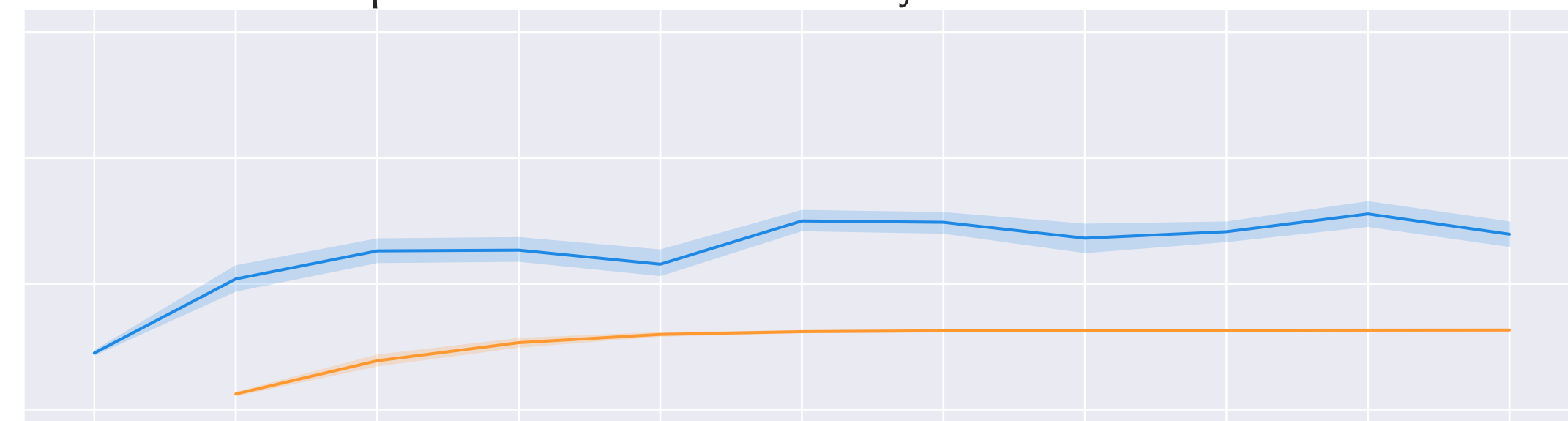


Component = Swiglu W_2 – Scale = tinyComponent = Swiglu W_2 – Scale = small

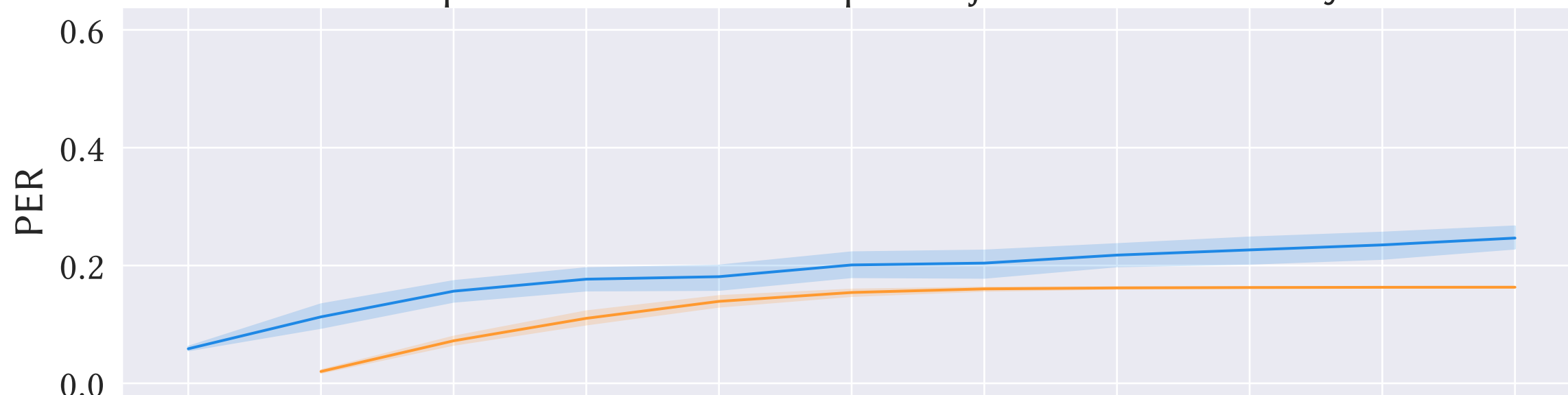
Component = Attention Value Projection – Scale = tiny



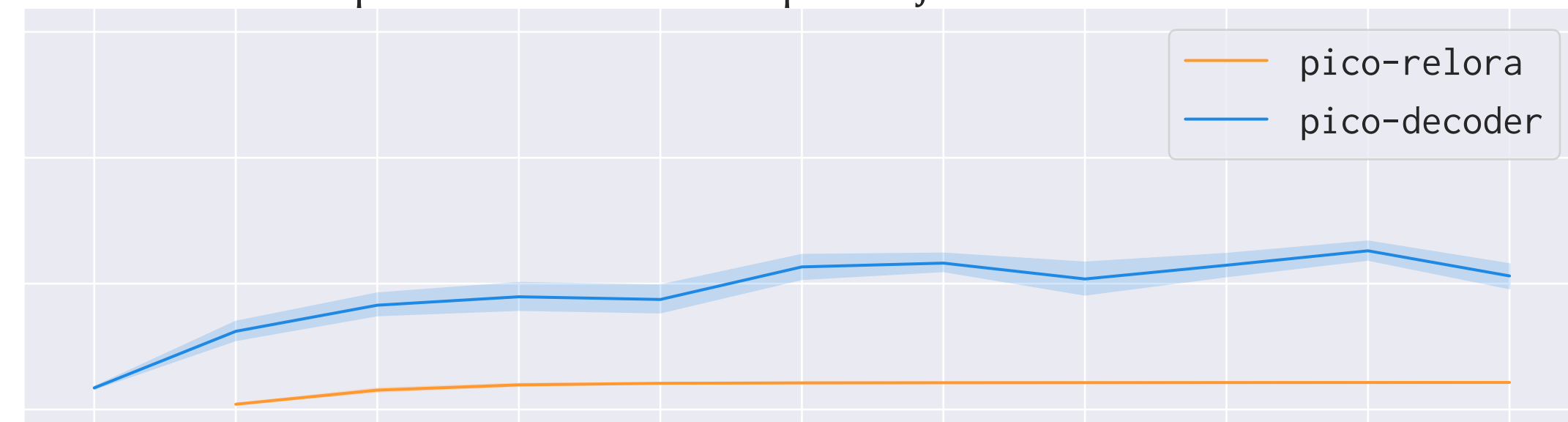
Component = Attention Value Projection – Scale = small



Component = Attention Output Projection – Scale = tiny



Component = Attention Output Projection – Scale = small



pico-relora
pico-decoder

Checkpoint step / 1000

Checkpoint step / 1000