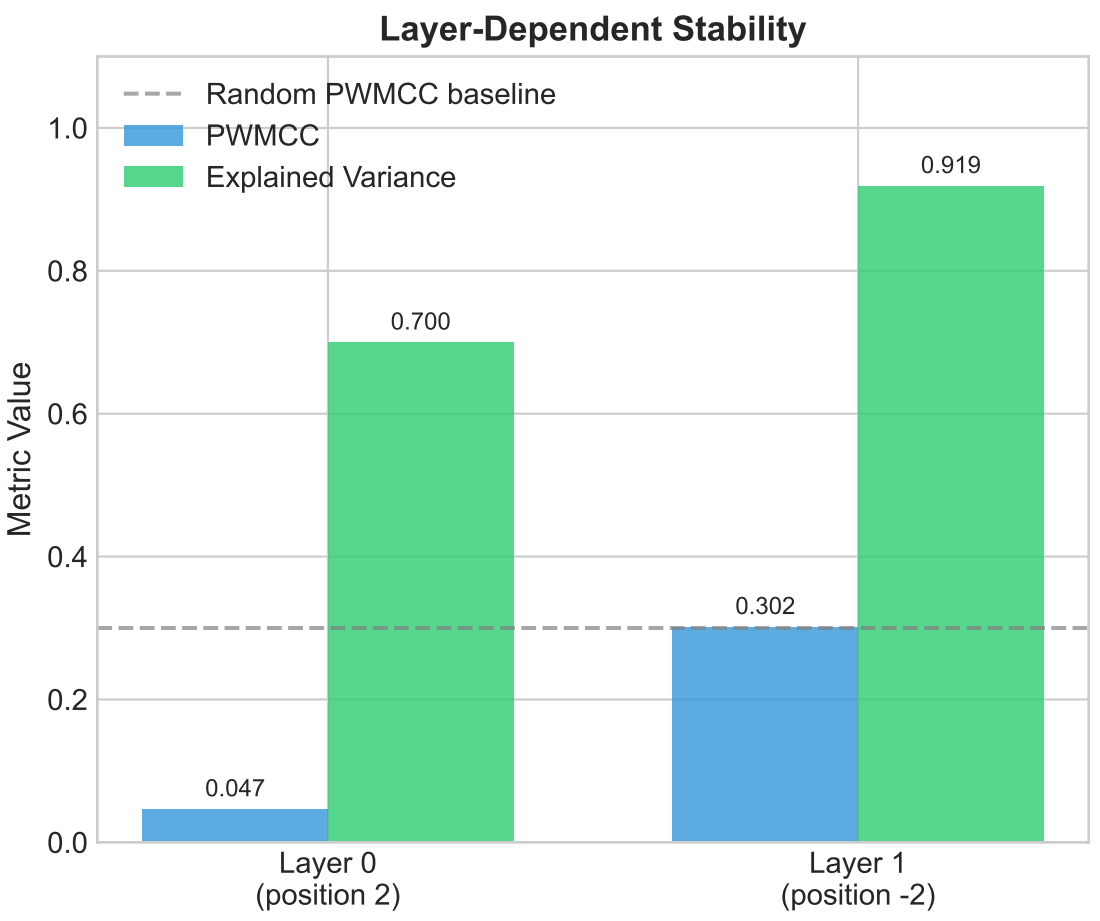

Interpretation

```
INTERPRETATION:

Layer 0 (PWMCC = 0.047, EV = 0.70):
• Features are nearly ORTHOGONAL across seeds
• Multiple equally-valid decompositions exist
• Good reconstruction, but different features

Layer 1 (PWMCC = 0.302, EV = 0.92):
• Features match random baseline
• No consistent feature learning
• Excellent reconstruction

IMPLICATION:
SAEs learn to RECONSTRUCT well, but don't
learn CONSISTENT features across seeds.
```