Performance on Multimodal Sequence Benchmarks