Test loss on two-hop CoT answers