

Score (harsh) vs Confidence Threshold: MSP, winogrande

