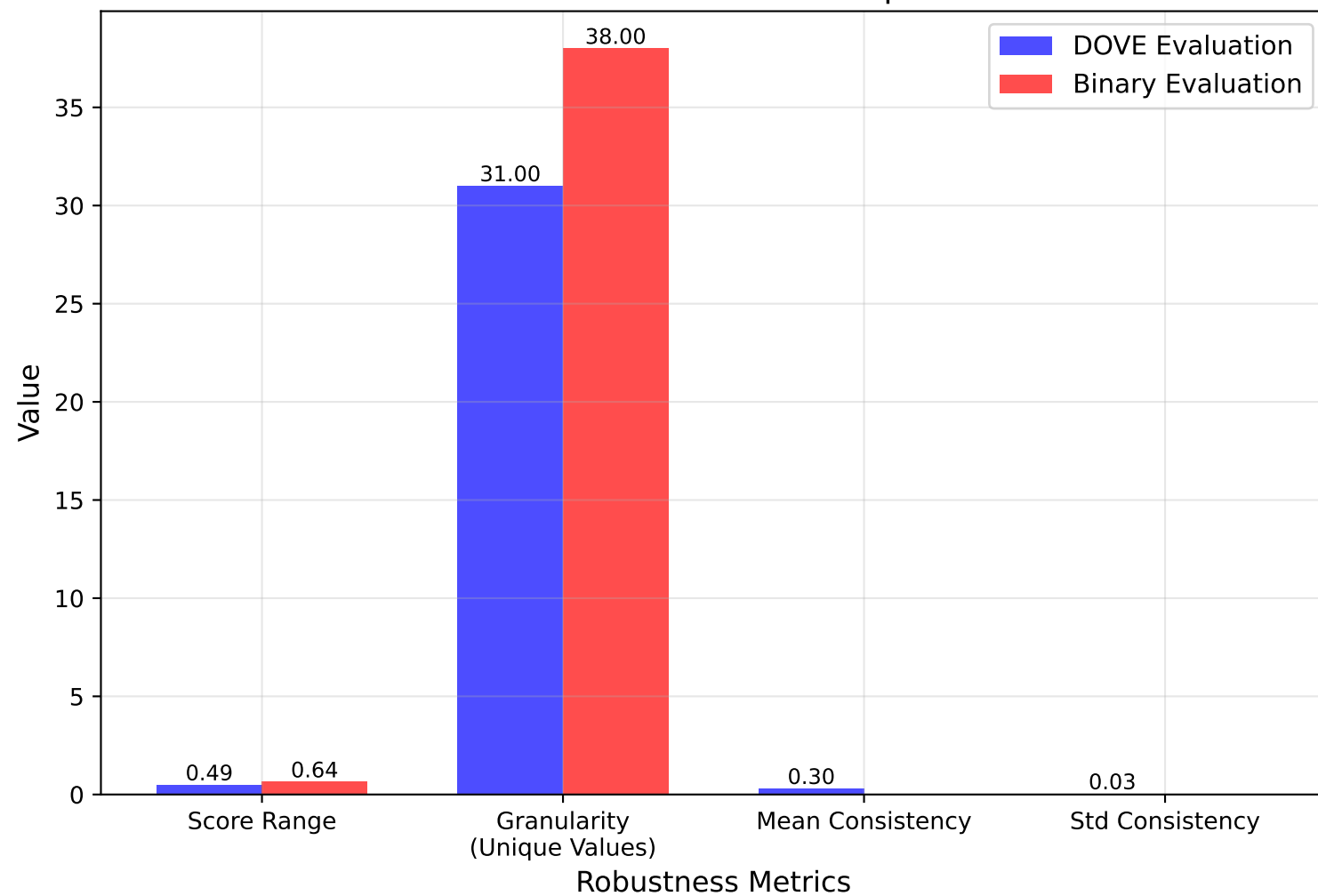
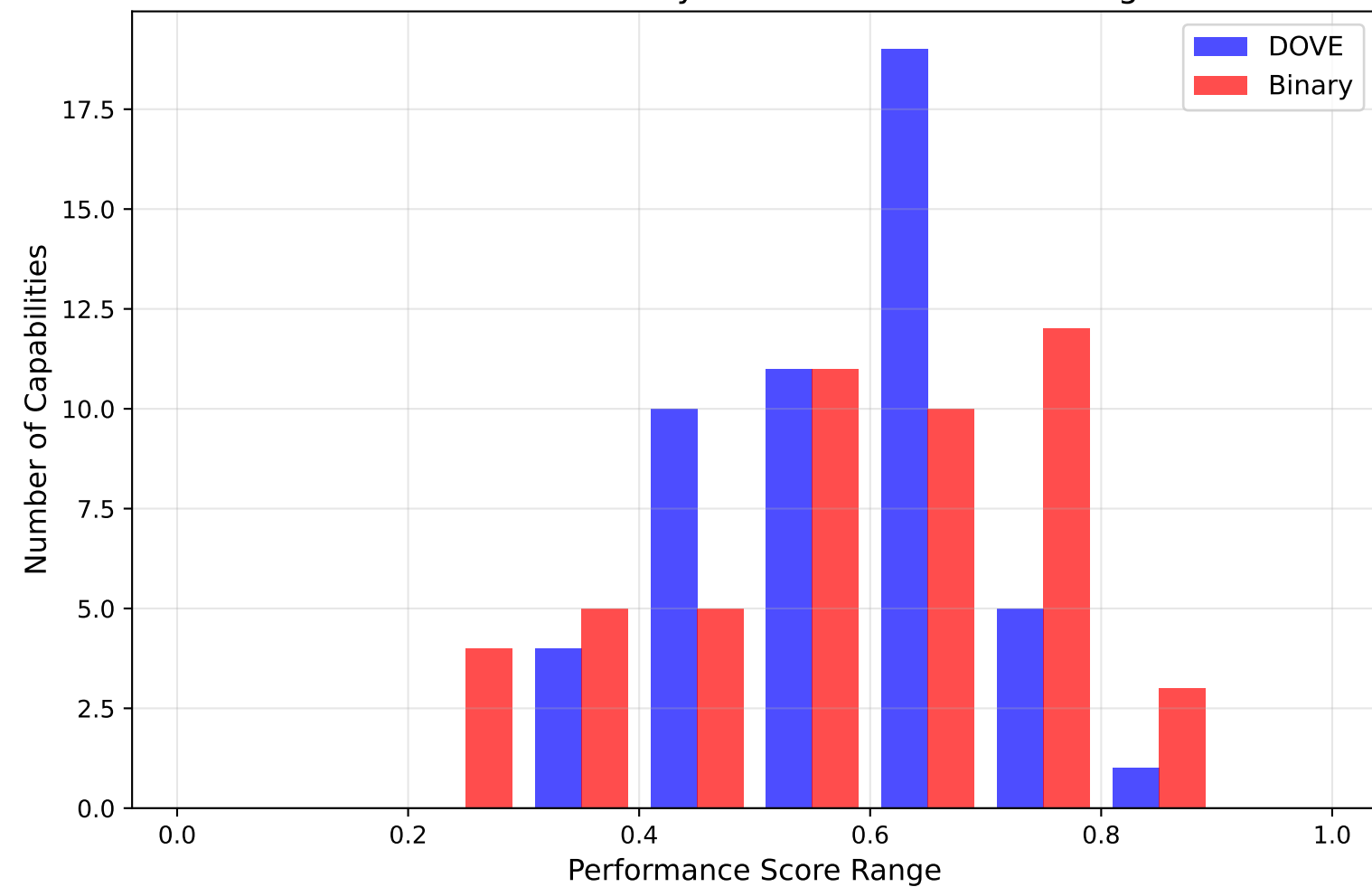


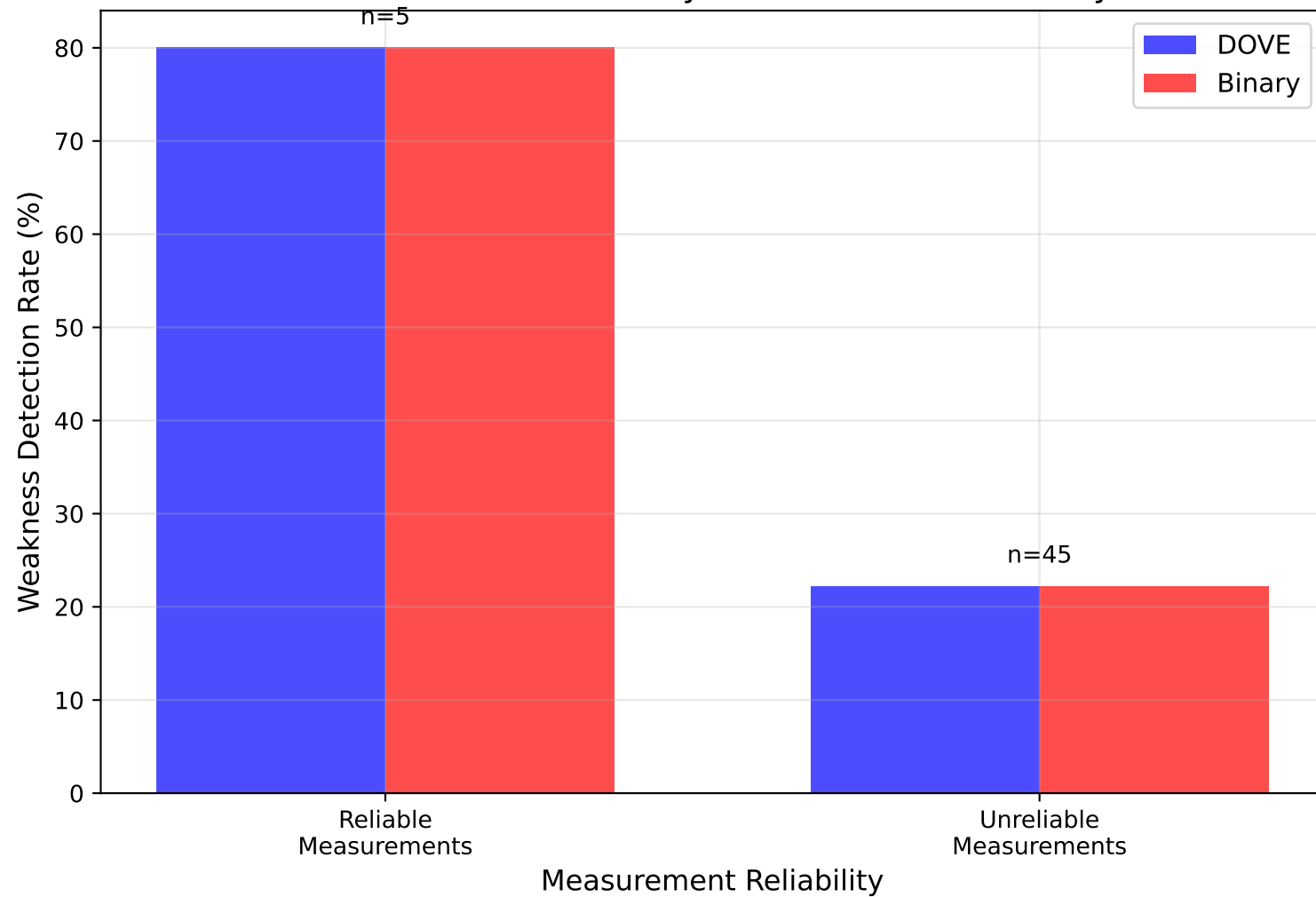
Evaluation Robustness Comparison



Evaluation Sensitivity Across Performance Ranges



Weakness Detection by Measurement Reliability



Overall Robustness Comparison

