Score vs Confidence Threshold: train, truthfulqa