# Peer Review 2:

CheckListing for Basic NLI Understanding

## Scope:

- **Rating:** Excellent
- **Justification:** The paper extends standard NLI evaluation by identifying multiple linguistic capabilities (NER, taxonomy, coreference, temporal, negation, SRL), systematically testing the model's weaknesses, and then improving its performance through targeted training. This multifaceted, capability-focused effort demonstrates a substantive scope.

## Implementation:

- **Rating:** Excellent
- **Justification:** The author systematically generated targeted test sets using CheckList templates, and then retrained or fine-tuned the ELECTRA model to improve its handling of identified weaknesses. Their approach appears technically sound, methodical, and well-documented, indicating a strong implementation.

## Results/Analysis:

- **Rating:** Excellent
- **Justification:** The author not only reports improved model performance in evaluation, but also identifies and addresses underlying data artifacts, revising templates for test cases to ensure more meaningful evaluation. This iterative, example-driven analysis and the subsequent improvement in results shows good insight.

## Clarity/Writing:

- **Rating:** Excellent
- **Justification:** The paper is well-structured, providing a logical progression. The writing is concise and coherent.