# VQA-E Dataset User Study Manual

#### Web Tool

- URL: <a href="http://172.21.150.216:5000">http://172.21.150.216:5000</a>
- Read the instructions and click 'Start Evaluation' button
- You will be assigned with a personal url:
  - o i.e <a href="http://172.21.150.216:5000/">http://172.21.150.216:5000/</a>
  - Please don't go to other user's url.
  - If you accidentally close the window, you can resume the assessment by input your personal url: <a href="http://172.21.150.216:5000/">http://172.21.150.216:5000/</a>
  - You will assess about 100 questions/answer/explanation sets.
  - You will need around 15-20 min to complete the assessment.
- Click 'Next' button to go to next one
- Click ← in your browser to go to previous one

# **Explanation Quality Assessment**

- Read image, Q(question), A(answer)
- Give scores to E(explanation) in four quality metrics
  - o Fluent:
    - Is this E a fluent English sentence/phrase? A fluent E should be correct in grammar and idiomatic in wording.
  - Correct:
    - Is this E is correct according to the image content?
  - Relevant:
    - Is this E relevant to the question? If an E is relevant, you should be able to infer the answer from E.
  - Complementary:
    - Is this E complementary to the Q&A? If an E is complementary, it should provide complementary information that describes the details related to the question and answer.

# **Explanation Quality Assessment**

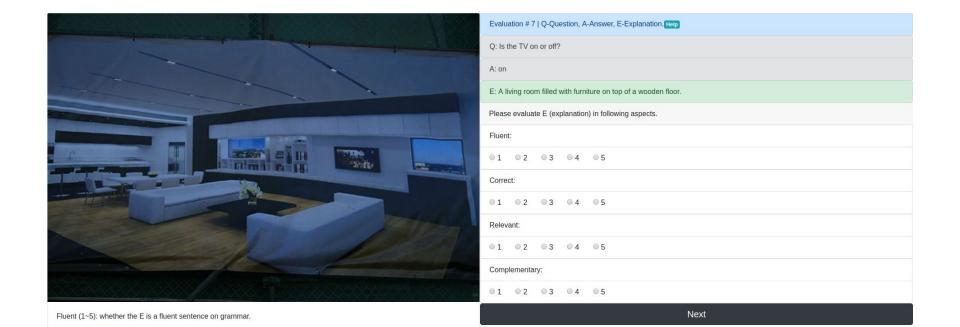
- Scores (1~5)
  - 1-Very Poor
  - o 2-Poor
  - 3-Barely Acceptable
  - o 4-Good
  - 5-Very Good

Not correct. The man is not playing tennis.



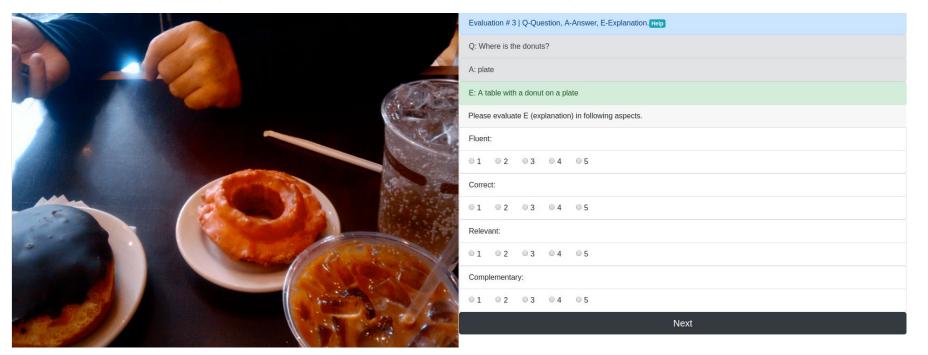
Not relevant. E is not relevant to question. We cannot infer the answer from E.

Not complementary. E does not provide related details to the Q&A.



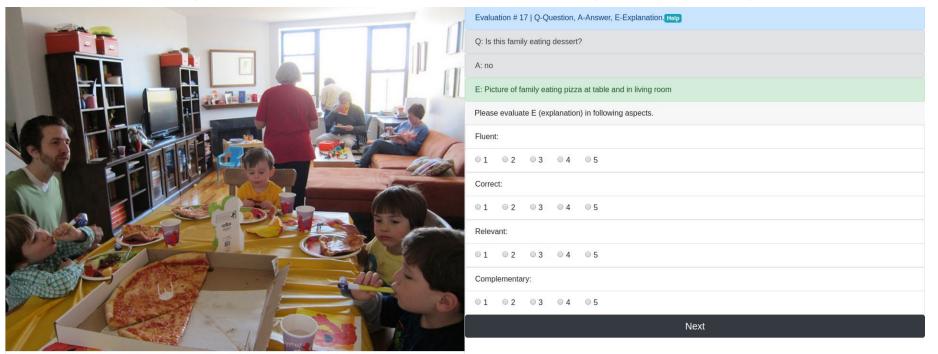
Relevant. E is relevant to question. E tells that the donut is on a plate.

Neutrally complementary. E does not provide many complementary details about answer.



Relevant. E is relevant to question. We can infer the answer from E.

Complementary. E gives related details about the answer (Not eating dessert but eating pizza).



Relevant. E is relevant to question. We can infer the answer from E.

Complementary. E gives details of the answer (because the man wears a boston jersey).

