## Yu et al. (2020) [yu-et-al-annotated.pdf](https://drive.google.com/file/d/1lFZN5h-_XEYdlbme5Sgc-HNa3_1lpaYl/view?usp=drive_link)

On the eighth page, we find the text:

*“Fluency with the value set {1, 2, 3}, is adopted for judging the question fluency. 1 means not fluent and 3 means very fluent.”.*

See also, Table 4.

There is no quality criterion definition given, and the labels do not provide any information on what the authors consider to be fluent (or not). Therefore, the only information we have is the quality criterion name. We are therefore unsure what has been evaluated.

For reference, information on Fluency can be found below.

**Fluency can be found in QCET at “QGO-b-2”**, found under the branch:

1. Q: Quality of outputs
2. QG : Goodness of outputs (excluding correctness)
3. QGO : Goodness of outputs in their own right
4. QGO-b: Goodness of outputs in their own right (both form and content)

Modes:

* **Subjective**: The opinion of evaluators is obtained.
* **Absolute**: Evaluators are shown outputs from a single system.
* **Intrinsic**: Since the system outputs are assessed only within the system context.

## 

## Qu et al. (2002) [qu-et-al-annoated.pdf](https://drive.google.com/file/d/1LMRfS85z40U5SoB-xnRUTKAKioQMokKF/view?usp=drive_link)

On the fourth page, we find the text:

*“eight users were given flight reservation tasks that required them to access the airline schedule database described earlier … Data were collected in two ways: dialogue logs and user survey data. The system logged the total completion time of a dialogue (Total Completion Time), the user actions, the system actions, and the effect of these actions.”*

On the fifth page, we find the text:

*“COMIX-MI improves the dialogue efficiency measures by reducing the average total*

*completion time by about 14%, the number of system turns by 35.2%, and the number of user turns by 49.6%.”*

This indicates that there are three ways in which dialogue efficiency is measured. See also, Table 3.

The quality criterion is **QGE-3-2: “Usability”**, because “system use” is the (external) frame of reference, and usability can be defined as:

“Usability can be described as the capacity of a system to provide a condition for its users to perform the tasks safely, effectively, and efficiently while enjoying the experience.” ([Wikipedia](https://en.wikipedia.org/wiki/Usability))

This is found under the branch:

1. Q: Quality of outputs
2. QG: Goodness of outputs (excluding correctness)
3. QGE: Goodness of outputs relative to external frame of reference
4. QGE-3: Goodness of outputs relative to system use

Modes:

* **Objective**: Measurements are automatically counted.
* **Extrinsic**: System performance is assessed in terms of the system’s effect on something external to the system.
* **Absolute**: From additional text on the fourth page:
  + *“Each user participated in three sessions: a training session and two evaluation sessions, one with COMIX-MI and another with COMIX-UI.”*.
  + Participants are shown one system at a time.