

# Ontology Development Tracker (OnDeT): Capture and Process Differences in Ontologies

No Author Given

No Institute Given

## 1 OnDeT Pre-Evaluation

For evaluating our tool, the so-called System Usability Scale (SUS)[2] is used. The SUS is a reliable and efficient method for gathering user feedback on the usability of the tool. The survey is composed of ten standardized questions that cover different aspects of the software application implementation and user-friendliness. The results provide valuable insights into the user's experience and can be obtained with relatively little effort or time. The standardized questions always have a certain basic tenor. This basic tenor is formulated either positively or negatively for the evaluation of the software application. For example,

- *Positive formulation:* I think that I would like to use Sandbox Semantic Difference frequently.
- *Negative formulation:* I found the Sandbox Semantic Difference Tool unnecessarily complex.

The SUS questionnaire provides for alternating positive and negative wording of the questions (5x5). A 5-point scale is utilized to rate the responses, with 1 denoting significant disagreement and 5 indicating substantial agreement (scale 1-5). The user's input is evaluated for each individual by tallying all the assigned scale values. Positive statements are standardized through the calculation of (scale value -1), and negative statements are normalized through (5 - scale value). Finally, we calculated and aggregated the System Usability Scale (SUS) value for all participants by normalizing and summing their individual results. To gather feedback, we utilized a Google docs, with both the responses and the SUS calculation documented in Table 1. As indicated in Table 1, participant five omitted the response for question 8. Consequently, the value for the question 6 was excluded from the SUS raw score calculation for that user.

The mean of final SUS score is 69.23, equating to a C grade (65.0 - 71.0) based on the GitLab proprietary SUS scale [1]. This indicates that users' experience with the tool ranges between Ok (51.7 - 62.6) and Good (71.1 - 72.5). Participants one, two, three and eight had experience with the tool marked as *best imaginable* [1]. However, participant eleven had a *poor* experience with the tool, scoring 30, which falls within F (25.1 - 51.6) range. Some of the reasons why participant eleven had poor experience with the tool functionalities are given below (see scale values of questions in Table 1). Letter N in the front of a question denotes negative formulation and P denotes positive formulation for the question.

**Table 1.** Survey responses and the SUS scores

Question / Participant ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS raw	SUS final
<b>1</b>	5	2	5	1	5	1	5	1	4	1	38	95
<b>2</b>	5	1	2	1	5	1	5	1	5	2	36	90
<b>3</b>	5	1	5	1	4	2	5	1	4	1	37	92,5
<b>4</b>	4	3	4	1	2	1	4	2	4	1	30	75
<b>5</b>	5	1	4	1	3	4	5		4	2	28	70
<b>6</b>	5	2	4	1	4	2	3	1	4	3	31	77,5
<b>7</b>	1	1	5	5	5	5	4	3	5	1	25	62,5
<b>8</b>	5	1	4	1	5	1	5	2	5	1	38	95
<b>9</b>	4	2	4	2	4	2	3	3	3	1	28	70
<b>10</b>	3	3	3	4	3	2	3	3	2	4	18	45
<b>11</b>	2	4	2	4	3	4	2	3	2	4	12	30
<b>12</b>	3	3	3	5	5	2	4	3	2	5	19	47,5
<b>13</b>	3	4	2	2	4	2	2	4	3	2	20	50
Average											27,69	69,23

N: I found the Semantic Difference Tool unnecessarily complex. (Q2 scale value is equal to 4)

P: I thought the Semantic Difference Tool was easy to use. (Q3 scale value is equal to 2)

N: I think that I would need the support of a technical person to be able to use the Sandbox Semantic Difference Tool. (Q4 scale value is equal to 4).

Observations in the form of free text given by participants about their experience in using our tool are important for further improvement of the tool. We quote here the most relevant one: "I like the idea and the human readable format of showing diffs. What was unclear for me: if the HTML representation of diffs is showing all triple changes or just based on owl terms. Also please add to the Query specific axiom that only commits could be selected where the in beforehand selected IRIs are part of a potential diff."

## References

1. The gitlab handbook, system usability scale, interpreting sus scores (2023), <https://handbook.gitlab.com/handbook/product/ux/performance-indicators/system-usability-scale/>, last accessed 2023-12-06
2. Lewis, J.R.: The system usability scale: past, present, and future. International Journal of Human-Computer Interaction **34**(7), 577–590 (2018), <https://www.tandfonline.com/doi/pdf/10.1080/10447318.2018.1455307>