



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

The evaluation Testing The SUS questionnaire

Fabio Vitali

The evaluation

The evaluation can take place:

- ◆ Internally to the development team, while the product is being developed. This is called inspection of the design
- ◆ Externally to the development team, with the participation of potential external users. This is called testing.



The Likert Scales

Likert scales are psychometric scale for questionnaires.

Named after its inventor, psychologist Rensis Likert (USA, 1903-1981), it is the most widely used in survey research.

Respondents specify their level of agreement on a symmetric scale for a series of statements. The range captures the intensity of their feelings.

Likert scales have either 5 or 6 (no neutral answers) or 7 levels

| Satiety | | | | |
|---|--------------------|---------|-----------------|-------------------|
| X = at meal initiation | | | | |
| Mobile phone use in on-campus computer labs should be prohibited. | | | | |
| 1 | 2 | 3 | 4 | 5 |
| strongly disagree | mildly disagree | neutral | mildly agree | strongly agree |
| 50% Negative | | | 50% Positive | |
| feeling | | | | |



System Usability Scale (SUS)

A very generic, super quick and super dirty testing method. Invented in 1986 by DEC, then adopted by IBM in 2006 because of its simplicity and reasonableness.

It is a test with a fixed protocol and a standard evaluation criteria, which can be algorithmised if needed.

Ten questions, alternatively proposed with positive and negative enunciation, which are asked to respond according to a 5 levels Likert scale.

Missing ratings are worth 3. No variation is allowed in the phrasing of the questions.

It generates a score from 0 to 100, which provides the total value of a usability of the system. Scores higher than 68 are said to indicate a good usability.

Details on the protocol <http://www.measuringusability.com/sus.php>



SUS - The questionnaire (2)

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system



SUS - The algorithm

For each positive question, we assign a score of (selected choice - 1) (going from 0 to 4)

For each negative question, we assign a score of (5 - selected choice) (again, from 0 to 4)

We sum scores and get a value going from 0 to 40. We multiply it by 2.5 and get a value between 0 and 100, with increments of 2.5.

There have been proposals in 2006 two subscales, Learnability and Usability, Learnability corresponds to questions 4, 8 and 10, and Usability at 1, 2, 3, 5, 6, 7, 8, 9 (Question 8 is in both scales) .

An analysis of thousands of test results indicate a good predictive ability of the two factors separately.

