Short Scales of Satisfaction Assessment:
A Proxy to Involve Disabled Users
in the Usability Testing of Websites

Simone Borsci[1], Stefano Federici[2,3(✉)], Maria Laura Mele[2,3],
and Matilde Conti[2]

[1] Human Factors Research Group, School of Mechanical, Materials
and Manufacturing Engineering, The University of Nottingham, Nottingham, UK
simone.borsci@gmail.com
[2] Department of Philosophy, Social & Human Sciences and Education,
University of Perugia, Perugia, Italy
stefano.federici@unipg.it, matildeconti92@hotmail.it,
marialaura.mele@gmail.com
[3] ECONA, Interuniversity Centre for Research on Cognitive Processing in
Natural and Artificial Systems, Sapienza University of Rome, Rome, Italy

Abstract. Short scales of user satisfaction analysis are largely applied in usability studies as part of the measures to assess the interaction experience of users. Among the traditional tools, System Usability Scale (SUS), composed of 10 items, is the most applied quick evaluation scale. Recently, researchers have proposed two new and shorter scales: the Usability Metric for User Experience (UMUX), composed of four items, and the UMUX-LITE, which consists of only the two positive items of UMUX. Despite their recent creation, researchers in human-computer interaction (HCI) have already showed that these two tools are reliable and strongly correlated to each other [1–3]. Nevertheless, there are still no studies about the use of these questionnaires with disabled users. As HCI experts claim [4–7], when disabled and elderly users are included in the assessment cohorts, they add to the overall analysis alternative and extended perspectives about the usability of a system. This is particularly relevant to those interfaces that are designed to serve a large population of end-users, such as websites of public administration or public services. Hence, for a practitioner adding to the evaluation cohorts a group of disabled people may sensibly extend number and types of errors identified during the assessment. One of the major obstacles in creating mixed cohorts is due to the increase in time and costs of the evaluation. Often, the budget does not support the inclusion of disabled users in the test. In order to overcome these hindrances, the administering to disabled users of a short questionnaire—after a period of use (expert disabled costumers) or after an interaction test performed through a set of scenario-driven tasks (novice disabled users)—permits to achieve a good trade-off between a limited effort in terms of time and costs and the advantage of evaluating the user satisfaction of disabled people in the use of websites. To date, researchers have neither analyzed the use of SUS, UMUX, and UMUX-LITE by disabled users, nor the reliability of these tools, or the relationship among those scales when administrated to disabled people.

In this paper, we performed a usability test with 10 blind and 10 sighted users on the Italian website of public train transportation to observe the differences between the two evaluation cohorts in terms of: (i) number of identified errors, (ii) average score of the three questionnaires, and (iii) reliability and correlation of the three scales.

The outcomes confirmed that the three scales, when administered to blind or sighted users, are reliable (Cronbach's $\alpha > 0.8$), though UMUX reliability with disabled users is lower than expected (Cronbach's $\alpha < 0.5$). Moreover, all the scales are strongly correlated ($p < .001$) in line with previous studies. Nevertheless, significant differences were identified between sighed and blind participants in terms of (i) number of errors experienced during the interaction and (ii) average satisfaction rated through the three questionnaires. Our data show, in agreement with previous studies, that disabled users have divergent perspectives on satisfaction in the use of a website. The insight of disabled users could be a key factor to improve the usability of those interfaces which aim to serve a large population, such as websites of public administration and services. In sum, we argue that to preserve the budget and even incorporate disabled users' perspectives in the evaluation reports with minimal costs, practitioners may reliably test the satisfaction by administrating SUS and UMUX or UMUX-LITE to a mixed sample of users with and without disability.

Keywords: Disabled user interaction · Usability evaluation · Usability Metric for User Experience · System Usability Scale

1   Introduction

Satisfaction is one of the three main components of usability [8], along with effectiveness and efficiency. Practitioners are used to testing this component through standardized questionnaires after that people have gain some experience in the use of a website. In particular, experts are used to applying short scales of satisfaction analysis to reduce the time and the costs of the assessment of a website. Among the quick satisfaction scales, the most popular tool of assessment is SUS [9]. SUS is a free and highly reliable instrument [10–14], composed of only 10 items on a five-point scale (1: Strongly disagree; 5: Strongly agree). To compute the overall SUS score, (1) each item is converted to a 0-4 scale for which higher numbers indicate a greater amount of perceived usability, (2) the converted scores are summed, and (3) the sum is multiplied by 2.5. This process produces scores that can range from 0 to 100. Despite the fact SUS was designed to be unidimensional, since 2009, several researchers have showed that this tool has two-factor structures: Learnability (scores of items 4 and 10) and Usability (scores of items 1-3 and 5-9) [2, 3, 13, 15–17]. Moreover, the growing availability of SUS data from a large number of studies [13, 18] has led to the production of norms for the interpretation of mean SUS scores, e.g., the Curved Grading Scale (CGS) [16]. Using data from 446 studies and over 5,000 individual SUS responses, Sauro and Lewis [16] found the overall mean score of the SUS to be 68 with a standard deviation of 12.5.

The Sauro and Lewis CGS assigned grades as a function of SUS scores ranging from 'F' (absolutely unsatisfactory) to 'A+' (absolutely satisfactory), as follows:

Grade F (0–51.7); Grade D (51.8–62.6); Grade C- (62.7–64.9); Grade C (65.0–71.0); Grade C+ (71.1–72.5); Grade B- (72.6–74.0); Grade B (74.1–77.1); Grade B+ (77.2–78.8); Grade A- (78.9–80.7); Grade A (80.8-84.0); Grade A+ (84.1–100).

Recently, two new scales were proposed as shorter proxies of SUS [17]: the UMUX, a four-item tool [1, 19], and the UMUX-LITE composed of only the two positive-tone questions from the UMUX [3]. The UMUX items have seven points (1: Strongly disagree; 7: Strongly agree) and both the UMUX and its reduced version, the UMUX-LITE, are usually interpreted as unidimensional measures. The overall scales of the UMUX and UMUX-LITE range from 0 to 100. Their scoring procedures are:

UMUX: The odd items are scored as [score − 1] and even items as [7 − score]. The sum of the item scores is then divided by 24 and multiplied by 100 [1].

UMUX-LITE: The two items are scored as [score − 1], and the sum of these is divided by 12 and multiplied by 100 [3]. As researchers showed [1, 3, 19], SUS, UMUX, and UMUX-LITE are reliable (Cronbach's α between .80 and .95) and correlate significantly ($p < .001$). However, for the UMUX-LITE, it is necessary to use the following formula (1) to adjust its scores to achieve correspondence with the SUS [3].

$$UMUX - LITE \; = \; .65([\text{Item 1 score}] + [\text{Item 2 score}]) + 22.9. \qquad (1)$$

Despite the fact short scale of satisfaction analysis is quite well known and used in HCI studies, rarely have the psychometric properties of these scales been analyzed by researchers when applied to test the usability of an interface with disabled users. This is because elderly and disabled people are often excluded from the usability evaluation cohorts because they are considered "people with special needs" [20], instead of possible end-users of a product with divergent and alternative modalities of interaction with websites. Nevertheless, as suggested by Borsci and colleagues [21], the experience of disabled users has a great value for HCI evaluators and for their clients. Indeed, to enrich an evaluation cohort with sub-samples of disabled users could help evaluators to run a sort of stress test of an interface [21].

The main complaint of designers, as regards the involvement of disabled people in the usability evaluation, is the cost of the test for disabled users. In fact, disabled users testing usually requires more time compared with the assessment performed by people without disability. The extra-time could be due to the following reasons. First, some disabled users need to interact with a website through a set of assistive technologies and this could require conducting the test in the wild instead of a lab. Second, evaluators need to set-up an adapted protocol of assessment for people with cognitive impairment, such as dementia [7]. Nevertheless, these issues could be overcome by adopting specific strategies. For instance, experts could ask for a small sample of disabled users, who are already customers of a website, to perform at their house a set of short interactions with a website driven by scenarios. Another approach could be to ask disabled users who are novices in the use of a website, to perform at home for a week a set of tasks by controlling remotely the interaction of these users [4]. Independently from the strategies, instead of fully monitoring the usability errors performed by disabled users, experts could just request from these end-users to complete a short scale after their experience with a system to gather their overall satisfaction. The satisfaction outcomes of disabled users' cohort could be then aggregated and compared

with the results of the other cohort of people without disability. Therefore, by using short scales of satisfaction evaluation, practitioners could save on costs and, with a minimal effort, report to designers the number of errors identified, the level of satisfaction experienced by users without disability, and a comparative analysis of the satisfaction with a mixed cohort of users. Thus, short scales could be powerful tools to include, at minimal cost, the opinions of disabled users in the usability assessment, in order to enhance the reliability of the assessment report for the designers.

Today, the possibility to include a larger sample of users with different kind of behaviors in the usability testing is particularly relevant to obtain a reliable assessment. In fact, in the context of ubiquitous computing people could access and interact through different mobile devices with websites, and a large set of information on public services (such as taxes, education, transport, etc.) is available online. Therefore, for the success of public services websites it is important to have an interface which is accessible to a wide range of possible users and usable in a satisfactory way.

Despite the growing involvement of disabled users in the usability analysis, there are no studies analyzing the psychometric properties of short scales of satisfaction and the use of these tools to assess the usability of website interfaces perceived by a sample of disabled users.

The aim of this paper is to propose a preliminary analysis of the use of SUS, UMUX, and UMUX-LITE with a small sample of users with and without disability. To reach this aim, we involved in a usability assessment two different cohorts (blind and sighted users), in order to observe the differences between the two samples in terms of number of errors experienced by the end-users during the navigation, and the overall scores of the questionnaires. Moreover, we compared the psychometric properties of SUS, UMUX, and UMUX-LITE when administered to blind and sighted participants in terms of reliability and scales correlation.

2   Methodology

Two evaluation cohorts composed of 10 blind-from-birth users (Age: 23.51; SD: 3.12) and 10 sighted users (Age: 27.88; SD: 5.63) were enrolled through advertisements among associations of disabled users, and among the students of the University of Perugia, in Italy. Each participant was asked to perform on the website of the Italian public train company (http://www.trenitalia.it) the following three tasks, presented as scenarios:

– Find and buy online a train ticket from "Milan – Central station" to "Rome – Termini station."
– Find online and print the location of info-points and ticket offices at the train station of Perugia.
– Use the online claim form to report a problem about a train service.

Participants were asked to verbalize aloud their problems during the navigation. In particular, sighted users were tested through a concurrent thinking aloud protocol, while blind users were tested by a partial concurrent thinking aloud [7].

After the navigation each participant filled the Italian validated version [14] of three scales, presented in a random order.

2.1   Data Analysis

For each group of participants there were descriptive statistics (mean [M], standard deviation [SD]). An independent t-test analysis was performed to test the differences between the two evaluation cohorts in terms of overall scores of the three questionnaires. Moreover, a Cronbach's α and Pearson correlation analyses were performed to analyze the psychometric properties of the scales when administered to different end-users. All analyses were performed using IBM® SPSS 22.

3   Results

3.1   Usability Problems and User Satisfaction

The two evaluation cohorts identified, separately, a total number of 29 problems: Blind users experienced 19 usability issues, while sighted users experienced only 10 issues. Of the 29 issues reported by the two cohorts, eight issues were identified by both blind and sighted users; two problems only by sighted users; and 11 only by blind users. Therefore, a sample of 21 unique usability issues was identified testing 20 end-users. As reported in Table 1, an independent t-test analysis showed that for each of the questionnaires there was a significant difference between the overall satisfaction in use experienced by blind and sighted users.

Table 1. Differences among SUS, UMUX, and UMUX-LITE administered to blind and sighted users.

Blind vs. Sighted users	Degree of Freedom	t	p
SUS	17	6.469	.001
UMUX		4.876	.001
UMUX-LITE		4.319	.001

As can be seen in Table 2, while blind users assessed the website as not usable (Grade F), sighted users judged the interface as having an adequate level of usability (Grades for C- to C). By aggregating the two evolution cohorts, the website could be judged as a product with a low level of usability (Grade F).

3.2   Psychometric Properties of Questionnaires

The Cronbach's α analysis showed that all the questionnaires are reliable when administered to both sighted and blind users (Table 3). Nevertheless, in the specific case of blind users, UMUX reliability is lower than expected (.568).

Table 2. Average score, standard deviation (SD) and average aggregated scores of the SUS, UMUX, and UMUX-LITE of blind and sighted users. For each scale the Curved Grading Scale (CGS), provided by Sauro and Lewis [16], was also used to define the grade of website usability.

	Sighted	Blind	Av. aggregated scores
SUS	67.75 (SD:20.83)	15.25 (SD:11.98)	41.5 (SD:31.6)
	Grade C	Grade F	Grade F
UMUX	62.02 (SD:17.91)	32.10 (SD:11.99)	46.27 (SD:21.21)
	Grade C -	Grade F	Grade F
UMUX-LITE	68.52 (27.48)	17.54 (14.24)	41.66 (34.27)
	Grade C	Grade F	Grade F

Table 3. Reliability of SUS, UMUX, and UMUX-LITE for both blind, and sighted users.

	Blind	Sighted
SUS	.837	.915
UMUX	.568	.898
UMUX-LITE	.907	.938

Table 4. Correlations among SUS, UMUX, and UMUX-LITE for both blind and sighted users.

Types of end-users	Scales	SUS	UMUX
Blind	SUS	1	.948**
	UMUX	.935**	1
	UMUX-LITE	.948**	.928**
Sighted	SUS	1	.890**
	UMUX	.890**	1
	UMUX-LITE	820**	.937**

**. Correlation is significant at the 0.01 level (2-tailed).

As Table 4 shows, all the questionnaires, independently from the evolution cohort, are strongly correlated (p < .001).

4   Discussion

Table 2 clearly shows that while sighted users judged the website as quite a usable interface (Grades from C- to C), disabled users assessed the product as not usable (Grade F). This distance between the two evaluation cohorts is perhaps due to the fact that blind users experienced 11 more problems than the cohort of sighted participants. These results indicate that a practitioner adding to an evaluation cohort a sample of disabled users may drastically change the results of the overall usability assessment, i.e., the average overall score of the scales (Table 1).

The three scales were very reliable for both the cohorts (Cronbach's $\alpha > 0.8$; Table 3), however, the UMUX showed a low reliability when administered to blind users (Cronbach's $\alpha > 0.5$). This low level of reliability of UMUX was unexpected, considering also that UMUX-LITE composed of only the positive items of UMUX – i.e., items 1 and 3 – was very reliable (Table 3). Perhaps the negative items of UMUX – i.e., items 2 and 4 – were perceived by disabled users as complex or unnecessary questions, or this effect is an artifact of the randomized presentation of the questionnaires to the participants. Finally, for both the cohorts, the three scales were strongly correlated – i.e., $p<.001$ (see Table 4).

5   Conclusion

Quick and short questionnaires could be reliably used to assess the usability of a website with blind users. All the three tools reliably capture the experience of participants with and without disability, by offering to practitioners a good set of standardized results about the usability of a website.

Although further studies are needed to clarify the reliability of UMUX when administered to disabled users, our results suggest that UMUX-LITE and SUS might be applied by practitioners as good scales of satisfaction analysis. The use of these short scales may help practitioners to involve blind participants in their evaluation cohorts and to compare the website experience of people with and without disability. In fact, practitioners with a minimal cost may administer SUS and UMUX or UMUX-LITE to a mixed sample of users, thus obtaining an extra value for their report: the divergent perspectives of the disabled users. This extra value is particularly important for websites of public administration and of those services, such as public transport, that have to be accessed by a wide range of people with different levels of functioning.

References

1. Finstad, K.: The Usability Metric for User Experience. Interacting with Computers **22**, 323–327 (2010)
2. Lewis, J.R., Sauro, J.: The Factor Structure of the System Usability Scale. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 94–103. Springer, Heidelberg (2009)
3. Lewis, J.R., Utesch, B.S., Maher, D.E.: Umux-Lite: When There's No Time for the Sus. In: Conference on Human Factors in Computing Systems: CHI '13, pp. 2099–2102 (2013)
4. Petrie, H., Hamilton, F., King, N., Pavan, P.: Remote Usability Evaluations with Disabled People. In: SIGCHI Conference on Human Factors in Computing Systems: CHI '06, pp. 1133–1141 (2006)
5. Power, C., Freire, A., Petrie, H., Swallow, D.: Guidelines Are Only Half of the Story: Accessibility Problems Encountered by Blind Users on the Web. In: Conference on Human Factors in Computing Systems: CHI '12, pp. 433 (2012)
6. Rømen, D., Svanæs, D.: Evaluating Web Site Accessibility: Validating the Wai Guidelines through Usability Testing with Disabled Users. In: 5th Nordic Conference on Human-Computer Interaction—Building Bridges: NordiCHI08, pp. 535–538 (2008)

7. Federici, S., Borsci, S., Stamerra, G.: Web Usability Evaluation with Screen Reader Users: Implementation of the Partial Concurrent Thinking Aloud Technique. Cogn. Process. **11**, 263–272 (2010)
8. ISO: Iso 9241-11:1998 Ergonomic Requirements for Office Work with Visual Display Terminals – Part 11: Guidance on Usability. CEN, Brussels, BE (1998)
9. Brooke, J.: Sus: A "Quick and Dirty" Usability Scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor & Francis, London (1996)
10. Lewis, J.R.: Usability Testing. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, pp. 1275–1316. John Wiley & Sons, New York (2006)
11. Sauro, J., Lewis, J.R.: When Designing Usability Questionnaires, Does It Hurt to Be Positive? In: Conference on Human Factors in Computing Systems: CHI '11, pp. 2215–2224 (2011)
12. Zviran, M., Glezer, C., Avni, I.: User Satisfaction from Commercial Web Sites: The Effect of Design and Use. Information & Management **43**, 157–178 (2006)
13. Bangor, A., Kortum, P.T., Miller, J.T.: An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction **24**, 574–594 (2008)
14. McLellan, S., Muddimer, A., Peres, S.C.: The Effect of Experience on System Usability Scale Ratings. Journal of Usability Studies **7**, 56–67 (2012)
15. Borsci, S., Federici, S., Lauriola, M.: On the Dimensionality of the System Usability Scale (Sus): A Test of Alternative Measurement Models. Cogn. Process. **10**, 193–197 (2009)
16. Sauro, J., Lewis, J.R.: Quantifying the User Experience: Practical Statistics for User Research. Morgan Kaufmann, Burlington (2012)
17. Lewis, J.R.: Usability: Lessons Learned … and yet to Be Learned. International Journal of Human-Computer Interaction **30**, 663–684 (2014)
18. Kortum, P.T., Bangor, A.: Usability Ratings for Everyday Products Measured with the System Usability Scale. International Journal of Human-Computer Interaction **29**, 67–76 (2012)
19. Finstad, K.: Response to Commentaries on 'The Usability Metric for User Experience'. Interacting with Computers **25**, 327–330 (2013)
20. Biswas, P., Langdon, P.: Towards an Inclusive World – a Simulation Tool to Design Interactive Electronic Systems for Elderly and Disabled Users. In: 2011 Annual SRII Global Conference, pp. 73–82 (2011)
21. Borsci, S., Kurosu, M., Federici, S., Mele, M.L.: Computer Systems Experiences of Users with and without Disabilities: An Evaluation Guide for Professionals. CRC Press, Boca Raton, FL (2013)