Do not Shoot the Messenger: Effect of System Critical Feedback on User-Perceived Usability

Georgios Melissourgos[1][0000-0002-9973-0104] and Christos Katsanos[2][0000-0001-9031-3083]

[1] Multitude SE, Helsinki, Finland
georgios.melissourgos@multitude.com
[2] Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
ckatsanos@csd.auth.gr

Abstract. Measuring perceived usability with questionnaires is a common practice for usability researchers and practitioners. This paper investigates whether there is any user bias towards the perceived usability of a system, when this system administers critical feedback to its users. These systems make decisions that substantially affect their users' lives, such as automated medical diagnosis, bank loan approval etc. In our study, we gathered data from three, almost identical, systems used to apply for a consumer loan and communicate the decision to the applicant. Our dataset involves a total of 332 applicants who completed the UMUX-LITE questionnaire after receiving the system decision (approved, rejected) for their loan. Results showed that participants who had their loans approved (positive system critical feedback) provided significantly higher UMUX-LITE scores compared to participants who had their loans rejected (negative system critical feedback). This finding suggests that one should pay attention when measuring perceived usability of critical feedback administering systems as it tends to be biased from the critical feedback that users received.

Keywords: System Critical Feedback, Perceived Usability, Usability Questionnaire, UMUX-LITE.

1 Introduction

1.1 Perceived Usability and Its Measurement

Perceived usability reflects users' subjective assessments of the usability of a system. Hertzum [1] considers perceived usability as one of the six images of usability which "concerns the user's subjective experience of a system based on his or her interaction with it".

Measuring perceived usability with questionnaires is a common practice for usability researchers and practitioners. Examples of such usability questionnaires are the System Usability Scale (SUS) [2], the Post-Study System Usability Questionnaire (PSSUQ) and the Computer System Usability Questionnaire (CSUQ) [3], the Usefulness, Satisfaction, and Ease of use (USE) [4], the Purdue Usability Testing Questionnaire (PUTQ) [5], and the Questionnaire for User Interface Satisfaction (QUIS) [6]. Popular usability

questionnaires have been translated into languages other than English. For instance, SUS is available in Greek [7], Slovenian [8], Polish [9], Persian [10], Chinese [11, 12], Danish [13], Arabic [12], French [12], German [12], Hindi [12] and Spanish [12]. As another example, PSSUQ and CSUQ have been translated into Greek [14] and into Turkish [15] and Arabic [16] respectively.

An important trend to perceived usability measurements are short usability questionnaires, which are supported by a growing body of evidence regarding their reliability and validity. The SUS consists of 10 questions and is one of the most used such questionnaires [17] that remains reliable even for a small sample size [18]. Still, there are cases where an even shorter questionnaire is required, such as when the perceived usability measurement is part of a larger questionnaire, or when multiple systems (e.g., versions of a system, competitors' systems) must be evaluated in the same usability study, or when the users' time is in deficit (e.g., studies in the wild). As a result, questionnaires have been developed to measure perceived usability with even less questions than the 10 of SUS. The Usability Metric for User Experience (UMUX) [19] has 4 questions and the UMUX-LITE [20] has only 2 questions.

All questionnaires mentioned so far are administered post study, that is after the participant has finished all the tasks. However, there are also short post-task questionnaires that are completed immediately after finishing a task, such as the After-Scenario Questionnaire (ASQ) [21] with 3 items and the Single Ease Question (SEQ) [22] with 1 item. Even though the latter two questionnaires are very short, they must be completed multiple times in a usability-study session that typically involves multiple tasks. Thus, any time savings from the smaller number of questions might be offset by the additional time required for multiple completions of the same questionnaire.

1.2 The UMUX-LITE Questionnaire

UMUX-LITE [20] is a two-items standardised questionnaire designed to measure perceived usability of a system. It comprises of the following two questions answered on a scale from 1 (strongly disagree) to 7 (strongly agree): "Q1. This system's capabilities meet my requirements" and "Q2. This system is easy to use". These questions are the two positively worded questions of UMUX. A UMUX-LITE score ranges from 0 to 100 and is calculated using Equation 1, where Q1 and Q2 are the participant's answers in Question 1 and Question 2 respectively.

$$\text{UMUX-LITE} = (Q1 + Q2 - 2) * (100/12) \text{ [20, 23]} \quad (1)$$

In the initial paper that presented UMUX-LITE [20], the scale was found to have high reliability (alpha coefficient from 0.82 to 0.83) and validity (approximately 1% difference with SUS scores). Several other studies have replicated the psychometric properties of UMUX-LITE with alpha coefficients ranging between 0.65 and 0.86 [23–25], which is excellent for a two-items questionnaire, and correlations with the SUS ranging from 0.74 to 0.83 [23, 25]. Given their short form, good psychometric properties, and correspondence to SUS, both UMUX and UMUX-LITE are quickly emerging questionnaires for measuring perceived usability; the paper introducing UMUX-LITE [20] has already 223 citations in Google Scholar as of 07/01/2023, and the initial paper on

UMUX [19] has 597 citations in Google Scholar as of 07/01/2023 compared to 144 on 9/12/2017 (as mentioned in [23]).

1.3 Research Motivation

It has been demonstrated that perceived usability questionnaires are affected by user's characteristics. Studies [26, 27] report a significant negative correlation between SUS score and age. By contrast, a more recent study [28] found that SUS scores and age are not significantly associated, while at the same time mentioning that the age distribution of the participants was not varied enough. Additional recent studies have found that SUS, UMUX and UMUX-LITE scores are not affected by age [29]. In addition, numerous studies [17, 27, 28, 30] have consistently shown that there is no effect of respondent's gender on SUS scores. Kortum and Oswald [31] found that SUS scores are affected by gender; females tended to have higher SUS scores, however this result was better explained by their personality traits. They state that SUS is affected by personality traits. Furthermore, previous experience with the evaluated system has been found to significantly increase both the obtained SUS scores [27, 28, 32, 33] and UMUX-LITE scores [25, 29].

However, there is little research on how the system's nature might affect the perceived usability measured. Bangor and colleagues [27] found that the system type (i.e., cell phones, customer equipment such as modems, GUIs, interactive voice response systems, and web pages/applications) significantly affects the obtained SUS scores. One other case of system type that remains rather unexplored is systems used to decide on issues that have a drastic effect on users' life activities, hereafter systems providing critical feedback. Examples of such systems are websites for loan application, automated diagnostic systems (e-health), and online examination systems.

This paper investigates whether there is any user bias towards the perceived usability of a system, when the system in question administers critical feedback. Critical feedback is commonly studied on systems that administer critical feedback for learning purposes to improve students' performance [34]. Our aim is to study the effect of critical feedback on perceived usability, when that is not administered with the objective of learning, rather as a final response. Our hypothesis is that if a system administers positive critical feedback to the user, then its perceived usability will be higher than the perceived usability for a user that received negative critical feedback.

2 Methodology

2.1 Evaluated System

The evaluated system is a web-based system that allows users to apply for a loan and receive a decision (critical feedback) on their application. This system is offered by Ferratum, a business unit within Multitude which is an international provider of digital financial services. Applying for a loan can be quite a complicated process, thus the company is highly interested to measure and improve its usability. The study reported in this paper is only one of the actions that the company takes in order to achieve this.

To apply for a loan, users go through a process of steps filling relevant information and uploading required documents. At the end of the process, the system approves or rejects the loan request based on a proprietary algorithm that calculates a risk score for each applicant. Each operating country is being served in the most commonly spoken languages, most of the time just the native language. Multitude is operating in multiple countries, but this study concerns the system available in three specific countries in Europe (exp1, exp2, exp3); countries are anonymized for privacy reasons as requested by the company. The systems between the countries are almost identical, except being served in different languages, and the existence or absence of some input fields due to local credit regulation.

2.2 Participants

A total of 332 participants from three countries were involved in the study. The participants were actual users of the web-based system who had just applied for a loan and received the systems' decision on their application. They were recruited by a message that asked them to evaluate the usability of the system.

From the 332 participants, 108 had their loan accepted (positive feedback) and 224 had their loan rejected (negative feedback). The participants in exp1 were 74 (positive feedback: 25 vs. negative feedback: 49), in exp2 they were 155 (43 vs. 112), and in exp3 they were 103 (40 vs. 63).

2.3 Instruments

UMUX-LITE was used to measure the perceived usability of the evaluated web-based system. Given that the questionnaire would be addressed to actual customers and administered through the live website, we chose a short questionnaire because it would have a higher likelihood of being answered voluntarily.

The study took place from August to September of 2022 and till that timepoint there were no validated translations of UMUX or UMUX-LITE in the languages used by each country's system. Thus, the English UMUX-LITE was translated by the native speakers responsible for company communications in each of the countries and was provided in the language used on the system serving each country.

2.4 Procedure

The evaluation study took place in the live website of each system for a period of two months. After each loan applicant finished the loan application procedure and received the critical feedback from the system (approved or rejected), they were shown the UMUX-LITE questionnaire that they could optionally complete in their native language. An introductory text informed participants that they participate on a voluntary basis, that their replies to the questionnaire are anonymous and that they would not affect the decision already communicated to them. For each participant, we collected the replies to the UMUX-LITE questionnaire and the system's critical feedback (positive or negative).

3 Results

We collected three datasets of UMUX-LITE scores, one per system used in each country of our participants.

An initial analysis was conducted to investigate whether a cross-country aggregated dataset of UMUX-LITE scores could be compiled. The assumption of normality was violated for all three countries; exp1: $W(74)=0.892$, $p<0.001$, exp2: $W(155)=0.894$, $p<0.001$, and exp3: $W(103)=0.869$, $p<0.001$ respectively. Thus, Kruskal-Wallis one-way ANOVA, a non-parametric test, was applied. Results found no significant differences among the UMUX-LITE scores for the systems used in the three countries; $H(2)=0.854$, $p=0.652$. Hence, a cross-country dataset can be compiled.

In the following, we report analyses both for the cross-country aggregated dataset and for each system (experiment) separately for completeness purposes. Table 1 presents descriptive statistics of the UMUX-LITE scores for each study dataset segmented by type of system critical feedback.

Table 1. Descriptive statistics of the UMUX-LITE scores for each study dataset segmented by type of system critical feedback.

Dataset	Critical feedback	Mean	Median	SD	95% C.I.
Exp1	Negative	45.75	50.00	33.68	(36.07, 55.42)
Exp1	Positive	85.67	91.67	18.25	(78.14, 93.20)
Exp2	Negative	47.02	50.00	35.51	(40.37, 53.67)
Exp2	Positive	75.97	83.33	27.53	(67.50, 84.44)
Exp3	Negative	38.76	33.33	36.02	(26.69, 47.83)
Exp3	Positive	79.58	87.50	25.53	(71.42, 87.75)
Aggregated	Negative	44.42	50.00	35.29	(39.77, 49.07)
Aggregated	Positive	79.55	91.67	24.97	(74.79, 84.32)

3.1 Questionnaire Reliability Analysis

Reliability analysis on the cross-country dataset showed that the UMUX-LITE scale had high reliability; $\alpha=0.828$, $N=332$ respondents. Cronbach's alpha coefficients for all three experiments were also above the 0.70 threshold for adequate internal consistency [35]; exp1: $\alpha=0.829$, $N=74$, exp2: $\alpha=0.774$, $N=155$, and exp3: $\alpha=0.904$, $N=103$ respectively. These results align with previous research [23–25] on the good reliability of the UMUX-LITE scale.

3.2 Effect of System Critical Feedback on Perceived Usability

A two-tailed Mann-Whitney test found that the UMUX-LITE scores provided by participants who had received positive critical feedback by the system ($M=79.55$, $SD=24.97$) were significantly higher compared to the ones provided by participants

who had received negative critical feedback by the system (M=44.42, SD=35.29); z=8.301, p<0.001, r=0.460. A non-parametric test was used because the assumption of normality was violated for both the negative and positive feedback groups; W(224)=0.898, p<0.001 and W(108)=0.809, p<0.001 respectively.

Fig. 1 presents the mean UMUX-LITE scores per experiment segmented by critical feedback received by the participants. In all three cases, the assumption of normality was violated for at least one level of the independent variable, thus non-parametric tests were used; Shapiro-Wilk tests, p<0.001. Two-tailed Mann-Whitney tests found a significant effect of system critical feedback on loan applicants' UMUX-LITE scores; exp1: z=4.629, p<0.001, r=0.538, exp2: z=4.524, p<0.001, r=0.363, and exp3: z=5.323, p<0.001, r=0.524, respectively. Analytically for each experiment, the mean UMUX-LITE score of the participants who received positive critical feedback (i.e., had their loans approved) was rather high; exp1: M=85.67, SD=18.25, exp2: M=75.97, SD=27.53, and exp3: M=79.58, SD=25.53. By contrast, the participants in the negative feedback condition gave significantly lower UMUX-LITE scores; exp1: M=45.75, SD=33.68, exp2: M=47.02, SD=35.51, and exp3: M=38.76, SD=36.02.

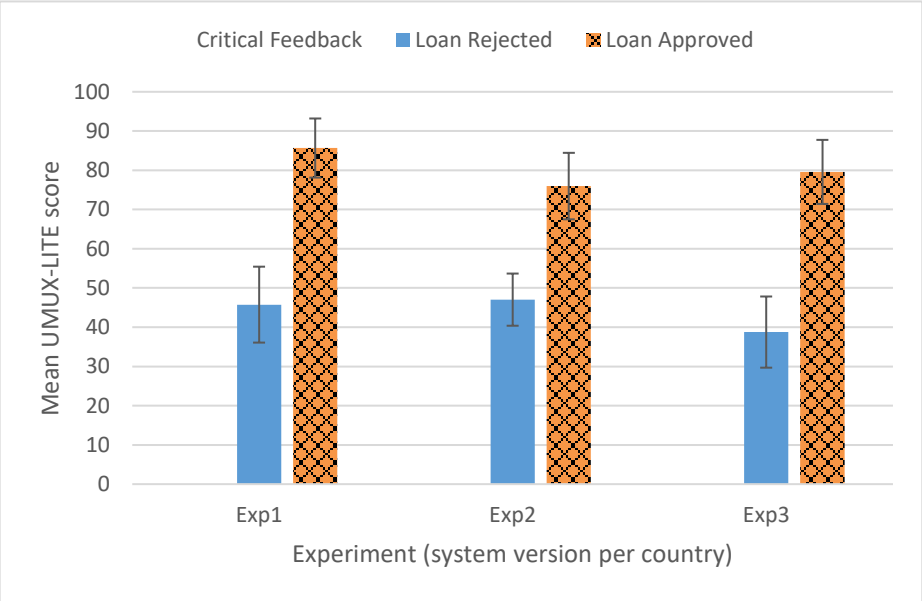

Fig. 1. The mean UMUX-LITE scores per experiment segmented by critical feedback received by the participants. Error bars represent the 95% confidence interval.

In sum, we consistently found that the participants who received negative critical feedback rated the perceived usability significantly lower than the participants who received positive critical feedback. In all the analyses, a medium to large effect size was observed [36], which demonstrates the magnitude of bias introduced to the participants' perceived usability ratings due to the critical feedback they had received.

3.3 System Critical Feedback and UMUX-LITE Questions

Our study participants interacted with the evaluated system to apply for a loan. Strictly speaking, this is the user task they had to perform. However, participants' overall goal was to get a loan. The latter goal was either met or not met based on the system decision which either approved (positive feedback) or rejected (negative feedback) their loan application.

UMUX-LITE is constructing a single usability score from the replies to two questions: "Q1. This system's capabilities meet my requirements" and "Q2. This system is easy to use". Although UMUX-LITE is a unidimensional measure of perceived usability, users who received negative critical feedback could interpret the first question as that the system didn't have enough capabilities to meet their requirements, that is getting a loan. At the same time, ease-of-use which is measured directly by the second item shouldn't be affected by the system's critical feedback. Thus, the observed differences in UMUX-LITE scores could be affected by lower scores in the first question only, which in turn might raise concerns for using this scale in our study and for systems providing critical feedback in general.

We conducted hypothesis testing on the cross-country aggregated dataset to investigate whether the type of system critical feedback (positive, negative) had a significant effect on the ratings of each UMUX-LITE question. The assumption of normality was violated for at least one level of the independent variable for both UMUX-LITE questions, thus non-parametric tests were used; Shapiro-Wilk tests, $p<0.001$. Two-tailed Mann-Whitney tests found a significant effect of system critical feedback on participants ratings for both UMUX-LITE questions; Q1: $z=8.073$, $p<0.001$, $r=0.443$, and Q2: $z=6.756$, $p<0.001$, $r=0.371$ respectively. Study participants who had their loans rejected provided significantly lower ratings for both questions (Q1: M=3.33, SD=2.33, Q2: M=4.00, SD=2.37) compared to users who had their loans approved (Q1: M=5.67, SD=1.78, Q2: 5.88, SD=1.40).

The same pattern was observed after analysing each of the three datasets separately (see Fig. 2 and Fig. 3). Again, the assumption of normality was violated in all cases and thus non-parametric tests were used; Shapiro-Wilk tests, $p<0.001$. Two-tailed Mann-Whitney tests found a significant effect of system critical feedback on loan applicants' ratings for UMUX-LITE Q1; exp1: $z=4.663$, $p<0.001$, $r=0.542$, exp2: $z=4.532$, $p<0.001$, $r=0.364$, and exp3: $z=4.798$, $p<0.001$, $r=0.473$, respectively. Participants' ratings for UMUX-LITE Q2 were also significantly affected by the system critical feedback; exp1: $z=3.680$, $p<0.001$, $r=0.428$, exp2: $z=3.337$, $p<0.001$, $r=0.268$, and exp3: $z=5.185$, $p<0.001$, $r=0.511$, respectively.

In sum, we consistently found that participants' ratings in both UMUX-LITE questions were significantly lower for participants in the negative feedback condition compared to ones in the positive feedback condition.

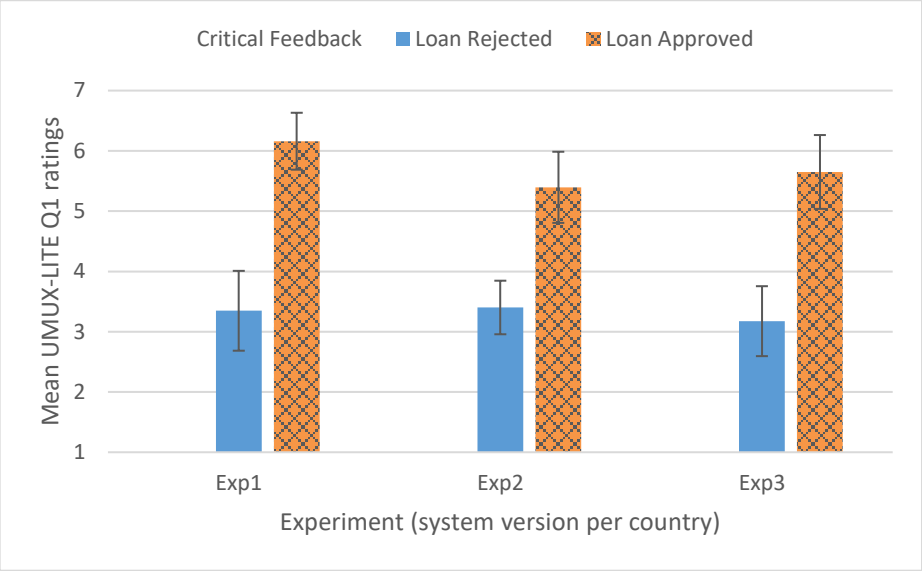

Fig. 2. The mean ratings for UMUX-LITE Q1 ("This system's capabilities meet my requirements") per experiment segmented by critical feedback received by the participants. Error bars represent the 95% confidence interval.

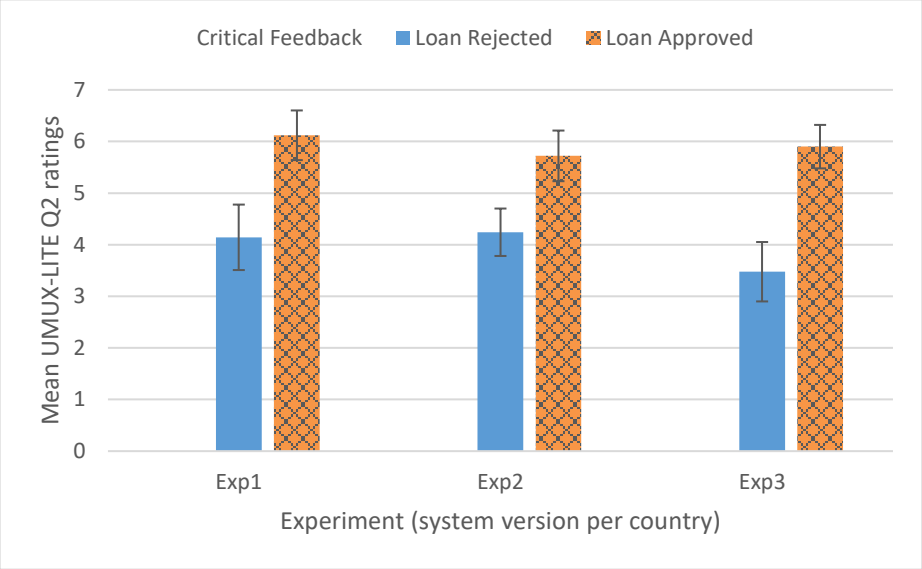

Fig. 3. The mean ratings for UMUX-LITE Q2 ("This system is easy to use") per experiment segmented by critical feedback received by the participants. Error bars represent the 95% confidence interval.

4 Discussion

4.1 Why Is Perceived Usability Affected by System Critical Feedback?

Subjective over Objective Effort. We can define "objective effort" as the observed user effort, measured by objective, deterministic metrics like time-to-complete-task, number-of-errors, etc. By contrast, "subjective effort" is the mental effort required to complete the task as perceived by the user. Subjective effort doesn't always correlate with objective effort. For example, it has been found that uncertain choices are significantly more subjectively effortful than random [37].

A possible explanation of the observed difference in perceived usability ratings might be that users tend to assess usability compared to final outcome. In our study, the objective effort required to get to the end of the loan application process was the same regardless of the critical feedback received. However, the subjective effort of the users might be different due to the final outcome (i.e., loan application approved or rejected). For users who received positive feedback, the subjective effort could be characterized as low, given that they received a loan for their effort. For users who received negative feedback, the subjective effort could be characterized as high since they didn't receive a loan regardless of their effort. That differentiation in perceived effort could have driven the differentiation of the perceived usability scores we observed in our study results.

In Section 3.3, we presented an analysis focusing on each separate UMUX-LITE question, which might be viewed as a first attempt to explore this differentiation in perceived usability scores. However, this perspective assumes that the first question of UMUX-LITE reliably captures only the subjective effort, whereas the second question measures only the objective effort, both of which are assumptions that we cannot support by referring to the literature or our own data analysis.

Complementary explanations. Based on the data presented in Table 1, one can observe that the variance of scores tends to be higher for users who received negative critical feedback compared to the ones who received positive critical feedback.

That could be an indication that users who received positive critical feedback evaluated the system from a rather utilitarian point of view (i.e., "I needed to do these actions to receive the loan, and for that reason I evaluate the usability as x"), thus relying more on the system functionality for their perceived usability rating, which might also explain the lower variance of the values. By contrast, users who received negative critical feedback might have relied more on their expectation of the minimum objective effort to receive negative feedback and/or their ratings might have been hedonically driven by their disappointment for having their loan rejected, thus making their evaluations more varied.

In any case, sentiment could play a particularly important role in usability ratings of systems providing critical feedback. Users who get positive critical feedback might tend to have more positive view of the system; users who get negative critical feedback might tend to have more negative view of the system.

4.2 Issues to Consider When Measuring Perceived Usability of Systems Providing Critical Feedback

When to Collect Perceived Usability Ratings? One might argue that perceived usability of systems providing critical feedback should be measured before users receive any type of feedback (positive or negative). This would certainly nullify the effect of system critical feedback on users' perceived usability ratings. However, this excludes from the evaluation the UI part that communicates the decision to the user and any task(s) that might rely on the user knowing the system feedback (e.g., issue the loan online after the loan application has been approved).

In addition, measuring perceived usability before users receive any critical feedback might be particularly challenging to achieve in some cases due to practical constraints (e.g., unmoderated remote user testing and the questionnaire can be made available only through a separate URL/screen after interacting with the evaluated system). At the same time, measuring perceived usability pre-admission of the critical feedback is adding unnecessary delay towards achieving the user's goal (e.g., get the loan). The user is asked to answer a usability questionnaire before getting the outcome, which potentially increases the risk of skipping the questionnaire completion or answering the questions without paying the required attention to do so.

Should We Split Perceived Usability Ratings by Feedback Type? It could be argued that splitting participants' ratings by feedback type (positive, negative) would provide a cleaner measure of perceived usability. However, this might not always be practical (e.g., too few users in one of the feedback types), and we would still be missing a single score that reflects the overall perceived usability of the system. To this end, the perceived usability scores per feedback type group could be first calculated and then they could be combined in a (weighted) average. Finding what weights to use, if any, could be the objective of future research.

5 Conclusion and Future Work

Questionnaires, such as SUS, UMUX and UMUX-LITE, are typically used to measure perceived usability in HCI practice. Previous research has explored how ratings in usability questionnaires are affected by user characteristics, such as gender, age, personality traits and previous experience with the evaluated system. However, there has been little investigation into how the nature of the system may impact the assessed perceived usability.

This paper explores whether there is any user bias towards the perceived usability of a system, when the system in question administers critical feedback, not administered with the objective of learning, rather as a final response. To this end, we gathered perceived usability data from three, almost identical, web-based systems used to apply for a consumer loan and communicate the decision to the applicant. The dataset involves a total of 332 applicants, 108 with approved loans and 224 with rejected loans, who had

also answered the UMUX-LITE after receiving the system decision for their loan. Results showed a significant effect of system critical feedback on loan applicants' UMUX-LITE scores. Participants who received positive system feedback provided significantly and largely higher UMUX-LITE scores compared to participants who received negative system feedback.

One possible limitation of the presented research is that it adopts a between-subjects research design: one group of participants received the positive system feedback and another one received the negative system feedback. This means that individual user characteristics might have affected the findings. Future research could employ a one-group pretest-posttest research design by asking from the same group of participants to provide perceived usability ratings both before and after being exposed to the system critical feedback.

Additional research with systems providing critical feedback in other domains is also needed to confirm that our findings are generalizable. In case this study's results are reproducible in other domains, it is worth investigating how to systematically analyse perceived usability ratings for such systems in order to calculate the overall "real" perceived usability. Furthermore, we know from our previous research that user-reported emotional ratings (e.g., valence-arousal ratings) are also affected by user characteristics [38, 39]. Our future work also involves investigating the effect of system critical feedback on user-reported emotional ratings.

For the time being, we advise usability researchers and professionals to pay attention when measuring critical feedback administering systems as users' perceived usability tends to be significantly biased from the critical feedback they received. We encourage anyone who is evaluating such systems to segment the perceived usability data by feedback type to further investigate the possible bias.

Acknowledgments. We thank Multitude for allowing us to use the collected usability data for our research purposes in this paper. We also thank the anonymous participants that volunteered to assess the perceived usability of the evaluated system and thus made this research possible.

References

1. Hertzum, M.: Images of usability. International Journal of Human–Computer Interaction. 26, 567–600 (2010). https://doi.org/10.1080/10447311003781300.
2. Brooke, J.: SUS: a "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., and McClelland, A.L. (eds.) Usability Evaluation in Industry. Taylor and Francis, London (1996).
3. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction. 7, 57–78 (1995). https://doi.org/10.1080/10447319509526110.
4. Lund, A.M.: Measuring usability with the USE questionnaire. Usability interface. 8, 3–6 (2001).

5. Lin, H.X., Choong, Y.-Y., Salvendy, G.: A proposed index of usability: A method for comparing the relative usability of different software systems. Behaviour & Information Technology. 16, 267–277 (1997). https://doi.org/10.1080/014492997119833.
6. Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 213–218. ACM, New York, NY, USA (1988). https://doi.org/10.1145/57167.57203.
7. Katsanos, C., Tselios, N., Xenos, M.: Perceived usability evaluation of learning management systems: a first step towards standardization of the System Usability Scale in Greek. In: Proceedings of the16th Panhellenic Conference on Informatics (PCI 2012). pp. 302–307 (2012).
8. Blažica, B., Lewis, J.R.: A Slovene translation of the System Usability Scale: the SUS-SI. International Journal of Human–Computer Interaction. 31, 112–117 (2015). https://doi.org/10.1080/10447318.2014.986634.
9. Borkowska, A., Jach, K.: Pre-testing of Polish translation of System Usability Scale (SUS). In: Proceedings of 37th International Conference on Information Systems Architecture and Technology. pp. 143–153. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-46583-8_12.
10. Taheri, F., Kavusi, A., Faghihnia Torshozi, Y., Farshad, A.A., Saremi, M.: Assessment of validity and reliability of Persian version of System Usability Scale (SUS) for traffic signs. Iran Occupational Health. 14, 12–22 (2017).
11. Wang, Y., Lei, T., Liu, X.: Chinese System Usability Scale: translation, revision, psychological measurement. International Journal of Human–Computer Interaction. 36, 953–963 (2020). https://doi.org/10.1080/10447318.2019.1700644.
12. Gao, M., Kortum, P., Oswald, F.L.: Multi-language toolkit for the System Usability Scale. International Journal of Human–Computer Interaction. 36, 1883–1901 (2020). https://doi.org/10.1080/10447318.2020.1801173.
13. Hvidt, J.C.S., Christensen, L.F., Sibbersen, C., Helweg-Jørgensen, S., Hansen, J.P., Lichtenstein, M.B.: Translation and validation of the System Usability Scale in a Danish mental health setting using digital technologies in treatment interventions. International Journal of Human–Computer Interaction. 36, 709–716 (2020). https://doi.org/10.1080/10447318.2019.1680922.
14. Katsanos, C., Tselios, N., Liapis, A.: PSSUQ-GR: a first step towards standardization of the Post-Study System Usability Questionnaire in Greek. In: CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter. p. Article23:1-Article23:6. ACM, New York, NY, USA (2021).
15. Erdinç, O., Lewis, J.R.: Psychometric Evaluation of the T-CSUQ: The Turkish version of the Computer System Usability Questionnaire. International Journal of Human–Computer Interaction. 29, 319–326 (2013). https://doi.org/10.1080/10447318.2012.711702.
16. Al-Hassan, A.A., AlGhannam, B., Naser, M.B., Alabdulrazzaq, H.: An Arabic translation of the Computer System Usability Questionnaire (CSUQ) with psychometric evaluation using Kuwait university portal. International Journal of Human–Computer Interaction. online, 1–8 (2021). https://doi.org/10.1080/10447318.2021.1926117.
17. Lewis, J.R.: The System Usability Scale: past, present, and future. International Journal of Human–Computer Interaction. 34, 577–590 (2018). https://doi.org/10.1080/10447318.2018.1455307.
18. Tullis, T., Stetson, J.: A comparison of questionnaires for assessing website usability. In: Usability Professionals Association (UPA) 2004 Conference. pp. 7–11 (2004).

19. Finstad, K.: The Usability Metric for User Experience. Interacting with Computers. 22, 323–327 (2010). https://doi.org/10.1016/j.intcom.2010.04.004.
20. Lewis, J.R., Utesch, B.S., Maher, D.E.: UMUX-LITE: when there's no time for the SUS. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 2099–2102 (2013).
21. Lewis, J.R.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. SIGCHI Bull. 23, 78–81 (1991). https://doi.org/10.1145/122672.122692.
22. Sauro, J., Dumas, J.S.: Comparison of three one-question, post-task usability questionnaires. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1599–1608. ACM, New York, NY, USA (2009). https://doi.org/10.1145/1518701.1518946.
23. Lewis, J.R.: Measuring perceived usability: The CSUQ, SUS, and UMUX. International Journal of Human–Computer Interaction. 34, 1148–1156 (2018). https://doi.org/10.1080/10447318.2017.1418805.
24. Borsci, S., Buckle, P., Walne, S.: Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? Applied Ergonomics. 84, 103007 (2020). https://doi.org/10.1016/j.apergo.2019.103007.
25. Lewis, J.R., Utesch, B.S., Maher, D.E.: Investigating the correspondence between UMUX-LITE and SUS Scores. In: Marcus, A. (ed.) Design, User Experience, and Usability: Design Discourse. pp. 204–211. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-20886-2_20.
26. Granic, A., Cukusic, M.: Usability testing and expert inspections complemented by educational evaluation: a case study of an e-Learning platform. Educational Technology & Society. 14, 107–123 (2011).
27. Bangor, A., Kortum, P., Miller, J.: An empirical evaluation of the System Usability Scale. International Journal of Human-Computer Interaction. 24, 574–594 (2008). https://doi.org/10.1080/10447310802205776.
28. Orfanou, K., Tselios, N., Katsanos, C.: Perceived usability evaluation of learning management systems: Empirical evaluation of the System Usability Scale. The International Review of Research in Open and Distributed Learning. 16, 227–246 (2015).
29. Berkman, M.I., Karahoca, D.: Re-Assessing the Usability Metric for User Experience (UMUX) scale. Journal of Usability Studies. 11, 89–109 (2016).
30. Kortum, P., Bangor, A.: Usability ratings for everyday products measured with the system usability scale. International Journal of Human-Computer Interaction. 29, 67–76 (2013). https://doi.org/10.1080/10447318.2012.681221.
31. Kortum, P., Oswald, F.L.: The impact of personality on the subjective assessment of usability. International Journal of Human–Computer Interaction. 34, 177–186 (2018). https://doi.org/10.1080/10447318.2017.1336317.
32. Sauro, J.: Does prior experience affect perceptions of usability?, https://measuringu.com/prior-exposure, last accessed 2023/01/30.
33. McLellan, S., Muddimer, A., Peres, S.C.: The effect of experience on system usability scale ratings. Journal of usability studies. 7, 56–67 (2012).
34. Cutumisu, M., Schwartz, D.L.: The impact of critical feedback choice on students' revision, performance, learning, and memory. Computers in Human Behavior. 78, 351–367 (2018).
35. Nunnally, J., Bernstein, I.: Psychometric theory. McGraw-Hill Humanities/Social Sciences/Languages (1994).
36. Cohen, J.: A power primer. Psychological Bulletin. 112, 155–159 (1992).

37. Robinson, M.M., Morsella, E.: The subjective effort of everyday mental tasks: Attending, assessing, and choosing. Motivation and Emotion. 38, 832–843 (2014).
38. Liapis, A., Katsanos, C., Xenos, M., Orphanoudakis, T.: Effect of personality traits on UX evaluation metrics: a study on usability issues, valence-arousal and skin conductance. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. p. LBW2721:1-LBW2721:6. ACM, New York, NY, USA (2019). https://doi.org/10.1145/3290607.3312995.
39. Liapis, A., Katsanos, C., Sotiropoulos, D., Xenos, M., Karousos, N.: Stress recognition in human-computer interaction using physiological and self-reported data: A study of gender differences. In: Proceedings of the 19th Panhellenic Conference on Informatics. pp. 323–328. ACM, New York, NY, USA (2015). https://doi.org/10.1145/2801948.2801964.