



# Unlocking the Mysteries of Voice User Interfaces: Subjective Workload and Usability Insights



Pushkar Chavan

College of Social Sciences & Law, University College Dublin, Belfield, Dublin 4

## Introduction

Voice user interfaces (VUIs) have become increasingly popular in our lives over the past few years. It is essential to comprehend how the subjective workload and usability of these VUIs are interconnected to enhance the user experience.

In the pursuit of uncovering this intricate connection, a study involving 100 participants was conducted and their experiences were recorded while completing everyday tasks using Amazon Alexa via an Echo smart speaker. Following these tasks, participants provided valuable insights through two distinct questionnaires: the Raw NASA Task Load Index (RTLX) and the System Usability Scale (SUS). The RTLX assessed the subjective workload across various dimensions including mental demand, physical demand, temporal demand, performance, effort, and frustration, while the SUS measured the perceived usability of the system.

This poster aims to show if there is a statistically significant relationship between the subjective workload and the usability of VUIs and throw light on the relationship between the two variables while also discussing the design, limitations, and future work.

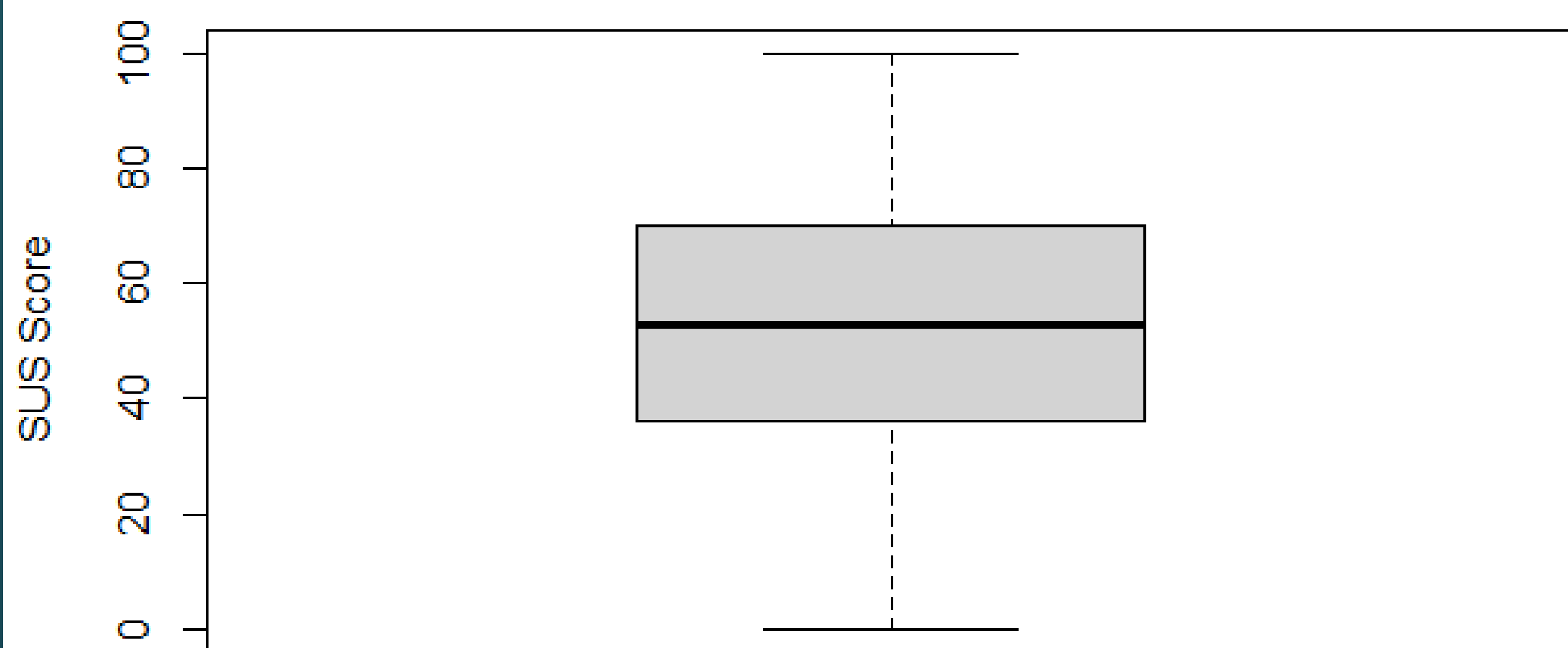
## Results

Table 1. Descriptive Statistics of the recorded scores

Parameters	SUS Score	RTLX Score
Minimum value	0	20
Maximum value	100	62
Mean	53.698	42.62
Median	53.075	42.5
Interquartile Range	33.125	13.25

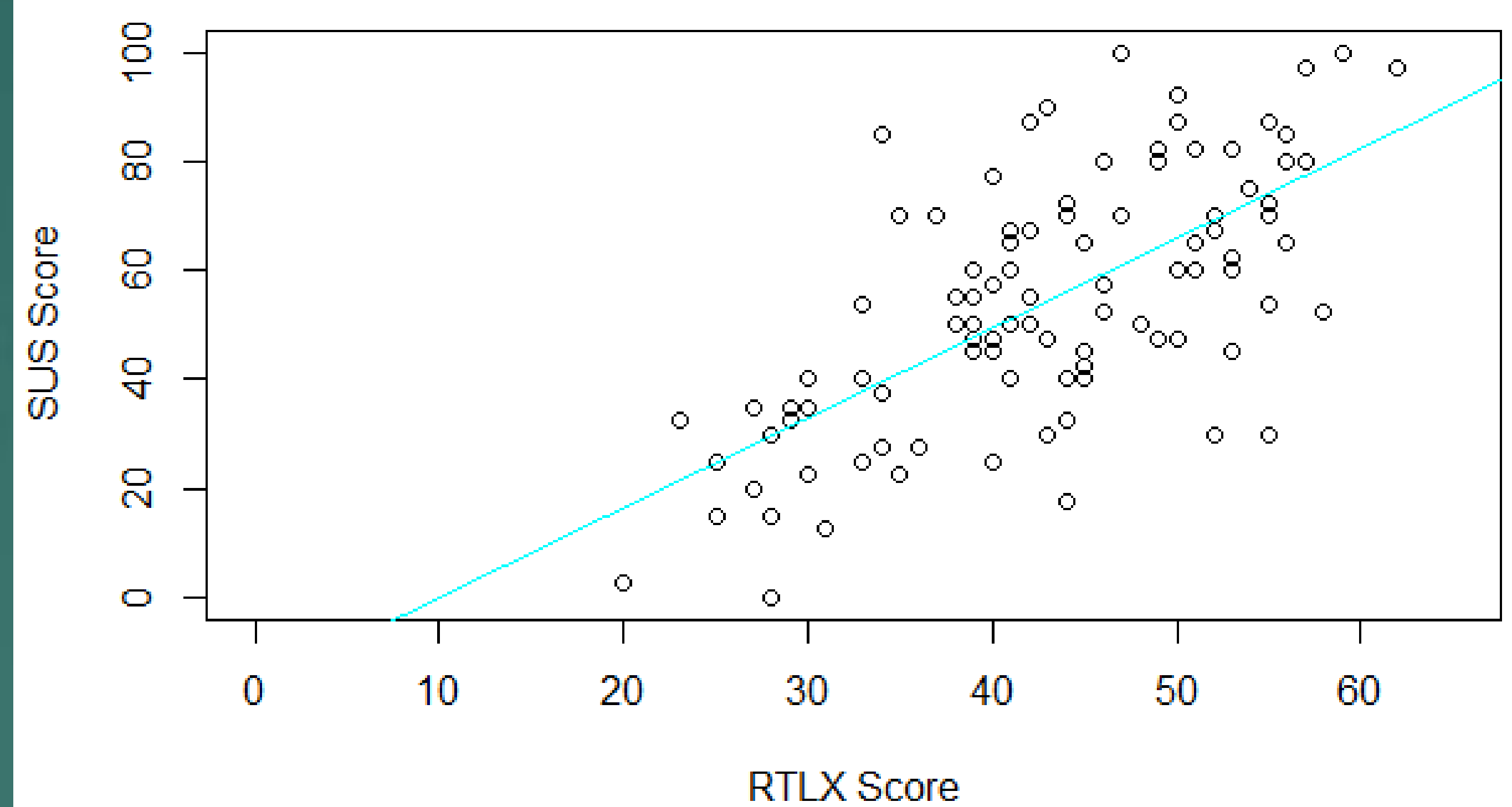
- Initial analysis showed that the mean and the median for both scores were quite close to each other indicating a near-even distribution of the scores. This can be visualized in the Figure 1 below where we see that there is a symmetrical distribution of SUS scores relative to the median line.

Figure 1. Boxplot of the System Usability Scale score



- Pearson's product-moment correlation test between the two recorded scores indicated that there is an evidence of a linear relationship. The p-value of the test was  $2.12 \times 10^{-14}$ , which was less than the significance level alpha of 0.05. It was concluded that SUS Score and RTLX Score were significantly correlated where  $r(98) = 0.67$ ,  $p < 0.05$ .
- This supports the research hypothesis H1, which states that 'There will be a statistically significant relationship between RTLX and SUS'. It implies that as the Subjective workload on a user using a VUI increases, the usability of that system also increases.

Figure 2. Scatter plot of the System Usability Scale score



- The study shows that participants' level of Relative task load predicted their System Usability,  $R^2 = 0.445$ ,  $F(1,98) = 80.39$ ,  $p < .001$ .
- The p-value of the model is  $< .001$  so the model is statistically significant in explaining variance in the SUS score.
- The straight line which predicts the System Usability Score is:  
 $SUS = 1.6472(RTLX) - 16.5056$

## Discussion

- The results support the hypothesis that a significant correlation is seen wherein as the task load increases, the system becomes increasingly usable.
- It would be insightful to know how the system usability fares against higher values of the RTLX spectrum which would likely occur with a sample of technologically averse or uninitiated participants. These findings may enhance system accessibility.
- Results from a study(Longo, Dondio,2015) advise that considering an increase in mental workload, if the interface is slightly altered and the tasks are enjoyable then the perception of good usability is strengthened. Whereas when tasks are not straightforward, perception of usability can be adversely affected with even a slight alteration of the interface.
- The unavailability of a breakdown of workload and usability findings for each task made it challenging to discern if certain tasks disproportionately influenced the overall results.
- SUS is primarily meant for testing GUI-based systems. A novel approach by one study(Zwakman et al., 2021) proposes a VUS(Voice Usability Scale) which is relatively efficient at testing VUIs. This, along with a larger sample size and the use of different voice assistants would be beneficial for mitigating the limitations of this study.

## References

- Longo, L., & Dondio, P. (2015). On the relationship between perception of usability and subjective mental workload of web interfaces. In 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) (Vol. 1, pp. 345-352). IEEE.
- Zwakman, D.S., Pal, D., Arpikanondt, C.(2021). Usability Evaluation of Artificial Intelligence-Based Voice Assistants: The Case of Amazon Alexa. SN Computer Science;2(1):28. doi: 10.1007/s42979-020-00424-4.

## Contact Information

Pushkar Chavan

Phone: +353 899643126

Email: [pushkar.chavan@ucdconnect.ie](mailto:pushkar.chavan@ucdconnect.ie)