An Analysis on the Relationship between System Usability Scale and Raw NASA Task Load Index

Data Cleaning

It is important to note that the scale of RTLX ranges from a minimum total score of 0 to a maximum total score of 126. And the scale of SUS ranges from a minimum total score of 0 to a maximum total score of 100. But we obtained two values on the SUS column in our dataset that had a negative value and a value over 100. Thus, some data cleaning was required.

Since our total number of participants were 100, and only 2 values needed to be cleaned, the data has been deleted rather than being modified or manipulated.

Results

The task of this paper is to research the following experimental Hypothesis H1: There will be a statistically significant relationship between RTLX and SUS. Indicating that a positive relationship between the System Usability Scale and Raw NASA Task Load Index, a statistically significant positive correlation was discovered [$r(96) = 0.68$, $p<0.001$]. H1 is therefore supported and the null hypothesis (H0) is rejected.

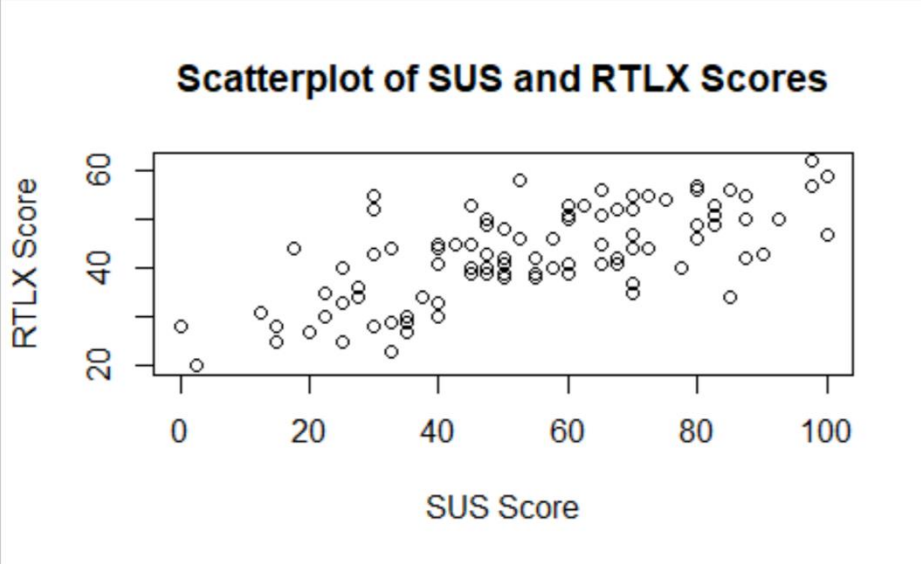

Figure 1: Scatterplot of SUS and RTLX Scores

Discussion and Related Works

This paper aims at developing on a lot of existing research on the relationship between usability of artefacts and the mental workload of humans interacting with them, specifically via the scale of SUS and RTLX. Our analysis indicates that the increase in usability of an artefact leads to the increase in difficulty of mental workload of the human. This result was obtained by using Pearson's correlation which resulted in a significant positive relationship between SUS and RTLX

Usability and Workload have been major factors that govern the efficiency of labour power. Research in the past has indicated that these 2 factors may be correlated (Giovanni, 2022). Research has shown us that the usability of an artefact based on its language has significant influence on the mental workload that the user experiences. Usability as a term has no specific measure and is accumulated by various factors. But tools such as SUS have managed to discover ways that can claim justice in their methods to measure usability. (Brooke, 1996)

Multi-year research programs have led to the development of NASA-TLX (Task Load Index) that aims at understanding and measuring workload from the experience of the human rather than the tasks. *"Workload is not an inherent property, but rather emerges from the interaction between the requirements of a task."* (Hart & Staveland, 1988). Expanding on such concepts, Mental workload has been observed to be correlated to accessibility to the language in use (Wu et al., 2020).

Research exists from the past in particular fields such as web interfaces that challenge the above observations. It has been determined that there might be no correlation between the usability of a web interface to the mental workload of a user (Longo & Dondio, 2015).

Limitations and Future Research

While the topic of discussion in this paper revolves around workload, it is important to note that we do not have information on the kind of mental workload the questionnaires could have put on the participants of the survey. There is no information on the accessibility of the survey to its participants in its language or the usability of the medium through which it was carried out. Since we are discussing the question of Usability's influence on Workload, we do not have evidence to comment on the survey's usability and the workload it created on the participants. Research in the past has indicated that workload has a negative impact on performance (Janib et al., 2021).

The study in this paper is cross sectional and cannot make any causal conclusions. No direct manipulations of variables were possible which

could allow us to make any required causal inferences. While SUS and RTLX are comprehensive tools to measure usability and workload respectively, future research could benefit well by treating them as two of many other factors rather than the primary factors.

Further, the RTLX remains a product of a program that is over three decades old. While newer versions have been proposed, the makers have noted that a useful ‘redline’ that can be applied across all applications and tasks is still far away (Hart, 2006). Research around SUS has observed that developments like Hybrid-SUS could significantly improve in obtaining more accurate information. While there is limited research that compares SUS to H-SUS, studies claim the possibility that H-SUS will positively affect psychometric properties and help individuals gain confidence in making their choices (Baumgartner et al., 2021).

Conclusion

By observing a statistically significant positive relationship between System Usability Scale (SUS) and Raw NASA Task-Load Index (RTLX) using data collected from multiple participants, this study found that there is a positive correlation between Usability and the Workload experienced by a human.

References

Brooke, John. (1996). SUS -- a quick and dirty usability scale.

Baumgartner, J., Ruettgers, N., Hasler, A., Sonderegger, A., & Sauer, J. (2021). Questionnaire experience and the hybrid System Usability Scale: Using a novel concept to evaluate a new instrument. *International Journal of Human - Computer Studies*, *147*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139-183). North-Holland.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Janib, J., Rasdi, R. M., Omar, Z., Alias, S. N., Zaremohzzabieh, Z., & Ahrari, S. (2021). The Relationship between Workload and Performance of Research University Academics in Malaysia: The Mediating Effects of Career Commitment and Job Satisfaction. *Asian Journal of University Education*, *17*(2), 85–99.

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.

Serra, G., Falco, F. D., Maggi, P., Piano, R. D., & Nocera, F. D. (2022). Website complexity and usability: is there a role for mental workload? *International Journal of Human Factors and Ergonomics*, *9*(2), 182–199.

Wu, Y., Edwards, J., Cooney, O., Bleakley, A., Doyle, P. R., Clark, L., ... & Cowan, B. R. (2020). Mental workload and language production in non-native speaker IPA interaction. In Proceedings of the 2nd Conference on Conversational User Interfaces (pp. 1-8).

R Script

```
#This is the R Script for Assignment 1

setwd("~\\Documents\\RQDA")

#The dataset that we shall be working on is the SusRtlx.csv

SusRtlx = read.csv("SusRtlx.csv")

head(SusRtlx)
tail(SusRtlx)

#From analyzing Head and Tail, we can see that there are 100 rows signifying the 100 participants that we have in our survey.

#We shall now extract all our relevant variables from this dataset.
#First we shall do it on SUS.Score

min(SusRtlx$SUS.Score)
max(SusRtlx$SUS.Score)
mean(SusRtlx$SUS.Score)
median(SusRtlx$SUS.Score)
IQR(SusRtlx$SUS.Score)
summary(SusRtlx$SUS.Score)
sd(SusRtlx$SUS.Score)

#From here we can see that there is a negative value. While we are not sure how many more negative values are there, this is a sign that we need to clean this data.

#We also notice that the max value is over 100 though the Maximum SUS Score is 100.

#Before cleaning the data, let's check RTLX.Score as well.

min(SusRtlx$RTLX.Score)
max(SusRtlx$RTLX.Score)
mean(SusRtlx$RTLX.Score)
median(SusRtlx$RTLX.Score)
IQR(SusRtlx$RTLX.Score)
summary(SusRtlx$RTLX.Score)
sd(SusRtlx$RTLX.Score)

#Check how many values in SUS are below zero and above 100.

which(SusRtlx$SUS.Score<0)
which(SusRtlx$SUS.Score>100)
which(SusRtlx$RTLX.Score<0)
```

```
#Only One value in SUS.Score is negative and that is the 92nd entry. And only one value is over 100 which is the 88th entry.
#Since it is just 2 values, we can delete them.

SusRtlx[92,2]
SusRtlx$SUS.Score[92]
tail(SusRtlx)
View(SusRtlx)

data.clean = SusRtlx[-92,]
data.clean = data.clean[-88,]
View(data.clean)

#After viewing our cleaned data, we can see that there is no entry with ID 92. Neither is there an entry with 88.

#Let's now plot a histogram on SUS Score and RTLX Score

hist (data.clean$SUS.Score, main="Histogram of SUS Score", xlab='SUS Score', breaks=c (0,10,20,30,40,50,60,70,80,90,100))

hist (data.clean$RTLX.Score, main="Histogram of RTLX Score", xlab='RTLX Score', breaks=c (0,10,20,30,40,50,60,70,80,90,100,110,120))

#Both of our histograms look somewhat like bell curves. Let's now plot a boxplot.

boxplot (data.clean$SUS.Score, main="boxplot of SUS Score", ylim= c(0,100), ylab = "SUS Score")

#There don't seem to be any outliers in our box plot.

boxplot (data.clean$RTLX.Score, main="boxplot of RTLX Score", ylim= c(0,126), ylab = "RTLX Score")

#There don't seem to be any outliers in here either.

#What would our data look like if we plot the 2 values against each other?

plot (data.clean$SUS.Score, data.clean$RTLX.Score, xlab="SUS Score", ylab="RTLX Score", main="Scatterplot of SUS and RTLX Scores")

#While the Scatterplot doesn't give us significant data, there seems to be a slight tendency of an imperfectly positive covariance.

#Let's perform a correlation test to investigate this.

cor.test(data.clean$SUS.Score, data.clean$RTLX.Score)

#Our p value is <0.001
#r(96)=0.68, p<0.001
```

#We can say that our correlation is Moderately strong.