Evaluation of System Usability Scale as A Marker of Non-Human Computer Interface's Usability: A Sanitizer Container-Based Study

Banerjee S[1], Dey S[2], Gangopadhyay S[2*]

[1]Research scholar, Department of Physiology, University of Calcutta, Kolkata, India

[2]Professor, Department of Physiology, University of Calcutta, Kolkata, India

*Corresponding author:** Gangopadhyay S, Professor, Department of Physiology, University of Calcutta, 92 A.P.C Road, Occupational Ergonomics Laboratory, Kolkata-700009, India. Tel: +91 9432290030, Email: ganguly1961@gmail.com

Abstract

The use and the production of sanitizers have increased in the post-pandemic situation to prevent the further spread of COVID-19. Usability assessment of sanitizer containers is essentially required to evaluate the effectiveness, efficiency, and satisfactory use of the sanitizer containers. This study aimed to evaluate the system usability scale (SUS) as a marker of perceived usability of non-human computer interfaces, such as - sanitizer containers. The perceived usability of three types of sanitizer containers was evaluated using SUS. The authors have measured the reliability, convergent, and discriminant validity to evaluate SUS as a marker of usability of sanitizer containers. The result showed that SUS lacks convergent validity though it has a higher reliability coefficient. Thus, this is not the best measure of usability for non-human computer interfaces like-sanitizer containers. The SUS was applied on the flip cap, finger-pressure pump, and spray-type sanitizer containers. The finger-pressure pump container has exhibited a higher SUS score than others. The results of this study have given an idea about the usability of three different kinds of sanitizer containers. This research provides an overview of the application of SUS on non-Human computer interfaces like sanitizer containers. This study provides limitations like lack of convergent validity and ideas to overcome common method bias.

Keywords: Usability Testing and Evaluation, Sanitizer Usability, Validity, Reliability, Gender

Introduction

The deadly novel SARS-CoV-2 (COVID-19) originated in Wuhan, China, which resulted in a massive outbreak around the globe [1]. The World Health Organization (WHO) declared COVID-19 as a pandemic on 11th March 2020 [2]. COVID-19 spreads through different human activities like- talking, coughing, sneezing, etc [3]. Apart from the vaccination, some precautionary measures like-wearing the mask(s), maintaining social (physical) distancing, and washing hands including repeated hand sterilizing with alcohol-based sanitizer are required to stop COVID-19 from further being spread. Alcohol-based sanitizer denatures the natural structure of the protein envelope of SARS-CoV-2 [4]. As a result, the sales of hand sanitizer have hiked by 600% [5]. The hand sanitizer production capacity of India has grown 1,000 times during the Covid-19 pandemic [6]. Usability assessment of sanitizer containers is necessary for three main reasons. Firstly, usability marks out the effectiveness of a product. Secondly, usability describes the efficiency of a product [7]. Third, usability can lead to users' satisfaction. Lack of usability is the source of users' dissatisfaction [8]. Moreover, usability plays an important role in influencing the choices of users while purchasing a product [9].

Research Hypotheses and Questions

Participants of this study were divided into two groups based on their gender and their type of work. Therefore, the objectives of this study were to compare the perceived usability score of different sanitizer containers and to find out the differences in the perceived usability of sanitizer containers between the males & females as well as work from home participants & commuting worker participants. Thus, the hypotheses of this study are-

H1: System Usability Scores of different sanitizer containers are different.
H2: There is different perceived usability of sanitizer containers between males and females.
H3: There is different perceived usability of sanitizer containers between work from home participants & commuting worker participants.

Brooke developed the System Usability Scale (SUS) [10]. Initially, it was unidimensional [11]. Two latent factors were identified in 2009. The first factor comprised of eight items (items 1,2,3,5,6,7,8 and 9) concerning usability and the second factor comprised of two items (items 4 and 10) related to learnability [12, 13]. But most satisfactory model revealed that the two-factor structure was based on positive (odd-numbered items) and negative (even-numbered items) items [14]. This system usability scale can be used to assess the usability of various products, software, hardware, and websites [15]. These products include- everyday products (microwave, landline, automated teller machine, etc.), customer equipment, face coverings, and safety signs [16-19]. Nonetheless, there is no evidence of using SUS as a measure of perceived usability of different sanitizer containers. Therefore, this study aimed to evaluate the system usability scale as a marker of sanitizer containers' usability. Thus, the research question of this study is-

Q: Can SUS be a reliable and valid marker of the perceived usability of non-human computer interfaces such as sanitizer containers?

The distinctiveness of this study is that exploratory and confirmatory factor analysis of system usability scale (SUS) items along with validity and reliability tests were used to interpret the result instead of accounting only the descriptive statistics. Different statistical analyses (discussed later) were performed to compare the usability of different sanitizer containers.

Materials and Methods

Study Design

This is a descriptive type of cross-sectional study. The authors have categorized perceived usability as a dependent variable that was measured by using SUS. Three independent variables of this study are - gender, type of work, and type of sanitizer containers.

Participants Selection

Participants were randomly selected from different cities in India. A total of 135 participants (N) were selected for the usability measurement, among them 72 (53.33%) were males and 63 (46.67%) were females (there was no third gender). Further, participants were divided into two groups based on their work type- commuting workers and work from home employees. The number of commuting workers and work-from-home employees were 57 and 78 respectively. The age range of all the participants is 18-65 years. The mean (±Standard Deviation) age of participants was 28.42±8.282 years. Inclusion & exclusion criteria were satisfied based on the information acquired from the participants.

Inclusion Criteria: Daily sanitizer users were included in this study. Participants in the age group of 18-65 years were included in this study. Participants who have used all three types of containers were included in this study.

Exclusion Criteria: Participants who do not use sanitizer and those participants who do not know how to use sanitizer containers were excluded from this study.

Ethical Approval: This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Human Ethical Committee at the University of Calcutta. Informed consent was obtained from each participant.

Materials

Sanitizer Container Selection: Three (3) sanitizer containers were selected through an online market with different usability. These sanitizer containers are designated as Type 1, Type 2 & Type 3. Type 1 is a flip cap container with gel-based alcohol sanitizer. Type 2 container consists of a finger pressure pump with gel-based alcohol sanitizer. Type 3 is a spray container with liquid-based alcohol sanitizer. Figure 1 shows three kinds of sanitizer containers and their usability.

Figure 1: Three different types of sanitizer containers with their usability. (a) Type 1 is a flip cap container, (b) Type 2 is a finger pressure pump container, (c) Type 3 is a spray container in nature

Measurement of Usability: The system usability scale (SUS) was adapted to measure the usability of sanitizer containers [18]. It was a 5-point rating scale that includes - *"I strongly disagree", "I disagree", "I don't disagree nor agree", "I agree", and "I strongly agree". SUS consists of 10 reliable (Cronbach's alpha = 0.85) questions*[20, 21].

The SUS questions are as follows [15, 22]:

Item 1. I think that I would like to use this system frequently.

Item 2. I found the system unnecessarily complex.

Item 3. I thought the system was easy to use.

Item 4. I think that I would need the support of a technical person to be able to use this system.

Item 5. I found the various functions in this system were well integrated.

Item 6. I thought there was too much inconsistency in this system.

Item 7. I would imagine that most people would learn to use this system very quickly.

Item 8. I found the system very cumbersome to use.

Item 9. I felt very confident using the system.

Item 10. I needed to learn a lot of things before I could get going with this system.

The odd number questions (1,3,5,7 & 9) were positive statements and the even number questions (2,4,6,8 & 10) were negative statements [12]. To compute the overall SUS score, the following method was applied- The item score based on positive statements or odd number questions was deducted by 1 (x-1) and the item score based on negative statements or even number questions was calculated by deducting the score from 5 (5-x). The summation of these item scores was then multiplied by 2.5 to provide a total SUS score. The range of the SUS score is between 0 (extremely poor usability) and 100 (excellent usability)[23].

Table 1 represents comparison of System Usability Scale with other usability assessment methods

Questionnaire	No of items	Type of Rating Scale	No of Sub scales	Time of administration the method	Application(s)
System Usability Scale (SUS)	10	5-point Likert scale	2	Assessment of the perceived usability at the end of a study	Subjective assessments of perceived usability of products
Questionnaire for User Interaction Satisfaction (QUIS)	31	9-point Likert scale	6	Assessment of the perceived usability at the end of a study	Usability assessment of Human-computer interface
Software Usability Measurement Inventory (SUMI)	50	3-point rating scale	5	Assessment of the perceived usability at the end of a study	Usability assessment of software
Post-Study System Usability Questionnaire (PSSUQ)	16	7-point Likert scale	3	Assessment of the perceived usability at the end of a study	Perceived satisfaction assessment with computer systems or applications
After-Scenario Questionnaire (ASQ)	3	7-point Likert scale	1	Assess immediately after the completion of a usability task scenario	Assessment of overall ease of task completion, satisfaction with the completion time
Expectation ratings (ER)	2	5 or 7-point Likert scale	1	Assess immediately after the completion of a usability task scenario	Assessment of the task difficulty before and after performing the task
Single Ease Question (SEQ)	1	5 or 7-point Likert scale	1	Assess immediately after the completion of a usability task scenario	Assessment of the overall ease of completing a task
Usability Magnitude Estimation (UME)	1	Open-ended question	1	Assess immediately after the completion of a usability task scenario	Measurement of relationships between the physical dimensions of a stimulus and its perception
Subjective Mental Effort Question (SMEQ)	1	7-point rating scale	1	Assess immediately after the completion of a usability task scenario	Assessment of subjective mental effort of a task

Table 1: Comparison of System Usability Scale with other usability assessment methods

For this study, the word "System" from the questionnaire was replaced with "Sanitizer container".

Procedure

Participants were randomly asked to use three different types of sanitizer containers one after another. Because the random presentation of the containers will counterbalance and minimize the carrying over effects of the participants. The participants used sanitizer containers using the following steps-

1. Participants held the container,
2. Then participants have used the head (flip cap or pressure pump or spray) of the sanitizer container to dispense the sanitizer,
3. Participants have poured sanitizer into their palms.
4. The SUS was applied after the use of each sanitizer container.
5. This study was documented using a google form. The contents of the form were a few demographic questions along with a SUS questionnaire.

Statistical Analysis

Descriptive Statistics and Normality Test: After the collection of data, total SUS scores of three different types of sanitizer containers were expressed in terms of average (mean) with standard error (SE) and percentile values. Kolmogorov-Smirnov test was performed to check the normality of the variables [24, 25].

Comparison Test: The related-sample Friedman's two-way ANOVA (two-tailed) by ranks was performed to compare the distributions of three SUS scores [8]. After obtaining significance, Bonferroni correction was performed as a post-hoc test. The Mann-Whitney U test (two-tailed and unpaired) was performed to compare the SUS scores distributions of three types of sanitizer containers between males and females as well as between commuting workers and work from home employees. The strength of association and effect size between the parameters were measured using Pearson's correlation coefficient (r) [26]. The effect size (r) is small if the value is within 0.1 to 0.3, medium if the value is within 0.3 to 0.5 and large if r varies more than 0.5 irrespective of sign (negative or positive) [27].

Exploratory Factor Analysis (EFA): EFA was performed to evaluate the system usability scale as a marker of sanitizer containers' usability. The exploratory factor analysis (EFA) of SUS was performed by using principal component analysis (PCA) with ProMax rotation [28, 29]. EFA identifies the number of latent factors which are involved in SUS and the correlation between each item of the SUS questionnaire [11].

Confirmatory Factor Analysis (CFA): CFA has done using maximum likelihood (ML) estimation to confirm the latent factors which are involved in SUS [30]. The fitness of the model was assessed by using Chi-square (χ^2)/degrees of freedom(df), Comparative Fit Index (CFI, close to 0.95 or greater), Root Mean Square Error of Approximation (RMSEA, close to 0.06 or below) along with upper and lower limit as suggested by Brown [31].

Test for Validity: Convergent and discriminant validity of the confirmatory factor analysis model was measured by construct reliability (CR>0.70), average variance extracted (AVE>0.50) and maximum shared squared variance (MSV<AVE) calculation [32]. Heterotrait-Monotrait (HTMT) ratio of correlations was also calculated for discriminant validity (HTMT ratio correlation coefficient<0.850) [33].

Test for Reliability: The reliability assessment of the SUS was done by using the Omega hierarchical (HA) coefficient [34]. Omega is a reliability evaluation that does not pivot on the assumption of tau equivalence. There is no specific benchmark for omega to evaluate acceptable reliability but a minimum of 0.50 and values with closer proximity to 0.75 are recommended for acceptable and good reliability. The Cronbach's alpha was measured to check the reliability of this questionnaire. A Cronbach's alpha coefficient higher than 0.70 is to be considered reliable [35].

Test for Common Method Bias Detection: Common method bias (CMB) was detected using an unmeasured latent marker construct (ULMC) [30]. A common latent factor (CLF) is used to represent the CMB[36]. The standardized regression weights of the bi-factor model with and without CLF were subtracted to measure the effect of the CMB. The difference of less than 0.200 indicates that there is a less weighted effect of the common factor [37].

Previously mentioned statistical tests were performed by using Microsoft Excel 2016 and Statistical Package for the Social Sciences (SPSS) software (version 26) including AMOS (version 23) [28, 38].

Results

Descriptive Statistics

Table 2 represents the mean (±SE) System Usability Scale (SUS) score of three types of sanitizer containers which are categorized as - general users, male users, female users, commuting workers, and working from home employees.

Category	Sample size (N)	Type 1 container SUS score	Type 2 container SUS score	Type 3 container SUS score
General users	135	63.78±1.531	71.27±1.632	41.46±1.688
Male users	72	63.30±2.077	70.56±2.392	40.69±2.370
Female users	63	64.33±2.281	72.10±2.196	42.34±2.413
Commuting workers	57	64.21±2.221	71.14 ±2.095	40.57±2.513
Working from home employees	78	63.46±2.106	71.38±2.384	42.12±2.283

Table 2: Mean SUS scores (**±SE)** of sanitizer containers are represented in different categories

Individual SUS scores were arranged according to the adjective rating scale [39]. It has been found that the maximum number (80) SUS scores in the acceptable range were for type 2 sanitizer containers (Table 3). These 80 individuals for type 2 sanitizer containers comprised 59.3% of the total participants. The lowest number (11) of SUS scores were for type 3 sanitizer container. These 11 individuals for type 3 sanitizer container comprised only 8.1% of the total participants.

Sanitizer container type	SUS score rating scale	Frequency (n)	Percent (%)
Type 1	Not acceptable	26	19.3
	Marginal	57	42.2
	Acceptable	52	38.5
Type 2	Not acceptable	14	10.4
	Marginal	42	31.1
	Acceptable	**79**	**58.5**
Type 3	Not acceptable	82	60.7
	Marginal	41	30.4
	Acceptable	12	08.9

Table 3: Descriptive Statistics of SUS Scores for adjective ratings in different sanitizers containers for general users

Percentile values of three types of sanitizer containers indicated that the number of distributions of individual SUS scores is greater in type 2 sanitizer containers in the 25th, 50th (median), 75th, and 95th percentiles than other two types (Table 4). Only the 95th percentile distributions of type 1 and type 2 were the same. Whether these distributions were significantly different or not is discussed in later sections of the result section.

Sanitizer container type	Percentile			
	25th	50th (Median)	75th	95th
Type 1	50.00	62.50	75.00	95.00
Type 2	60.00	75.00	85.00	95.00
Type 3	25.00	42.50	52.50	73.00

Table 4: Percentiles of SUS Scores in different sanitizers containers for general users

Comparison Tests

The type 2 sanitizer container had a higher SUS score than type 1 and 3. This result has indicated that the participants preferred type 2 sanitizer container over type 1 and 3 in terms of usability. To confirm the differences between the above-mentioned SUS scores, several tests were performed. Kolmogorov-Smirnov normality test was performed to check the type of distribution. It has specified that the set of observations was not normally distributed ($p<0.05$) [25]. Related-samples Friedman's Two-Way ANOVA test by ranks has indicated that the distributions of the SUS scores of type1, type 2, and type 3 are significantly different ($p<0.05$). Post-hoc paired wise comparison (with Bonferroni's correction) of SUS scores of previously mentioned sanitizer containers also dissimilar with one another ($p<0.05$). Hence, the type 2 sanitizer container had the highest usability over the type 1 and type 3 sanitizer container respectively. The effect size of type 2 and type 3, type 1 and type 3 were very large (very strong) and large (strong) respectively. But the effect sizes of type 1 and type 2 were medium (moderate) [40]. The related-sample Friedman's two-way ANOVA by rank test has revealed significant differences ($p<0.05$) in perceived system usability scale scores among three different types of sanitizer containers (N= 135, Chi-square= 82.093, df= 2). Table 5 indicates the Wilcoxon signed-rank pair-wise comparisons of type 3 vs type 1 SUS score, type 3 vs type 2 SUS score, and type 1 vs type 2 SUS score along with significance level including Bonferroni adjustment correction and effect sizes.

Sample 1 vs Sample 2	Test Statistic	Standard Error	Z score Statistic	Significance	Adjusted Significance[a]	Effect Size (r)
Type 1 vs Type 3 SUS score	0.681	0.122	5.599	0.000*	0.000*	0.48
Type 2 vs Type 3 SUS score	1.063	0.122	8.733	0.000*	0.000*	0.75
Type 2 vs Type 1 SUS score	-0.381	0.122	-3.134	0.002*	0.005*	-0.27

a= Significance values have been adjusted by the Bonferroni correction for multiple tests

*=significant (level of significance=0.05)

Table 5: Post-hoc Wilcoxon signed-rank pair-wise comparison tests for type 1, type 2 & type 3 sanitizer containers along with Bonferroni adjustment correction and effect sizes

The Mann-Whitney U test has revealed that the distribution of the SUS scores was the same between the genders across all three types of sanitizer containers (intra-type) which are type 1, type 2, and type 3 respectively. In other words, there were no significantly different ($p>0.05$) distributions of SUS scores observed between males and females in the same type of sanitizer container. The comparison was as follows:

1. Type 1 SUS score of males and Type 1 SUS score of females.
2. Type 2 SUS score of males and Type 2 SUS score of females.
3. Type 3 SUS score of males and Type 3 SUS score of females.

The Mann-Whitney U test has revealed that the distribution of the SUS scores was the same between the work types across all three types of sanitizer containers (intra-type) which are type 1, type 2, and type 3 respectively. In other words, there were no significantly different ($p>0.05$) distributions of SUS scores observed between commuting workers and work from home employees in the same type of sanitizer container. The comparison was as follows:

1. Type 1 SUS score of commuting workers and Type 1 SUS score of work from home employees

2. Type 2 SUS score of commuting workers and Type 2 SUS score of work from home employees

3. Type 3 SUS score of commuting workers and Type 3 SUS score of work from home employees

There were significantly different (p<0.05) distributions of SUS scores observed for inter-type of sanitizer containers (among type 1, type 2, and type 3) in the following categories:

1. Gender-based-

a. Male

b. Female

2. Work type based-

a. Commuting workers

b. Work from home employees

Table 6 represents Related-Samples Friedman's Two-Way ANOVA by Ranks test summary of both categories. From table 6, the authors obtained that the distributions of the SUS scores of type1, type 2, and type 3 are significantly different (p<0.05) in both categories.

Category	Sub-category	Total sample (N)	Chi-square	Degrees of freedom (df)	Asymptotic Significance (2-tailed test)
Gender	Male	72	39.007	2	0.000*
	Female	63	43.414	2	0.000*
Work type	Commuting workers	57	42.369	2	0.000*
	Work from home employees	78	40.409	2	0.000*

*=significant (level of significance=0.05)

Table 6: Related-Samples Friedman's Two-Way ANOVA by Ranks test summary of SUS scores based on gender and work categories

Pairwise comparison along with Bonferroni correction has been utilized as a post-hoc test. It has been found that the SUS scores of type 1 and type 2 sanitizer containers are not significantly different in both male and female users. In other words, the distribution of the SUS scores of type 1 and type 2 sanitizer containers under the male user sub-category were the same.

A similar result was also found in the gender-based category. The SUS scores of type 1 and type 2 sanitizer containers among work from home employees and commuting workers were insignificantly distributed. In other words, the distribution of the SUS scores of type 1 and type 2 sanitizer containers under the work from home employee sub-category were identical. Table 7 indicates the pairwise comparison of the SUS scores based on gender and work type categories respectively. The effect size of type 2 and type 3 in both gender-based and work-based categories was very large (very strong). The effect size of type 1 and type 3 in both gender-based and work-based categories was large (strong) respectively. But the effect size of type 1 and type 2 in both gender-based and work-based categories was medium (moderate) [40].

Category	Sub category	SUS score of Sample1vs Sample 2	Test Statistic	Standard Error	Z score Statistic	Significance level	Adjusted significance level[a]	Effect Size (r)
Gender	Male	Type 1 vs Type 3	0.639	0.167	3.833	0.000	0.000*	0.45
		Type 2 vs Type 3	1.007	0.167	6.042	0.000	0.000*	0.71
		Type 2 vs Type 1	-0.368	0.167	-2.208	0.027	**0.082**	-0.26
	Female	Type 1 vs Type 3	0.730	0.178	4.098	0.000	0.000*	0.50
		Type 2 vs Type 3	1.127	0.178	6.325	0.000	0.000*	0.80
		Type 2 vs Type 1	-0.397	0.178	-2.227	0.031	**0.092**	-0.28
Work type	Commuting workers	Type 1 vs Type 3	0.746	0.187	3.980	0.000	0.000*	0.52
		Type 2 vs Type 3	1.175	0.187	6.275	0.000	0.000*	0.80
		Type 2 vs Type 1	-0.430	0.187	-2.295	0.022	**0.065**	-0.29
	Work from home employees	Type 1 vs Type 3	0.635	0.160	3.963	0.000	0.000*	0.45
		Type 2 vs Type 3	0.981	0.160	6.125	0.000	0.000*	0.69
		Type 2 vs Type 1	-0.346	0.160	-2.162	0.031	**0.092**	-0.25

a= Significance values have been adjusted by the Bonferroni correction for multiple tests

*=significant (level of significance) = 0.05

Table 7: Post-hoc Wilcoxon signed-rank pair-wise comparisons for type 1, type 2 & type 3 sanitizer containers for different gender

Exploratory Factor Analysis

The factor analysis of SUS was done by using principal component analysis with pro-max rotation (κ, kappa=4). The initial inspection of the correlation matrix of SUS questionnaire items indicated considerable numbers of the coefficients were above ±0.30 (Table 8). Table 8 indicates that the correlation between an odd item with other odd items was positive and an odd item with other even items was negative, and vice versa.

		Item1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Correlation co-efficient	Item1	1.000									
	Item 2	-0.329	1.000								
	Item 3	0.543	-0.446	1.000							
	Item 4	0.017	0.288	0.001	1.000						
	Item 5	0.487	-0.266	0.575	0.060	1.000					
	Item 6	-0.226	0.433	-0.198	0.347	-0.182	1.000				
	Item 7	0.426	-0.325	0.463	-0.144	0.602	-0.192	1.000			
	Item 8	-0.341	0.515	-0.301	0.476	-0.250	0.551	-0.319	1.000		
	Item 9	0.541	-0.319	0.570	-0.067	0.625	-0.242	0.573	-0.384	1.000	
	Item 10	-0.023	0.373	-0.023	0.635	0.024	0.426	-0.154	0.501	-0.074	1.000

Determinant = 0.015

Table 8: Correlation Matrix of SUS questionnaire items

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy index was 0.846 which was above the recommended value of 0.60 [41]. Bartlett's test of sphericity has reached statistical significance (χ^2=1679.873, p<0.001) [42, 43]. Communalities were the proportion of variances of each variable that can be explained by the components or factors [44]. No item had a communality (h^2) below 0.30. These results infer that the data obtained from the SUS questionnaire was suitable for principal component analysis [45, 46]. There were two factors with Eigenvalues over 1. Factor 1 and 2 have initial eigenvalues of 4.029 and 2.190, respectively. These factors cumulatively explained 62.19% of the total variances. Pattern matrix of exploratory factor analysis of ten SUS questionnaire items has revealed that there are two latent components/factors (Table 9). The first component consisted of item numbers 5, 9, 3, 7, and 1 and the second component consisted of item numbers 10, 4, 8, 6, and 2. Factor 1 or component 1 symbolized the user-friendliness of a system (sanitizer container) and Factor 2 or component 2 symbolized the perplexity (confusion and complexity) of the system (sanitizer container).

|		Item 5	Item 9	Item 3	Item 1	Item 7	Item 10	Item 4	Item 8	Item 6	Item 2
Component	1	0.863	0.818	0.809	0.756	0.717					
	2						0.880	0.844	0.727	0.672	0.531

Table 9: Pattern matrix for SUS questionnaire items

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) has confirmed the previously-mentioned issue. Further, two models were developed. The first model was developed using two latent factors- user-friendliness and perplexity as a bifactor model. The second model indicated that there may be a common method bias that was linked with the items. Table 10 indicates the model fit indices for both models. The model with a common method bias exhibited a better model fitness than the bifactor model.

	CMIN/df	CFI	RMSEA	90% lower	90% upper
Bifactor model	4.750	0.928	0.096	0.081	0.112
With Common method bias	2.593	0.979	0.063	0.043	0.083

Table 10: Model fit indices for Confirmatory factor analysis

Path diagrams of the bifactor model with and without common method bias (CMB) are provided in figure 2. Error terms (e) of item 4 and item 10 and item 5 and item 7 are interrelated because they had a 0.49 and 0.20 correlation value respectively. As suggested by Brooke, SUS items were dependent on- the product's effectiveness, efficiency, and user satisfaction. Item 5 was scored for the system integration and item 7 was scored for quick learnability. A well-integrated system can cause better learnability of that system [10]. On the other hand, learnability has a relation to usability [47], and items 4 and 10 are scored for learnability. Thus, error terms of items 4 and 10 and 5 and 7 were interlinked to form the model.
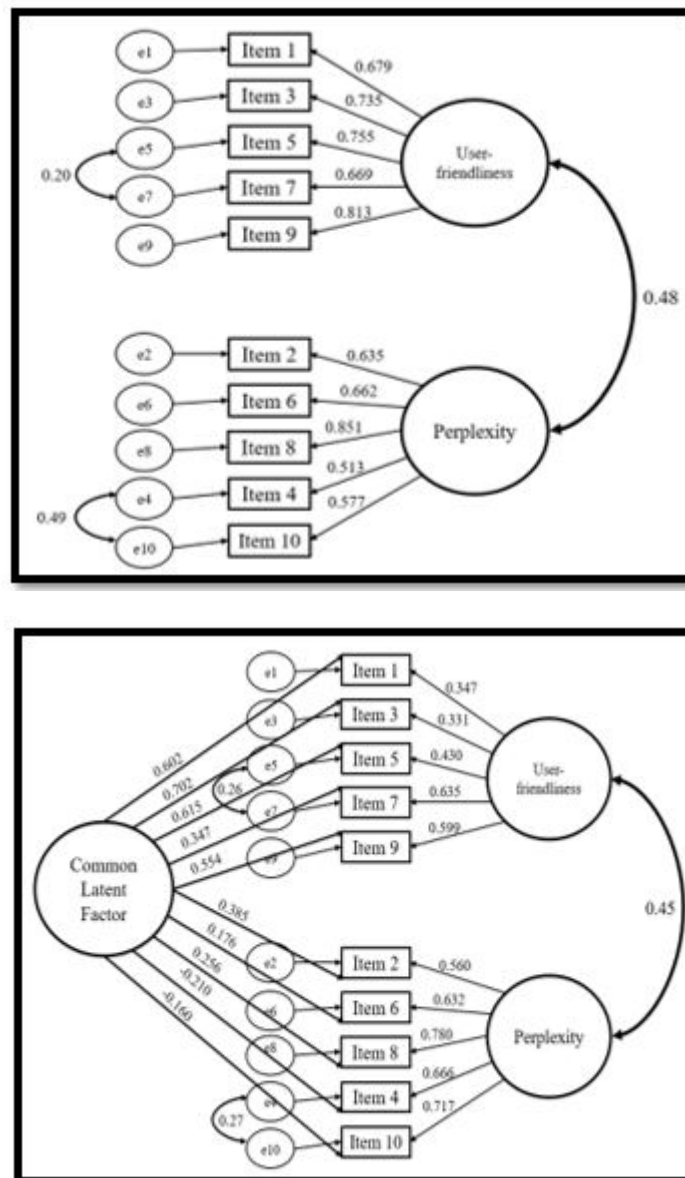
Figure 2: Path diagram of (a) bi-factor model without common latent factor (b) bi-factor model with common latent factor. The number along with unidirectional arrows are indicating the standard loadings (R) of the factors while bidirectional arrows are indicating the correlation coefficient between the factors. "e" represents the error terms

Validity

The convergent and divergent validity of the common latent factor model was explained using table 11. Construct reliability (CR) was higher (>0.700) in both of the factors (User-friendliness and perplexity) but very low for the common bias. The average variance extracted (AVE) was almost the same as the cut-off value (0.50) for the factor user-friendliness, but slightly lower than the cut-off value for the perplexity factor. The maximum shared squared variance (MSV) of the User-friendliness and perplexity was 0.225 which was lower than the AVE. These factors indicated that the convergent and divergent validity were above the threshold level. The effect of common factor bias was low.

Factors	CR	AVE	MSV
User-friendliness	0.852	0.536	0.225
Perplexity	0.787	**0.433**	0.225

Table 11: Convergent and divergent validity for the model

There was a confirmed discriminate validity among user-friendliness and perplexity because the HTMT correlation ratio of these two was 0.456 which is lower than 0.850.

Reliability

Omega and Cronbach's alpha coefficients of the SUS were reliable because both of these coefficients were well above the acceptable (>0.70) limit. Table 12 represents the Omega and Cronbach's alpha coefficients.

	All items	User-friendliness	Perplexity
Omega Coefficient (HA)	0.795	0.853	0.803
Cronbach's Alpha Coefficient	0.827	0.853	0.808

Table 12: different reliability coefficients of SUS

Common Method Bias

The post-hoc ULMC method has detected the presence of common method bias (CMB). Standardized estimates of "user-friendliness" factor items have changed more than 0.2 in the presence of CMB (Table 13).

Factor	Item no	Estimate with CLF	Estimate No CLF	Differences
User-friendliness	ITEM9	0.599	0.813	**0.214**
	ITEM7	0.635	0.669	0.034
	ITEM5	0.430	0.755	**0.325**
	ITEM3	0.331	0.735	**0.404**
	ITEM1	0.347	0.679	**0.332**
Perplexity	ITEM10	0.717	0.577	-0.140
	ITEM4	0.666	0.513	-0.153
	ITEM8	0.780	0.851	0.071
	ITEM6	0.632	0.662	0.030
	ITEM2	0.560	0.635	0.075

Table 13: Difference of standardized estimates between with and without CLF model

The chi-square (χ^2) and degrees of freedom (df) values were significantly different($p<.05$) between the bifactor model and the bifactor model with CMB.

Discussion

In this study, the authors have found the presence of two latent factors which can affect perceived usability. These factors are perceived user-friendliness and perplexity (confusion and complexity) of the product. This bifactor model has comprised of positive and negative items of the SUS which represent user-friendliness and perplexity respectively. Further, confirmatory factor analysis has established the presence of two latent factors in SUS. Model fit indices revealed that it is moderately fit. Post-hoc unmeasured latent marker construct has confirmed the presence of common method bias in the "user-friendliness" latent factor items. This bias may be due to the common scale format, common scale anchor (5-point scale), and positive-negative wording [30]. The order of presentation of the products can be a potential source of method bias. The construct reliability (CR) of convergent validity has construed that the system usability scale (SUS) can be used as a marker of sanitizer containers' usability. But the average variance explained (AVE) has failed to establish a convergent validity for the "perplexity" factor items. AVE is a strict measure of convergent validity because consideration of CR only will conclude that the convergent validity of a construct is adequate, even if more than 50% of the variance is due to the error [48]. Maximum shared squared variance (MSV) and heterotrait-monotrait (HTMT) ratio have successfully established discriminant validity. The contents of the SUS questionnaire were reliable enough to assess the usability of sanitizer containers because Omega hierarchical and Cronbach's alpha coefficients were well above the threshold limit. Furthermore, the previously-stated model has indicated the relationship between user-friendliness, perplexity, and usability. Therefore, the answer to our research question is that the SUS cannot be a valid marker of the perceived usability of sanitizer containers though it has a decent reliability coefficient. Table 14 represents the comparison of the reliability coefficients (Cronbach's alpha) of different studies to validate the effectiveness of the proposed experimental condition [18].

Study	Sample size	Reliability (Coefficient alpha)
Present Study	135	0.827
Bangor et al. (2008)	2324	0.91
Berkman and Karahoca (2016)	151	0.83
Finstad (2010)	558	0.97
Kortum and Sorber (2015)	3575	0.88
Lewis and Sauro (2009)	324	0.92
Lewis et al. (2013)	389	0.89
Lewis et al. (2015)	471	0.90
Lewis (2018)	618	0.93
Sauro and Lewis (2011)	107	0.92

Table 14: Comparison of the Cronbach's alpha reliability coefficients of different studies to validate the effectiveness of the proposed experimental condition

The system usability scale has been applied to real-time applications such as- Web sites, Cell phones, Television, Interactive Voice Responses (IVR), Graphical User Interface (GUI) and the mean SUS scores of these products were 68.2, 65.9, 67.8, 72.4, 76.2 respectively. The perceived usability of 14 everyday products has been measured using the SUS. Previously mentioned study has shown that Excel, Global Positioning System (GPS), Digital Video Recorder (DVR), PowerPoint presentation (PPT), Word, Wii (game console), iPhone, Amazon, Automated teller machine (ATM), Gmail, Microwaves, Landline, Browser, Google search have the mean SUS scores of 56.5, 70.8, 74.0, 74.6, 76.2, 76.9, 78.5, 81.8, 82.3, 83.5, 86.9, 87.7, 88.1, 93.4 respectively. Hence, the afore-mentioned study validates the acceptability of SUS. It has been obtained from the previous study that the SUS score does not interpret whether the usability of a product is poor or good. This decision requires some comparison with other products. One form of the comparison is to statistically compare the SUS score of two or more groups. The comparison can be done among diverse products or user groups [18]. The percentile values and descriptive statistics of adjective ratings for individual System

Usability Scale (SUS) scores have shown that the type 2 sanitizer container has better usability over type 1 and type 3. Because a higher SUS score means better usability [11]. According to the comparative chart of mean SUS score, type 2 sanitizer container has 'good' usability of 71.39 which is higher than type 1(63.50) and type 3(41.06) which are in the range of 'ok' and 'poor' respectively [11, 39]. Friedman's ANOVA has revealed that the result is not only true for the general users but also true for the user categories based on gender and work type. But post-hoc analysis indicated that the females and commuting workers have found that the usability of type 1 and 2 sanitizer containers were similar. All of these outcomes indicate that the type 2 sanitizer container is much easier to use, less complex, and has better learnability. On the other hand, there was no significant difference in perceived usability of a particular sanitizer container between the genders and between the type of work. Additional reasons for the greater SUS score of type 2 sanitizer are- more accurate dispense of sanitizer solution and anyone can comfortably use it in public or at home. Usability influences the choice of a product [8]. The higher usability of type 2 can result in more preference for pressure pumps like sanitizer container (type 2).

Conclusion

It can be concluded that the system usability scale is not the best measure of usability for sanitizer containers though it has a good reliability coefficient. One of the reasons is that the SUS has a poor convergent validity. Another reason is that it can be affected by CMB. Additionally, this research revealed that the type 2 sanitizer container has better usability, learnability, user-friendliness, and lower perplexity irrespective of the users. These factors helped users to give a higher SUS score to the type 2 sanitizer container when compared to the other two. The result of this study suggested that both the genders (male and female) and workers (commuting and working from home) widely accepted the type 2 sanitizer container for their day-to-day lives.

Limitations and Drawbacks

● Initially, the SUS was developed for HCI. It was applied to sanitizer containers that were non-HCI in the current study. For that reason, it has faced convergent validity issues and it has made the SUS a poor marker of usability for non-HCI.
● The present study has used a post-hoc unmeasured latent marker construct (ULMC) method to detect Common method bias (CMB). ULMC is not the best method to detect and minimize the effect of CMB.
● Only three types of sanitizers have been used as non-HCI to evaluate SUS as a marker of non-human computer interface's usability. The application of SUS on other products with non-HCI has needed to confirm the existing results.

Recommendations

Usability researchers should not rely on SUS only for non-human computer interfaces (non-HCI) like- sanitizer containers. They should use more than one usability scale or questionnaires precisely for non-human computer interface (non-HCI). If practitioners are unable to use more than one usability questionnaire, then they are recommended to use a-priori ideal markers along with the SUS questionnaire as an effective measure against common method bias [49]. Because it will statistically remove the effect of CMB from the SUS responses [50].

SUS with all positive items can be an effective measure of perceived usability. The previous study has shown that it has better reliability [51]. All positive SUS questionnaire was adapted to measure perceived usability for non-human computer interface (non-HCI) like face coverings[16].

Practitioners should separate the usability measurement of each product to reduce common method bias. This separation can be done through psychological barriers. For example, the psychological barrier can be created by using a story to make it appear that the usability of the first product is not related to the usability of the other product [30].

The randomized or counterbalance order of the presentation of the products along with a higher number of participants can minimize the effect of common method variance [52].

Acknowledgment

References

1. Guan W, Ni Z, Hu Y et al. (2020) Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 382:1708-1720.

2. W.H.O. (2020) WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020.

3. Haider N, Rothman-Ostrow P, Osman AY et al. (2020) COVID-19 Zoonosis or Emerging Infectious Disease? Front Public Heal 8:1-8.

4. Bawankar V, Sawarkar G (2020) Overview of sanitizer usability in COVID-19 pandemic scenario. Indian J Forensic Med Toxicol 14:6636-6640.

5. Terlep S (2021) Hand Sanitizer Sales Jumped 600% in 2020. Purell Maker Bets Against a Post-Pandemic Collapse. wall Str. J.

6. Jayan TV (2021) India's hand sanitizer production capacity grew 1,000 times during Covid-19 pandemic. Hindu Bus.

7. Drew MR, Falcone B, Baccus WL (2018) What Does the System Usability Scale (SUS) Measure? In: Marcus A, Wang W (eds) Design, User Experience, and Usability: Theory and Practice. Springer International Publishing, Cham, 356-366

8. Mack Z, Sharples S (2009) The importance of usability in product choice: A mobile phone case study. Ergonomics 52:1514-1528.

9. Chowdhury A, Karmakar S, Reddy SM, Ghosh S, Chakrabarti D (2014) Usability is more valuable predictor than product personality for product choice in human-product physical interaction. Int J Ind Ergon 44:697-705.

10. Brooke J (1995) SUS: A 'Quick and Dirty' Usability Scale. Usability Eval Ind 207-212.

11. Bangor A, Kortum PT, Miller JT (2008) An empirical evaluation of the system usability scale. Int J Hum Comput Interact 24:574-594.

12. Lewis JR, Sauro J (2009) The factor structure of the system usability scale. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 5619 LNCS:94-103.

13. Borsci S, Federici S, Lauriola M (2009) On the dimensionality of the System Usability Scale: A test of alternative measurement models. Cogn Process 10:193-197.

14. Lewis JR, Sauro J (2017) Revisiting the Factor Structure of the System Usability Scale. J Usability Stud 12:183-192

15. Mol M, Van Schaik A, Dozeman E et al. (2020) Dimensionality of the system usability scale among professionals using internet-based interventions for depression: A confirmatory factor analysis. BMC Psychiatry 20:1-10.

16. Robertson I, Kortum P (2021) The Usability of Face Coverings Used to Prevent the Spread of COVID-19. Hum Factors 1-16.

17. Kortum P., Bangor A (2013) Usability Ratings for Everyday Products Measured With the System Usability Scale. Int J Hum Comput Interact 29:67-76.

18. Lewis JR (2018) The System Usability Scale: Past, Present, and Future. Int J Hum Comput Interact 34:577-590.

19. Ng AWY, Lo HWC, Chan AHS (2011) Measuring the usability of safety signs: A use of System Usability Scale (SUS). IMECS 2011 - Int MultiConference Eng Comput Sci 2011 2:1296-1301

20. Peres SC, Pham T, Phillips R (2013) Validation of the system usability scale (SUS): Sus in the wild. Proc Hum Factors Ergon Soc 192-196.

21. Martins AI, Rosa AF, Queirós A, Silva A, Rocha NP (2015) European Portuguese Validation of the System Usability Scale (SUS). Procedia Comput Sci 67:293-300.

22. Grier RA, Bangor A, Kortum P, Peres SC (2013) The system usability scale: Beyond standard usability testing. Proc Hum Factors Ergon Soc 187-191.

23. Lewis JR, Sauro J (2018) Item Benchmarks for the System Usability Scale. J Usability Stud 13:158-167

24. Massey FJ (1951) The Kolmogorov-Smirnov Test for Goodness of Fit. J Am Stat Assoc 46:68-78.

25. Lilliefors HW (1967) On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. J Am Stat Assoc 62:399-402.

26. Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences, Second. LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS, New York

27. Cohen J (1992) A Power Primer. Psychol Bull 112:155-159.

28. Williams B, Brown Andrys Onsman T, Onsman A et al. (2010) Exploratory factor analysis: A five-step guide for novices. Australas J Paramed 8.

29. Ul Hadia N, Abdullah N, Sentosa I (2016) An Easy Approach to Exploratory Factor Analysis: Marketing Perspective. J Educ Soc Res.

30. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP (2003) Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. J Appl Psychol 88:879-903.

31. Brown TA (2006) Confirmatory factor analysis for applied research. The Guilford Press, New York

32. Hair JFJ, Black WC, Babin BJ, Anderson RE (2018) Maltivariate data analysis, 8th ed. Cengage Learning India Pvt. Ltd

33. Henseler J, Ringle CM, Sarstedt M (2015) A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 43:115-135.

34. Hayes AF, Coutts JJ (2020) Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…. Commun Methods Meas 14:1-24.

35. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297-334.

36. Richardson HA, Simmering MJ, Sturman MC (2009) A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. Organ Res Methods 12:762-800.

37. Serrano Archimi C, Reynaud E, Yasin HM, Bhatti ZA (2018) How Perceived Corporate Social Responsibility Affects Employee Cynicism: The Mediating Role of Organizational Trust. J Bus Ethics 151:907-921.

38. Field A (2017) Discovering Statistics using IBM SPSS Statistics, 5th ed. SAGE Publications Ltd, London

39. Bangor A, Staff T, Kortum P, Miller J, Staff T (2009) Determining what individual SUS scores mean: adding an adjective rating scale. J usability Stud 4:114-123

40. Rosenthal JA (1996) Qualitative descriptors of strength of association and effect size. J Soc Serv Res 21:37-59.

41. Kaiser HF (1974) An index of factorial simplicity. Psychometrika 39:31-36.

42. Jackson DA (1993) Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecol Soc Am 74:2204-2214.

43. Rojas-Valverde D, Pino-Ortega J, Gómez-Carmona CD, Rico-González M (2020) A systematic review of methods and criteria standard proposal for the use of principal component analysis in team's sports science. Int J Environ Res Public Health 17:1-13.

44. Ford JK, MacCallum RC, Tait M (1986) the Application of Exploratory Factor Analysis in Applied Psychology: a Critical Review and Analysis. Pers Psychol 39:291-314.

45. MacCallum RC, Widaman KF, Zhang S, Hong S (1999) Sample size in factor analysis. Psychol Methods 4:84-99.

46. Pearson RH, Mundfrom DJ (2010) Recommended sample size for conducting exploratory factor analysis on dichotomous data. J Mod Appl Stat Methods 9:359-368.

47. Goodwin NC (1987) Functionality and usability. Commun ACM 30:229-233.

48. Malhotra NK (2010) Marketing Research- An Applied Orientation, 6th ed. Pearson Education, Inc., publishing as Prentice Hall

49. Simmering MJ, Fuller CM, Richardson HA, Ocal Y, Atinc GM (2015) Marker Variable Choice, Reporting, and Interpretation in the Detection of Common Method Variance: A Review and Demonstration. Organ Res Methods 18:473-511.

50. Williams LJ, Hartman N, Cavazotte F (2010) Method variance and marker variables: A review and comprehensive cfa marker technique. Organ Res Methods 13:477-514.

51. Kortum P, Acemyan CZ, Oswald FL (2021) Is It Time to Go Positive? Assessing the Positively Worded System Usability Scale (SUS). Hum Factors 63:987-998.

52. Tehseen S, Ramayah T, Sajilan S (2017) Testing and Controlling for Common Method Variance: A Review of Available Methods. J Manag Sci 4:146-175.