Uşak Üniversitesi Sosyal Bilimler Dergisi
Yıl: XI Sayı:Eylül-2018 Sayfa:237-253
Geliş:11.06.2018
Kabul:14.09.2018

Usak University Journal Of Social Sciences
Year: XI /Issue:September-2018 Pages: 237-253
Received:11.06.2018
Accepted:14.09.2018

DOİ: http://dx.doi.org/10.29217/uujss.495
Bu makale benzerlik taramasına tabi tutulmuştur.

A TURKISH TRANSLATION OF THE SYSTEM USABILITY SCALE: THE SUS-TR[1]

Denizhan DEMİRKOL*
Çağla ŞENELER**

Abstract

Aim of developing usable and effective systems yield need for measuring quality of user experience with a system. The System Usability Scale (SUS) questionnaire that was developed in English has been widely used in the literature for assessing perceived usability of interactive systems. This study aimed to adapt the SUS into Turkish and investigate its validity and reliability. To verify the translation, two professional translation techniques were used and four translators were employed. The official Turkish version of the SUS (SUS-TR) conducted to 324 university students. The SUS-TR's reliability was studied and it was found at a high level. Furthermore, confirmatory factor analyses (CFA) successfully revealed the two-factor structure of the SUS-TR. The results of this study showed that the SUS-TR is a reliable and valid tool for measuring usability, with psychometric properties consistent with the original version. The SUS-TR was developed successfully to make it suitable for users from Turkey.

Keywords: System Usability Scale, Usability, Turkish, Reliability, Validity

SİSTEM KULLANILABİLİRLİK ÖLÇEĞİNİN TÜRKÇEYE ÇEVİRİSİ: SUS-TR

Özet

Kullanılabilir ve verimli sistemlerin geliştirmesi amacı, kullanıcının sistem deneyiminin kalitesinin ölçülme ihtiyacını açığa çıkarmaktadır. İngilizce olarak geliştirilen Sistem Kullanılabilirlik Ölçeği (SUS), interaktif sistemlerin algılanan kullanılabilirliğini değerlendirmek için literatürde yaygın olarak kullanılmıştır. Bu çalışma SUS'un Türkçeye uyarlanmasını, geçerliliğini ve güvenilirliğini araştırmayı amaçlamıştır. Çeviriyi doğrulamak için, iki profesyonel çeviri tekniği ve dört çevirmen kullanılmıştır. SUS-TR, 324 üniversite öğrencisine uygulanmıştır. SUS'un Türkçe versiyonu olan SUS-TR'ın güvenilirliği çalışılmış ve yüksek düzeyde bulunmuştur. Ayrıca, doğrulayıcı faktör analizi, SUS-TR'nin iki faktörlü yapısını başarılı bir şekilde ortaya koymuştur. Bu çalışmanın sonuçları, SUS-TR'nin kullanılabilirliği ölçmek için güvenilir ve geçerli bir araç olduğunu

---

[1] This work was produced from the thesis entitled "Evaluation of a Student Information System (SIS) in Employment, Performance and Perceived Usability: A Turkish University Case", which has been defended at the 07.06.2018 master thesis under Dr. Cagla Seneler academic consultancy.

* Management Information Systems Department, Adnan Menderes University, Aydın, Turkey, denizhan.demirkol@adu.edu.tr

** Management Information Systems Department, Yeditepe University, Istanbul, Turkey,cagla.seneler@yeditepe.edu.tr

ve orijinal İngilizce versiyonuyla tutarlı psikometrik özellikleri olduğunu göstermiştir. Sonuç olarak, SUS-TR başarılı bir şekilde geliştirilmiş ve Türk kullanıcılara uygun hale getirilmiştir.

Anahtar Kelimeler: Sistem Kullanılabilirlik Ölçeği, Türkçe, Geçerlilik, Güvenilirlik

1. Introduction

The International Organization for Standardization (ISO) standard 9241-11 Guidance on Usability (1998) defines usability as;

> *the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (ISO,1998).*

Furthermore, Jordan (1998, p: 5) listed three components of usability; effectiveness as the extent to which a goal or task is achieved, efficiency as the amount of effort required to accomplish a goal and satisfaction as the level of comfort that the users feel when using a product and how acceptable the product is to users as a means of achieving their goals.

The increase of developing and using complex technologies causes need for measuring the quality of users' experiences when they are interacting with a system. Therefore, researches on user experience and usability still get attention to gain a better understanding of these issues. There are lots of studies that showed usable systems' positive effects on users' experiences (Forlizzi and Battarbee, 2004; Kampf and Payne, 2000; Thüring and Mahlke, 2007). Nielsen (2012) stated that if a website is not easy to use, people leave. In today's world, people can easily switch to another system since there are lots of alternatives. Thus, it is highly critical to create and develop usable and effective systems. In addition, usability also can affect users' preferences on systems. According to Jordan (1998), if a system is not usable this can cause problems of a varying degree of severity. Besides, usability has financial implications for the commercial and industrial world (Mack and Sharples, 2009). Mack and Sharples (2009) studied the importance of usability on users' system preference related to system's attributes. They found that usability is important on system preference. Thus, usability evaluation and its measurement are extremely important both during the process of system development and after the system was released.

Hornbeak (2006) states that usability cannot be directly measured and selecting suitable usability measure is challenging. However, it is known that it is a necessity to find an appropriate and practical method for evaluating usability. For that reason, various methods and approaches have been defined in the literature. According to Sweeney, Maguire and Shackel (1993), to understand how well users learn and use a system, there are three different usability evaluation approaches that can be classified as expert-based, model-based, and user-based.

In expert-based (also called non-empirical) methods, an expert is making a judgement about system's usability or s/he is checking system's design qualities in a structured way. However, these methods have considerable disadvantages. To illustrate, it is hard to find trained usability experts. Even if you can find one, these experts are demanding high prices. Furthermore, multiple experts are needed to evaluate the system and aggregate the results (Nielsen, 1992). Second approach that is the model-based approach can be defined as using a model of how user would use planned system to acquire supposed system (Sears and Jacko, 2008). However, it has some disadvantages like measuring only some aspects of user performance and inadequate task applicability (Dillon, 2001).

Since systems are developed for users, usability mainly focuses on understanding users. That's why, if developers desire to develop a useful system, the understanding of potential users is vital. Hence, the last approach that includes user-based (also empirical) methods focuses on observing how users are interacting with systems or allows users to express their perceptions about systems' usability. According to several studies, most apparent way to learn about usability of tested system is asking to the participants about their experiences with it (Tullis and Albert, 2013) and this can be fulfilled in a structured way by using a questionnaire (Macleod, 1994; Tullis and Albert, 2013). Questionnaires support researchers to understand potential users clearly (Rubin and Chisnell, 2008). In a broad way, they are appropriate methods for interpreting how users use systems and what features they specifically like or dislike (Nielsen, 1993). For these reasons, questionnaires are greatly practical methods for assessing usability.

A variety of questionnaires have been used and reported in the literature for assessing perceived usability of interactive systems such as System Usability Scale (SUS) (Brooke, 1996), Questionnaire for User Interface Satisfaction (QUIS) (Harper and Norman, 1998) and Computer System Usability Questionnaire (CSUQ) (Lewis, 1995). The SUS questionnaire has been widely used in the literature (Barnum, 2011; Brooke, 2013; Finstad, 2006; Tullis and Albert, 2013, Tullis and Stetson, 2004). The SUS was created in the 1980s by Dr. John Brooke at Digital Equipment Co. Ltd. to give a universal view of subjective assessments of usability (Brooke, 1996). It is simple and practical scale as a result is widely accepted and used for assessing usability of diverse types of systems including websites, software products and hardware.

The SUS is a 10-item questionnaire (See Table A.1 in Appendix for original version of the SUS) in which respondents indicate their level of agreement with each item on a scale from 1 (strongly disagree) to 5 (strongly agree). The odd numbered items (1,3,5,7 and 9) have positive-tone words and the even-numbered items (2,4,6,8 and 10) have negative-tone words. For odd numbered items, the score contribution is the scale position minus 1 and for even-numbered items, the score contribution is 5 minus the scale position. Next, to obtain the overall value of SUS, the sum of the contributions should be multiplied by 2.5. The conversion of SUS scores to a scale that can range from 0 to 100 makes it easier for usability practitioners and product managers to communicate (Brooke, 2013). Brooke (1996) stated that higher scores indicate better system usability.

Nunnally (1978) stated that standardized usability questionnaires offer many advantages, which are objectivity, replicability, quantification, economy, communication and scientific generalization to practitioners. The SUS is potent, notorious and its use is well-established with more than 1200 publications (Brooke, 2013). There are few direct comparisons of the various standardized usability questionnaires (Sauro and Lewis, 2012). Moreover, its comparison with other usability questionnaires have been performed (Tullis and Stetson; 2004). Tullis and Stetson (2004) conducted a comparison of questionnaires for assessing website usability study. Five questionnaires (SUS, QUIS, CSUQ, a two variant of Microsoft's Product Reaction Cards) were compared with 123 participants. The SUS yielded among the most reliable results across sample sizes. Besides, the SUS was the only questionnaire that has questions all address different aspects of the user's reaction to the website as a whole.

The SUS questionnaire was developed in English. There are studies that showed participants respond better to scales in their own language (Bahrick, Hall, Goggin, Bahrick, and Berger, 1994; Delgado, Guerrero, Goggin and Ellis, 1999). The SUS is previously translated European Portuguese, Persian and Slovene languages (Blažica and Lewis, 2015; Dianat, Ghanbari, and AsghariJafarabadi, 2014; Martins, Rosa, Queirós, Silva, and Rocha, 2015). There are also various unofficial translations of the SUS. It's reliability and validity studies are also carried out. Bangor, Kortum and Miller (2008) described the results of 2324 SUS surveys from 206 usability tests collected over a 10-year period. In that study, it was found that the SUS was highly reliable (Cronbach's alpha = 0.91) and useful over a wide range of interface types. A standardized questionnaire has construct validity when a factor analysis demonstrates that its items align as expected with its hypothesized factors. A number of researchers have noted the tendency for positive-tone and negative-tone items of the SUS to load on separate factors (Barnette, 2000; Davis, 1989; Pilotte and Gable, 1990; Sauro and Lewis, 2011; Schmitt and Stuits, 1985; Schriesheim and Hill, 1981; Wong, Rindfleisch, and Burroughs, 2003).

Although a number of translations into other languages exist as mentioned at above, the authors of this paper have failed to find a Turkish version of the SUS (a fact confirmed by SUS's main author, Dr. John Brooke). Therefore, to make an official Turkish version of the SUS (SUS-TR) suitable for users from Turkey, it was translated into the Turkish language with permission of Dr. John Brokee by using professional translation techniques. Furthermore, the SUS-TR was studied for its reliability and validity.

2. Method

2.1. Translation Method Of The SUS Into Turkish

With permission from its main author, Dr. John Brooke, the SUS was translated into Turkish. In order to validate translations and to reduce the risks that can be faced while translating from one language to another, two different translation techniques were used. Most studies have translated questionnaires into other languages have applied one of the each translation techniques during the translation process (Isemonger and Sheppard, 2007). In this study, the use of both multiple forward and back-translation techniques prevented poor translations and enabled translations to be crosschecked. This translation approach was used before successfully in the literature (Seneler, 2014). In order to translate the SUS into Turkish, four translators who are native speakers of Turkish and advanced speakers of English were employed. These four translators will be referred to as Translator1, Translator2, Translator3 and Translator4 in this text.

In the first phase, a multiple forward translation technique was used. A multiple forward translation technique is the translation of a document from the source language into the target language independently by a number of translators (Maxwell, 1996). Translator1 and Translator2 undertook two independent translations. Then the authors as a native speaker of Turkish and fluent English speaker compared these translations on an item-to-item basis in order to identify any differences in meaning. Next, Translator3 was asked to translate only the different parts of the first two translations. After that, the efforts of all three translators were evaluated and these efforts produced an overall first translation (See Figure 1).

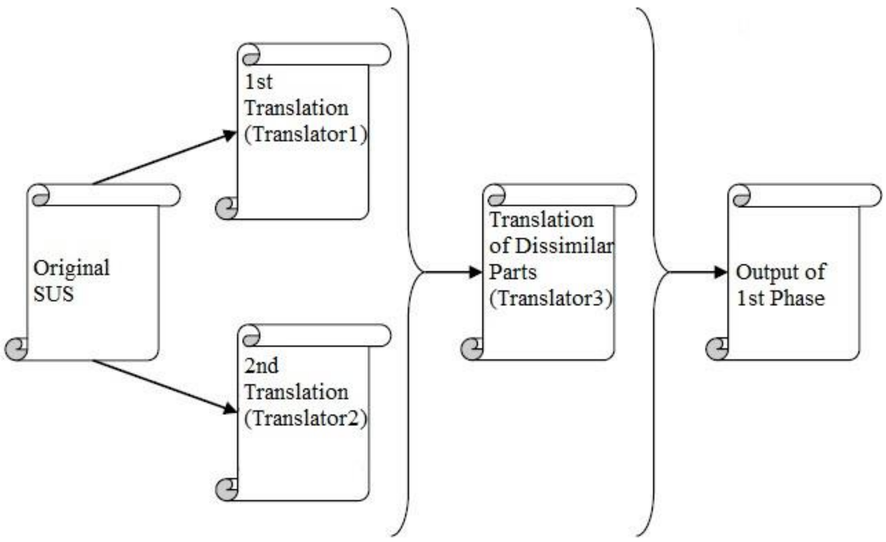

Figure 1: First phase of the translation process

In the second phase, a back-translation technique was used, that is a translation of a document that has been already translated into a target language back into the source language (Maxwell, 1996). Translator4 was asked to translate the output of first phase (the overall first translation of the SUS) back into English (See Figure 2).

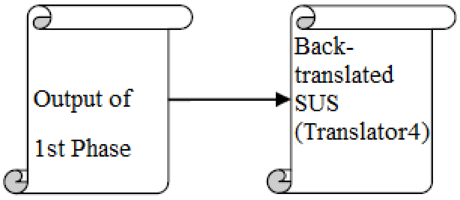

Figure 2: Second phase of the translation process

In the third and last phase of translation process, the original SUS and the back-translated SUS were compared (See Figure 3). Appropriate modifications were made and the Turkish version of SUS was finalized (See Table A.2 for finalized SUS-TR).

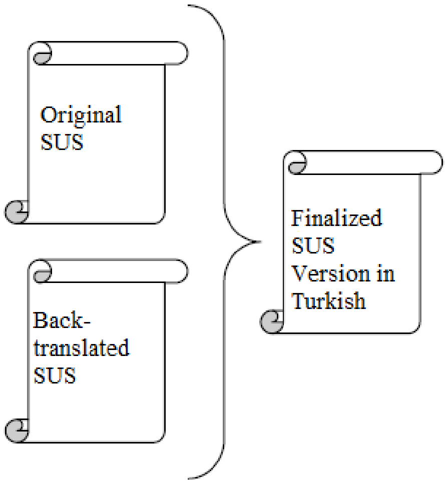

Figure 3: Third phase of the translation process

2.2. Reliability and Validity Study of the SUS-TR

2.2.1. Method

After the translation process has ended, the scale was conducted to study reliability and validation of it. Before putting the scale into practice, a pilot study was conducted with six undergraduate students from a Turkish University to determine unrecognized and unclear points and to get any feedback. After that, the data were collected by a two-part questionnaire. In the first part of the questionnaire, participants were asked to give answers on the SUS-TR. The questions in this section are of the Likert type of 5 (1 = "Strongly disagree" and 5 = "Strongly agree"). In the second part of the questionnaire, demographic information such as age, gender, department and questions that determine the duration of web usage were included.

In the data collection process, an electronic version of the questionnaire was prepared so that it was applied online. The address of the electronic questionnaire is sent by e-mail to the students and it was also implemented online in several lectures. The data collection process lasted four weeks in total.

2.2.2. Participants

The scale was applied to 324 undergraduate students from a Turkish University. As it is illustrated in the following table (See Table 1), 50% of the participants are female and 50% of the participants are male. Almost all participants' ages are between 18-25 years.

Table 1: Demographic information of the working group

		f	%
Gender	Female	162	50
	Male	162	50
Age	18	5	1.5
	19	31	9.5
	20	61	18.8
	21	62	19.1
	22	76	23.4
	23	38	11.7
	24	23	7.0
	25	15	4.6
	26	5	1.5
	27	4	1.2
	28	3	0.9
	32	1	0.3
Total		324	100

The average age of the participants is 20. The participants mentioned that 92,9% of them use the web every day. A large majority of participants (70.4%) use computers over 10 years. Almost half of the participants (55.2%) stated that their confidence about computer use could change depending on the given task. Furthermore, 40.1% of the participants expressed that they are confident about using computers. Only one participant mentioned that s/he does not feel confidence and trust about her/his computer usage at all (See Table 2).

Table 2: Web usage, computer usage and confidence statistics of the participants

	Items	f	%
Web usage	Everyday	301	92.9
	Several times a week	22	6.7
	Several times a month	1	0.3
Total		324	100
Computer usage	3-5 years	6	1.8
	5-7 years	22	6.7
	7-10 years	68	20.9
	10 years and over	228	70.4
Total		324	100
Computer usage confidence	Do not trust	1	0.3
	Usually need help	14	4.3
	Depends on the given task	179	55.2
	Trust	130	40.1
Total		324	100

2.2.3. Item analysis

Item analysis uses statistics and expert judgment to evaluate tests based on the quality of individual items, item sets, and entire sets of items, as well as the relationship of each item to other items. It explores the performance of items considered separately either in relation to some external criterion or in relation to the remaining items on the test (Thompson and

Levitov, 1985). The items should be extracted from scale if the relation of one item to the other items is below 0.30. It was decided not to extract any item from the scale since none of the relation of one item to the other items is below this value (Büyüköztürk, Çakmak, Akgün, Karadeniz and Demirel, 2017). Item-total statistics of the SUS-TR's scale were given in following table (See Table 3).

Table 3: Item-total statistics of the SUS-TR

ITEMS	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
*Item 4	29.417	59.160	0.397	0.841
Item 3	28.247	55.382	0.561	0.826
Item 7	28.164	54.986	0.587	0.823
*Item 10	29.056	56.926	0.515	0.830
*Item 6	28.321	54.417	0.665	0.816
*Item 8	28.531	56.423	0.583	0.824
*Item 2	28.500	53.594	0.653	0.816
Item 5	27.546	59.394	0.548	0.829
Item 9	27.907	58.555	0.516	0.830
Item1	27.617	60.807	0.371	0.842
*Rotated				

2.2.4. Reliability analysis

Reliability was assessed using coefficient alpha (Cronbach, 1951) and it is generally accepted method for measuring reliability (Sauro and Lewis, 2012). It was determined that the reliability level of the SUS-TR scale consisting of 10 items was at a high level which is 0.8 (See Table 4).

Table 4: Reliability coefficients

Cronbach's Alpha	N of Items	N of Cases
0.84	10	324

2.2.5. Assessment of appropriateness of data for factor analysis

The Kaiser-Meyer-Olkin (KMO) and Barlett's test can be used to determine whether the data obtained from the study group is consistent with the explanatory factor analysis (ECA) (Büyüköztürk, 2010; Çokluk, Şekercioğlu and Büyüköztürk, 2012; Karagöz and Kösterelioğlu, 2008). As a result of the KMO test, it is interpreted that factor analysis cannot be continued if the value is lower than 0.5 (Çokluk et al., 2012). The KMO value for this study was 0.8 that is quite adequate for research sample. Showing the suitability of the data for factor analysis, the Bartlett test result was also significant ($\chi2 = 1072.8$, $p = 0.000$) (See Table 5).

Table 5: KMO and Bartlett test results

Kaiser-Meyer-Olkin Test (KMO)		0.871
Bartlett's Test	Chi-Square	1072.869
	df	45
	p.	0.000

2.2.6. Construct validity

For evaluating construct validity, the results of both EFA and CFA were presented for the items of the SUS-TR. Based on the findings, the principal component method and varimax rotation were applied as EFA of the 10-item SUS-TR scale. A two-factor structure emerged that explains 56.7% of the total variance resulting from factor analysis with an eigenvalue greater than 1. The first factor and the second factor revealed 29.5% and 27.3% of the total variance, respectively (See Table 6).

Table 6: Total variance explained

Components	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	**% Of the variance**	**Cumulative %**	**Total**	**% Of the variance**	**Cumulative %**	**Total**	**% Of the variance**	**Cumulative %**
1	4.210	42.096	42.096	4.210	42.096	42.096	2.936	29.365	29.365
2	1.460	14.602	56.698	1.460	14.602	56.698	2.733	27.333	56.698
3	0.809	8.095	64.793						
4	0.674	6.738	71.531						
5	0.582	5.820	77.352						
6	0.541	5.411	82.763						
7	0.509	5.088	87.851						
8	0.447	4.475	92.326						
9	0.443	4.427	96.753						
10	0.325	3.247	100.000						

When the scree plot test graph is examined (See Figure 4), it is seen that limiting the factor number to two is sufficient. When the distribution of the items according to the factors after the varimax rotation method is examined, it is seen that all the materials provide logical integrity in terms of the factor structures.

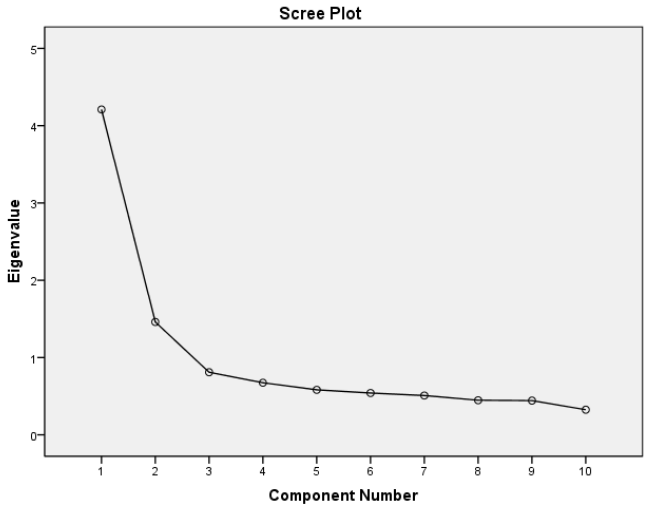

Figure 4: Scree plot test graph

According to the factor loadings which are shown in the following table (See Table 7), since the distances between the loads of the factors in which the items are collected must be at least 10% of the distance between them, there is no item which does not comply with this rule. The Cronbach Alpha value was used to calculate the internal consistency of the factors. If the Cronbach alpha value is over 0.7 , it means that the reliability is high (Cronbach, 1951). As shown in the following table, all items have Cronbach values greater than 0.7.

Table 7: Varimax-rotated two-factor solution for the Turkish version of the system usability scale

Items	**Factors**	
	1	**2**
Item 5	0.775	
Item 1	0.746	
Item 7	0.725	
Item 3	0.609	
Item 9	0.605	
Item 10		0.819
Item 4		0.753
Item 8		0.684
Item 6		0.640
Item 2		0.600

When the items belonging to the factors are examined, five expressions were collected under factor 1, and these expressions, load values and other statistical values are shown in the following table (See Table 8). This factor can be named as "Positive SUS-TR Factor". Factor loads of the positive SUS-TR factor ranged from 0.6 to 0.7. Positive SUS-TR factor has a high reliability, which is 0.7.

Table 8: Positive SUS-TR factor

Factor 1: Positive SUS-TR Factor	**Factor Load**	**Factor Reliability**
Item 5	0.775	0.769
Item 1	0.746	
Item 7	0.725	
Item 3	0.609	
Item 9	0.605	

Five expressions were collected under factor 2 (See Table 9), and these expressions and load values and other statistical values are shown in the following table. When the collected items of this factor are examined, the factor can be named as "Negative SUS-TR Factor". Factor loads of the negative SUS-TR factor ranged from 0.6 to 0.8. Negative SUS-TR factor has a high reliability, which is 0.8.

Table 9: Negative SUS-TR factor

Factor 2: Negative SUS-TR Factor	**Factor Load**	**Factor Reliability**
Item 10	0.819	0.805
Item 4	0.753	
Item 8	0.684	
Item 6	0.640	
Item 2	0.600	

2.2.7. CFA results

The accuracy of the two-dimensional factorial structure based on the results of EFA was tested by CFA. Since there is no single statistical significance test used to assess the fitness of the model generated using the obtained data, the fact that many measurements are considered simultaneously in the process of evaluating the model in the study has been taken into consideration. In the CFA, different indices were used to assess the fitness of a model, and the most commonly used ones were; "The Root Mean Square Error of Approximation (RMSEA)" and "Comparative Fit Index (CFI)" (Tabachnick and Fidell, 2007). In this study, the values of chi-square ($\chi 2$), RMSEA, CFI, "Goodness of Fit Index", "Adjusted Good Fit Index (AGFI)" and "Normed Fit Index (NFI)" were evaluated on the basis of.

When performing the CFA analysis of SUS-TR, the subscales of the scale were called as positive dimension and negative dimension. Necessary modifications have been carried out and the SUS-TR scale has been modified from SUS9 to SUS7 and from SUS10 to SUS4. As shown in the following figure (See Figure 5), SUS7 is the most important item with a coefficient of 0.7 in the positive factor whereas SUS6 is the most important item with a coefficient of 0.8 in the negative factor.

In addition, the relationship between positive and negative dimensions is 0.7 and the relationship between them is significant (p = 0.000). The results of the CFA analysis in which the subscales of the SUS-TR are included are given in the following table (See Table 10).

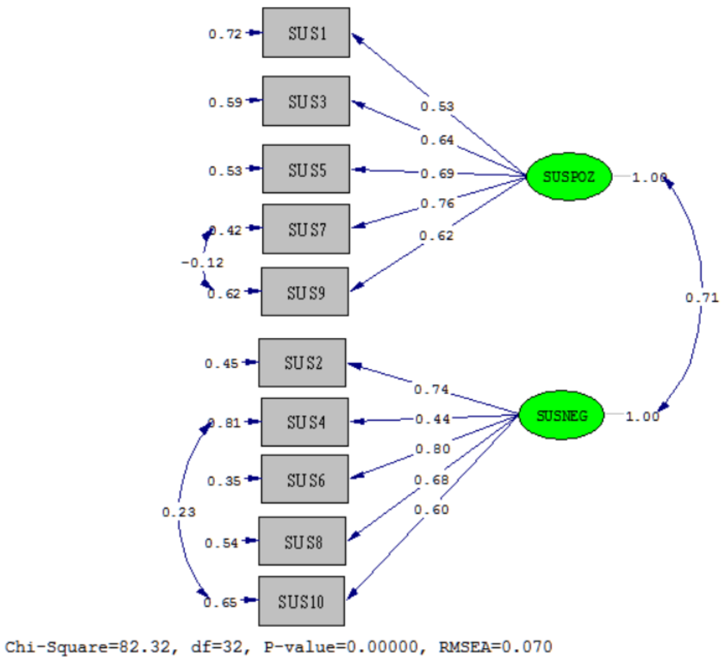

Figure 5: CFA results for SUS-TR scale

Table 10: SUS-TR scale model results

Factor/Item	**Standardized Loads**	**t-value**	**R^2**
Positive dimension			
SUS1	0.53	9.35	0.28
SUS3	0.64	11.85	0.41
SUS5	0.69	12.89	0.47
SUS7	0.76	14.32	0.58
SUS9	0.62	10.75	0.38
Negative dimension			
SUS2	0.74	14.40	0.55
SUS4	0.44	7.62	0.19
SUS6	0.80	15.95	0.65
SUS8	0.68	12.73	0.46
SUS10	0.60	10.84	0.35

It is possible to say that the compliance criteria for the CFA analysis of SUS-TR are among the acceptable limits. Other than these criteria, $\chi 2$ (32) = 82.32, $\chi 2$ / sd = 2.57 < 3 is another indicator used to determine model suitability and is another indicator that the model is perfectly in terms of statistics (See Table 11).

Table 11. Values of compliance criteria for the SUS-TR scale CFA model

Compliace Criteria	**Good Fit**	**Acceptable Fit**	**Values of Developed Scale (Current Study)**
RMSEA	$0 < RMSEA < 0.05$	$0.05 \leq RMSEA \leq 0.10$	0.070
SRMR	$0 \leq SRMR < 0.05$	$0.05 \leq SRMR \leq 0.10$	0.052
GFI	$0.95 \leq GFI \leq 1$	$0.90 \leq GFI \leq 0.95$	0.950

AGFI	0.90 ≤ AGFI ≤ 1	0.85 ≤ AGFI ≤ 0.90	0.920
RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual, GFI: Goodness of Fit Index, AGFI: Adjusted Goodness of Fit Index (Schermelleh-Engel, Moosbrugger and Müller, 2003)			

Table 11 shows that the RMSEA value of the developed scale is acceptable (RMSEA = 0.070 ; 0.00 ≤ RMSEA ≤ 0.05). Furthermore, the SRMR value is within the acceptable fit index (SMRM = 0.052 ; 0.00 ≤ SRMR ≤ 0.05) (GFI = 0.950 ; 0.95 ≤ GFI ≤ 1.00). Besides, the AGFI value was within the perfect fit index (AGFI = 0.920 ; 0.90 ≤ AGFI ≤ 1.00). As a result of the CFA analysis, the items were confirmed to have related factors at 95% confidence level ($p < 0.05$) and that the compliance indices were within acceptable values and the model consistency was within acceptable values ($\chi 2$ / dF = 2.57 < 3).

3. Findings

To make an official Turkish version of the SUS suitable for users from Turkey, it was translated into the Turkish language with permission of Dr. John Brokee by using professional translation techniques. In this study, the use of both multiple forward and back-translation techniques prevented poor translations and enabled translations to be crosschecked. In order to translate the SUS into Turkish, four translators who are native speakers of Turkish and advanced speakers of English were employed. With this effort, a definite translation has been performed and the SUS-TR was structured.

Furthermore, the scale was applied to 324 undergraduate students from a Turkish University. It was evaluated for its reliability and validity, and was found to have good reliability and validity. The reliability level of the SUS-TR scale consisting of 10 items was at a high level, which is 0.8. The KMO value for this study was 0.8 that is quite adequate for research sample. Showing the suitability of the data for factor analysis, the Bartlett test result was also significant ($\chi 2$ = 1072.8 , p = 0.000).

For evaluating construct validity, the results of both EFA and CFA were presented for the items of the SUS-TR. Based on the findings, the principal component method and varimax rotation were applied as EFA of the 10-item SUS-TR scale. A two-factor structure emerged that explains 56.7% of the total variance resulting from factor analysis with an eigenvalue greater than 1. The first factor revealed 29.5% of the total variance and the second factor revealed 27.3% of total variance. When the distribution of the items according to the factors after the varimax rotation method is examined, it is seen that all the materials provide logical integrity in terms of the factor structures. When the items belonging to the factors are examined, two factors were emerged. This finding is inline with studies in literature (Barnette, 2000; Davis, 1989; Pilotte and Gable, 1990; Sauro and Lewis, 2011; Schmitt and Stuits, 1985; Schriesheim and Hill, 1981; Wong et al., 2003). The compliance criteria for the CFA analysis of the SUS-TR are among the acceptable limits.

4. Discussion

In this study, the SUS (Brooke, 1996), which was developed to measure usability, was adapted to Turkish. The present study has resulted in the development and validation of the SUS-TR for usability studies in Turkey. The multi-stage translation process included the steps of initial translation, expert review, and back-translation. Psychometric evaluation of the SUS-TR indicated an acceptable level of reliability. SUS-TR scale appears to be a two-factor structure that is inline with the literature. It is possible to say that the compliance criteria for the CFA analysis of the SUS-TR are among the acceptable limits.

Usability studies, which are very important for both the designers of the products and consumers, have been very limited in Turkey. Therefore, Turkish communities have a great need for valid and reliable tools and instruments to measure users' perception of the usability of a wide range of products and services. The results of this study showed that the SUS-TR is a valid and reliable tool for measuring usability, with psychometric properties consistent with the original English version.

As a conclusion, now the SUS is usable with Turkish users. Turkish researchers who wish to undertake research with Turkish participants related to the SUS can use SUS-TR.

5. References

Bahrick, H. P., Hall, L. K., Goggin, J. P., Bahrick, L. E., & Berger, S. A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. Journal of Experimental Psychology: General, 123, 264–283.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation Of The System Usability Scale. International Journal Of Human-Computer Interaction, 24(6), 574–594. Https://Doi.Org/10.1080/10447310802205776

Blažica, B., & Lewis, J. R. (2015). A Slovene Translation Of The System Usability Scale: The SUS-SI. International Journal Of Human-Computer Interaction, 31(2), 112–117. Https://Doi.Org/10.1080/10447318.2014.986634

Brooke, J. (1996). SUS - A Quick And Dirty Usability Scale. Usability Evaluation İn Industry, 189(194), 4–7. Https://Doi.Org/10.1002/Hbm.20701

Brooke, J. (2013). SUS : A Retrospective. Journal Of Usability Studies, 8(2), 29–40. Https://Doi.Org/10.1074/Jbc.R115.675280

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2017). Bilimsel Araştırma Yöntemleri. Pegem Atıf İndeksi, 1-360.

Büyüköztürk, Ş. (2010). Sosyal Bilimler İçin Veri Analizi El Kitabı, Pegem Akademi Yayıncılık, Ankara.

Cronbach, L. J. (1951). Coefficient Alpha And The İnternal Structure Of Tests* Lf~ J. Cronbach, *16*(3), 297–298.

Çokluk, Ö., Şekercioğlu, G. Ve Büyüköztürk, Ş. (2012). Sosyal Bilimler İçin Çok Değişkenli İstatistik: SPSS Ve Lisrel Uygulamaları,Pegem Akademi Yayıncılık, Ankara.

Davis, D. (1989). Perceived Usefulness, Perceived Ease Of Use, And User Acceptance Of Information Technology. MIS Quarterly, 13, 319–339.

Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B.(1999). Self-Assessment Of Linguistic Skills By Bilingual Hispanics. Hispanic Journal Of Behavioral Sciences, 21,31–46.

Dillon, A. (2001) Usability Evaluation. In W. Karwowski (Ed.) Encyclopedia Of Human Factors And Ergonomics, London: Taylor And Francis.

Dianat, I., Ghanbari, Z., & Asgharijafarabadi, M. (2014). Psychometric Properties Of The Persian Language Version Of The System Usability Scale. Health Promotion Perspectives, 4(1), 82–9. Https://Doi.Org/10.5681/Hpp.2014.011

Forlizzi, J., & Battarbee, K. (2004). Understanding Experience İn Interactive Systems. Human-Computer Interaction Institute, 37(2), 261–268. Https://Doi.Org/10.1145/1013115.1013152

Harper, B. D., & Norman, K. L. (1998). Improving User Satisfaction: The Questionnaire For User İnteraction Satisfaction Version 5.5. Proceedings Of Mid Atlantic Human Factors Conference, 224–228.

Hornbæk, K. (2006). Current Practice İn Measuring Usability: Challenges To Usability Studies And Research. International Journal Of Human Computer Studies, 64(2), 79–102. Https://Doi.Org/10.1016/J.İjhcs.2005.06.002

Isemonger, I. & Sheppard, C. (2007). A Construct-Related Validity Study On A Korean Version Of The Perceptual Learning Styles Preference Questionnaire. Educational And Psychological Measurement, 67(2), 357-368

ISO, 1998. Ergonomic Requirements For Office Work With Visual Display Terminals (Vdts)-Part 11: Guidance On Usability—Part 11: Guidance On Usability (ISO 9241-11:1998).

Jordan, P. W. (1998). An Introduction To Usability. CRC Press.

Kampf, M., & Payne, S. (2000). What İs Beautiful İs Usable. Interacting With Computers, 13(December), 127–145.

Karagöz, Y. Ve Kösterelioğlu, İ. (2008). "İletişim Becerileri Değerlendirme Ölçeğinin Faktör Analizi Metodu İle Geliştirilmesi" Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 21, 81-98.

Lewis, J. R. (1995). IBM Computer Usability Satisfaction Questionnaires : Psychometric Evaluation And İnstructions For Use IBM Computer Usability Satisfaction Questionnaires : Psychometric Evaluation And İnstructions For Use. International Journal Of Human- Computer Interaction, 7(1), 57–78. Https://Doi.Org/10.1037/T32698-000

M.Sweeney, M.Maguire, B. S. (1993). Evaluating User-Computer İnteraction: A Framework. International Journal Of Man-Machine Studies, 38(4).

Mack, Z., & Sharples, S. (2009). The Importance Of Usability İn Product Choice: A Mobile Phone Case Study. Ergonomics, 52(12), 1514–1528. Https://Doi.Org/10.1080/00140130903197446

Macleod, M (1994) Usability İn Context: Improving Quality Of Use. In: G Bradley And HW Hendricks (Eds.) Human Factors İn Organizational Design And Management - IV - Proceedings Of The International Ergonomics Association 4th International Symposium On Human Factors İn Organizational Design And Management, (Stockholm, Sweden, May 29 - June 1 1994). Amsterdam, Elsevier.

Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., & Rocha, N. P. (2015). European Portuguese Validation Of The System Usability Scale (SUS). Procedia Computer Science, 67(Dsai), 293–300. Https://Doi.Org/10.1016/J.Procs.2015.09.273

Maxwell, B. (1996). Translation And Cultural Adaptation Of The Survey Instruments. In: Martin, M. O. & Kelly, D. L. (Eds.). Third International Mathematics And Science Study (TIMSS) Technical Report, Volume I: Design And Development. Chestnut Hill, MA: Boston College.

Nielsen,J (04.01.2018). Usability 101: Introduction To Usability [Available Online At: Https://Www.Nngroup.Com/Articles/Usability-101-İntroduction-To-Usability/], Retrieved On February 22th, 2018.

Nielsen, J. (1992). Finding Usability Problems Through Heuristic Evaluation. Proceedings Of

The SIGCHI Conference On Human Factors İn Computing Systems - CHI ’92, 373–380. Https://Doi.Org/10.1145/142750.142834

Nielsen, J. (1993). Usability Engineering. Morgan Kaufmann Pietquin O And Beaufort R, 44(1/2002), 362. Https://Doi.Org/10.1145/1508044.1508050

Nunnally, J. C. (1978). Psychometric Theory (2nd Ed.). New York: Mcgraw-Hill.Barnette, J. J. (2000). Effects Of Stem And Likert Response Option Reversals On Survey İnternal Consistency: If You Feel The Need, There İs A Better Alternative To Using Those Negatively Worded Stems. Educational And Psychological Measurement, 60, 361–370.

Pilotte, W. J., & Gable, R. K. (1990). The Impact Of Positive And Negative İtem Stems On The Validity Of A Computer Anxiety Scale. Educational And Psychological Measurement, 50, 603–610.

Rubin, J., & Chisnell, D. (2008). Handbook Of Usability Testing [Electronic Resource] : How To Plan, Design, And Conduct Effective Tests (2nd Ed.). Indianapolis, IN: Wiley Pub. Https://Doi.Org/10.1007/S13398-014-0173-7.2

Sauro, J., & Lewis, J. R. (2011). When Designing Usability Questionnaires, Does İt Hurt To Be Positive? In Proceedings Of CHI 2011 (Pp. 2215–2223). Vancouver, Canada: ACM.

Sauro, J., & Lewis, J. R. (2012). Quantifying The User Experience. Practical Statistics For User Research. Morgan Kaufmann. Https://Doi.Org/10.1007/S13398-014-0173-7.

Seneler, C. (2014). The Impact Of Learning Styles And Cultural Background On Users' Experience Of Websites. Phd Thesis. University Of York.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating The Fit Of Structural Equation Models: Tests Of Significance And Descriptive Goodness-Of-Fit Measures. Methods Of Psychological Research Online, 8(2), 23–74. Https://Doi.Org/10.1002/0470010940

Schmitt, N., & Stuits, D. (1985) Factors Defined By Negatively Keyed İtems: The Result Of Careless Respondents? Applied Psychological Measurement, 9,367–373.

Schriesheim, C. A., & Hill, K. D. (1981). Controlling Acquiescence Response Bias By İtem Reversals: The Effect On Questionnaire Validity. Educational And Psychological Measurement, 41, 1101–1114.

Sears, A., & Jacko, J. (2008). The Human – Computer İnteraction Handbook. 2008 By Taylor & Francis Group, LLC.

Tabachnick, B. G. Ve Fidell, L. S. (2007). Using Multivariate Statistics. Boston: Pearson Education Inc.

Thüring, M., & Mahlke, S. (2007). Usability, Aesthetics And Emotions İn Human-Technology İnteraction. International Journal Of Psychology, 42(4), 253–264. Https://Doi.Org/10.1080/00207590701396674

Thompson, B., & Levitov, J. E. (1985). Using Microcomputers To Score And Evaluate Test Items. Collegiate Microcomputer, 3, 163-168.

Tullis, T., & Albert, B. (2013). Measuring The User Experience: Collecting, Analysing, And Presenting Usability Matrics. Measuring The User Experience (Second Edition). Https://Doi.Org/Http://Dx.Doi.Org/10.1016/B978-0-12-415781-1.00007-8

Tullis, T. S., & Stetson, J. N. (2004). A Comparison Of Questionnaires For Assessing Website Usability.

Wong, N., Rindfleisch, A., & Burroughs, J. (2003). Do Reverse-Worded Items Confound Measures in Cross-Cultural Consumer Research? The Case Of The Material Values Scale. Journal of Consumer Research, 30, 72–91.

APPENDIX

Table A.1 Original version of the SUS

System Usability Scale (SUS)	**1= Strongly disagree**	**2**	**3**	**4**	**5 = Strongly agree**
1. I think that I would like to use this system frequently.					
2. I found the system unnecessarily complex.					
3. I thought the system was easy to use.					
4. I think that I would need the support of a technical person to be able to use this system.					
5. I found the various functions in this system were well integrated.					
6. I though there was too much inconsistency in this system.					
7. I would imagine that most people would learn to use this system very quickly.					
8. I found the system very cumbersome/awkward to use.					
9. I felt very confident using the system.					
10. I needed to learn a lot of things before I could get going with this system.					

Table A2: SUS-TR

Sistem Kullanılabilirlik Ölçeği	**1 = Hiç katılmıyorum**	**2**	**3**	**4**	**5 = Tamamen katılıyorum**
1. Bu sistemi sıklıkla kullanmak isteyeceğimi düşünüyorum.					
2. Bu sistemi gereksiz bir şekilde karmaşık buldum.					
3. Bu sistemin kullanımının kolay olduğunu düşündüm.					
4. Bu sistemi kullanabilmek için daha teknik bir kişinin desteğine ihtiyaç duyacağımı düşünüyorum.					
5. Bu sistemdeki çeşitli fonksiyonları iyi entegre edilmiş buldum.					
6. Bu sistemde çok fazla tutarsızlık olduğunu düşündüm.					
7. Birçok insanın bu sistemi kullanmayı çok çabuk öğreneceğini sanıyorum.					

8. Bu sistemin kullanımını çok elverişsiz buldum.					
9. Bu sistemi kullanırken kendimden çok emin hissettim.					
10. Bu sistemde bir şeyler yapabilmek için öncelikle bir çok şey öğrenmem gerekti.					