Perceived usability of Arabic System Usability Scale (A-SUS) of Microsoft Teams within Faculty Members

Manal Alsuwaidi
BSc, The Public Authority for Applied Education and Training
Kuwait

ABSTRACT
Computer and mobile applications (apps) were more beneficial for education and applied skills services during COVID-19. There are questions about both theoretical comprehension and actual application regarding how successfully these apps may be used, particularly in languages like Arabic. In order to understand faculty members' opinions on utilizing Microsoft Teams in the departments of basic education, business studies, technological studies, and administrative services, this study uses the Arabic System Usability Scale (A-SUS) survey in Kuwait. According to the A-SUS data, the business and technological faculties are relatively more useful. A master's degree is more common among academics who are between the ages of 30 and 39 in particular. To investigate the present theoretical complaint of the factoring structure of SUS, the data is further compressed into two composite components, usability and learnability. The data theory of SUS in general and A-SUS in particular is supported by these findings, which is important from the faculty's point of view because it considers A-SUS a crucial part of learning management software.

Keywords
Arabic, Applications, Usability, Microsoft Teams, Education, Training.

1. INTRODUCTION
Higher education and its systems have benefited greatly from the growth of information and communication technologies (ICT) during the past few decades [1,2]. ICT has benefitted higher education institutions, especially during the Covid-19 epidemic, when they used it to transform instruction. [3,4]. The development of smartphones and associated applications has further improved the convergence of ICT and education. (apps). The most recent incarnations of the quickly expanding systems are smartphones and applications. According to statistics from 2020, there will be 1.35 billion tablets and 2.7 billion smartphones. [5].

The System Usability Scale (SUS) is primarily used in ICT and systems research to evaluate the relative usability of systems and related programs [6]. Ten items make up the conventional form of the SUS, which are often negative and positive statements with 5-point Likert scales [7]. According to published research, the SUS has been used in a number of systems, including those that provide cognitive behavioral therapy, health care, automated teller machines, insulin pens, social media sites, smartphones, mobile housing applications, email clients, and word processors [8,9,10,11,12,13].

Utilizing SUS is essential for continuous system evaluation and scale evaluation, comparing results with those from prior studies to ensure system quality in terms of usability and sustaining or maintaining use [14]. It is crucial to assess the usefulness of ICT-based learning management systems from a variety of viewpoints, including those of professors and students [15]. If the perceived usefulness is consistently compared and checked, such systems and related applications for educational institutions can result in the satisfaction of relevant stakeholders while also enhancing the sustainability of the overall academic management. In order to improve its usefulness, urgent research is needed from a variety of angles [14].

This article is a component of a series of usability studies conducted on several applications used in a higher education academic context. This study uses the Microsoft Teams application, which is used by the Public Authority for Applied Education and Training. As published research has repeatedly failed to duplicate that factor structure of SUS, such research improves the general understanding of A-SUS in terms of its factor's segregation [7].

2. LITERATURE REVIEW

2.1 Usability Definition
Usability is defined as the "extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (International Organization for Standardization, 2018) p5. It is difficult to develop the same absolute sense and measures for usability's things because it is a concept that is understood and applied contextually. Therefore, operational definitions from their context are needed for both usability and measurements [16].

2.2 Standard Usability Tools
Since its inception, the term "usability," upon which the SUS is based, has been wholly contextual, and as a result, it is divisive in both its science and in the application of experimental psychology, measurement, and statistics [7]. The first systematic usability questionnaires for measuring usability first appeared in the late 19th century [16,17].

Psychometric tests, such as usability inspection techniques like heuristic evaluation and user testing, can also be used to evaluate an app's usability [18]. Heuristic evaluation is based on the judgment of an expert who examines the program and spots usability issues from the viewpoints of possible users. When an app is being tested, a sample user is chosen, instructed to use it, and asked to report any issues they encountered. In contrast, SUS, a psychometric measuring tool, uses the Likert scale to gather user feedback that will be utilized as the foundation for assessing usability and give critical information for the scale's factor structure.

2.3 SUS Specifically
The SUS, which has been introduced by largely functions as an opinion-based questionnaire [19]. It is a well-known

standardized test for judging how usable systems and applications are thought to be in a variety of industries [7]. The equipment is hailed as being extremely dependable, rapid, simple to use, and inexpensive, making it perfect for large-scale international surveys in which businesses are typically interested. SUS grades applications from 0-100 and consists of ten questions that are worded both positively and negatively. The questions have a Likert scale with a range of 5 to 7. Recent initiatives appear to divide the ten SUS sections into a more composite structure for the two components of usability and learnability [7]. As they appear to be more focused on scoring learning than usability, items 4 and 10 are kept in the learnability category. The remaining eight objects are preserved in usable condition at the same time.

2.4 Translated SUS into Arabic Language

Using structured translation and validation procedures, the standard SUS has recently been translated into other languages, including Arabic, Chinese, French, German, Hindi, and Spanish [6]. The primary goals of these ongoing translation projects for usability surveys and questionnaires are to maintain contextual relevance and to approach the usability community more confidently because they feel at home utilizing a translated SUS in their native language [14]. Additionally, they seek to document the rhetorical and contextual aspects [6], which will improve adherence to the written words. Compliance of adapted SUS with SUS can lead to confidence-building and eventually standardization of them, from which the A-SUS inquiry rises, as this study tries to test the factor structure of such a translated SUS.

However, since usability is contextual and difficult to quantify with precise figures, the usability evaluation of A-SUS for the Arabic region is modest and lacking in confidence [6,15]. As a result, the faculty context in the Arabic-speaking world can add intriguing elements, particularly the verification of a composite structure (usability and learnability) [7] and how the A-SUS scores fluctuate among faculty in the region with various ages, specializations, and qualifications. A way for achieving these goals is described in the next section.

3. METHODOLOGY

An enterprise-standard usability tool is implemented on the Microsoft Teams application to aid the Arabic region in usability studies. The Microsoft Teams application is discussed from the viewpoints of academic staff members in this study. The Arabic System Usability Scale (A-SUS) tool, a translation of the Conventional System Usability Scale into Arabic, is used to administer it. Data on Microsoft Teams' perceived usability are subjected to psychometric evaluation in this study. The methodology for this study is described in the next section.

Data collection and analysis for SUS research use the survey approach. Ten questions from the A-SUS scale, which is described in [24], are used in this study. There are studies that use a 5-point scale that ranges from strongly disagree to highly agree. Usability is defined as having a minimum score of 68 on a scale from 0 to 100 [1,14]. 54 active faculty members who used PAAET to fill out the A-SUS were chosen for the study because it attempts to develop a faculty viewpoint. A sample size of at least 200 people is ideally regarded as robust and sufficient [28]. However, there are studies that have produced strong arguments using samples under 200, such [5]'s usage of 50 replies. Additionally, the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy is conducted. According to a general rule, sampling is sufficient when KMO values fall between 0.8 and 1 [17]. The KMO value for this study is 0.855, demonstrating that the data are sufficient for the component analysis [18].

Principal component analysis (PCA), descriptive statistics, and Cronbach's alpha were all used in the statistical analysis. Ten SUS elements are factored into two primary components while doing the PCA (PC1 = Usability, all items other than 4 and 10, PC2 = Learnability (Q4 and Q10) [24]. To determine whether an instrument has internal consistency and is dependable, Cronbach's alpha is utilized. As opposed to this, descriptive statistics were employed to determine how the scores varied according on the age, field, and qualification of the professors. The investigation was able to validate the factor structuring thanks to principal component analysis. The study may transform and validate all ten A-SUS questions into two composite constructs, usability and learnability, which are factors in the SUS factor structure [24, 28].

4. RESULTS AND INTERPRETATIONS COLLEGE-WISE SCORES OF A-SUS

Table 1: Scale Reliability Statistics

	mean	sd	Cronbach's α
scale	3.50	0.750	0.907

The A-SUS items have a greater internal consistency, and as a result, the instrument is extremely dependable. According to the general norm, Cronbach's >.90 signifies "excellent" internal consistency [17]. Table 1 displays a Cronbach's alpha coefficient of =0.907, which is much higher than the minimally necessary standard value of 0.70.

Table 2: Frequentist Individual Item Reliability Statistics

	If item dropped			
Item	Cronbach's α	Item-rest correlation	mean	sd
Q1	**0.898**	**0.663**	**3.58**	**1.057**
Q2	**0.893**	**0.770**	**3.81**	**0.903**
Q3	**0.889**	**0.819**	**3.80**	**0.963**
Q4	**0.902**	**0.629**	**3.29**	**1.221**
Q5	**0.895**	**0.712**	**3.41**	**0.966**
Q6	**0.899**	**0.652**	**3.43**	**1.018**
Q7	**0.902**	**0.599**	**3.53**	**0.998**
Q8	**0.895**	**0.726**	**3.90**	**0.901**
Q9	**0.894**	**0.732**	**3.54**	**0.993**
Q10	**0.912**	**0.453**	**2.74**	**1.101**

Note. The following items were reverse scaled: Q2, Q4, Q6, Q8, Q10.

The item-wise co-efficient is shown in Table 2. Six of the ten A-SUS items (Q1, Q4, Q5, Q6, Q9, and Q10) have Cronbach's alpha values larger than 0.90, while the remaining four items (Q2, Q3, Q7, and Q8) have alpha values slightly below 0.90. As a result, since each item scored greater than the benchmark co-efficient value of 0.70, each item is also independently dependable. Additionally, this indicates that four items fit into the "Good" internal consistency category, while six items go into the "Excellent" category [18].

4.1 Demographics

Table 3: Demographics of the study

Frequencies of College

Levels	Counts	% of Total	Cumulative %
Basic Education College	49	22.2 %	22.2 %
College of Business Studies	63	28.5 %	50.7 %
College of Health Sciences	14	6.3 %	57.0 %
College of Technological Studies	37	16.7 %	73.8 %
Higher Institute for Administrative Services	25	11.3 %	85.1 %
Kuwait University	1	0.5 %	85.5 %
Nursing School	2	0.9 %	86.4 %
Others	30	13.6 %	100.0 %

Frequencies of Education

Levels	Counts	% of Total	Cumulative %
Bachelor	74	33.5 %	33.5 %
Master	55	24.9 %	58.4 %
Ph.D	92	41.6 %	100.0 %

Frequencies of Age

Levels	Counts	% of Total	Cumulative %
30-40	38	17.2 %	17.2 %
41-59	152	68.8 %	86.0 %
60 and above	31	14.0 %	100.0 %

Table 3 shows the complete demographics of the respondents, who attended seven colleges and fell into three different age groups and academic levels. First off, when looking at the participants by college, 28.5% (n=63) were from the College of Business Studies, 22.2% (n=49) from the Basic Education College, 16.7% (n=37) from the College of Technological Studies, and 11.3% (n=25) from the Higher Institute for Administrative Services.

Second, when looking at the respondents' educational backgrounds, 41.6% (n=92) had Ph.D.s, followed by 33.5% (n=74) who held bachelor's degrees, and the remaining 24.9% (n=55) who held master's degrees. Thirdly, when looking at participants' ages, 68.8% (n=152) were between the ages of 41 and 59, while 17.2% (n=38) were between the ages of 30 and 40, and just 14% (n=31) of respondents were over the age of 60.

5. PRINCIPAL COMPONENT ANALYSIS

Table 4: Component Loadings

Component Loadings		
	Component	
	1	**Uniqueness**
sus score	0.998	0.00356
1. I think this app is very easy to use.	0.873	0.23874
2. I found this app too complicated	-0.833	0.30656
3. I felt totally confident using this app.	0.801	0.35800
4. I found this app strange to use.	-0.795	0.36799
5. I found the multiple functions in this application consistent with each other.	0.778	0.39485
6. I think I like to use this app constantly.	0.735	0.45988
7. I thought there was a lot of conflict in using this app.	-0.722	0.47809
8. I think I need help from a technical person to use this app.	-0.700	0.50979
9. I imagine a lot of faculty / training will learn to use this app easily.	0.690	0.52404
10. You must know many things to facilitate the use of this application.	-0.527	0.72269

Note. 'varimax' rotation was used

Component loadings in principal component analysis are the correlations between the components (Usability & Learnability) and the variables (Q1-Q2) in the analysis. The percentage of that variable's volatility that the component accounts for is known as its loading. The maximum variance explained by PC1 (usability) in Q3 is 0.886, and the lowest variance explained by PC1 in Q6 is 0.464. Q4's variance is most strongly explained by PC2 (learnability), followed by Q10, in contrast. These findings support other research' assertions and interpretations [24] that learnability and usability are independent constructs that should be used in separate analyses.

Table 5: Component Characteristics

Initial Eigenvalues			
Component	**Eigenvalue**	**% of Variance**	**Cumulative %**
1	6.636	60.33	60.3
2	0.955	8.68	69.0
3	0.771	7.01	76.0
4	0.652	5.92	81.9
5	0.503	4.57	86.5
6	0.402	3.65	90.2
7	0.360	3.27	93.4
8	0.294	2.67	96.1
9	0.284	2.58	98.7
10	0.145	1.32	100.0
11	3.71e-16	3.37e-15	100.0

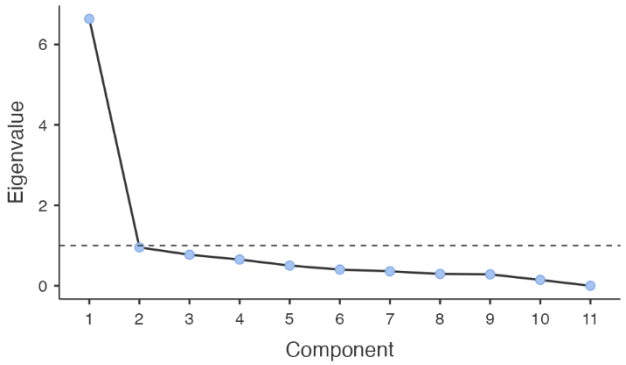

Figure 1: Scree plot

Each of the ten A-SUS questions' eigenvalues are displayed on a scree plot (see Table 10 and Figure 1) in Table 10. The varimax rotation is used to understand the eigenvalue, which is bigger than one. The subsequent varimax factor loadings are classified as "strong" if they are larger than 0.75, "moderate" if they are between 0.5 and 0.75, and "weak" if they are between 0.3 and 0.49. With an eigenvalue of 0.583, both PC1 (usability) and PC2 (learnability) fall into the category of moderate loading, whereas PC2's eigenvalue is just 0.164 and falls into the category of weak factor loading.

6. DISCUSSION

In this study, the faculty members' perceptions of the Microsoft Teams mobile application's usability are measured using the A-SUS instrument. The usability score is higher than the criterion for faculty from colleges of business and technology, particularly among senior faculty between the ages of 30 and 40 who have a Master's degree, according to subsequent analysis. The total usability score is below the average usability standard. This conclusion suggests that subjective usability is dependent on the user's attributes since they establish the context for the program [22, 30]. This finding also demonstrates the need for careful attention to detail because usability cannot be standardized for all sorts of people [16]. This study discovered that the context, in particular, derives from the users' age, their credentials, and the subject matter they teach.

Contextual usability increases the dimensionality of the A-SUS scores. More composite architectures are required in order to have a better halitotic comprehension of the system usability [13]. The two primary components in principle component analysis that account for variations in various A-SUS items are usability and learnability. According to recent discussions in the literature [24], Q4 and Q10 have shown increased loading to learnability, substantiating the idea that learnability should be researched and used in system studies, especially for Microsoft Teams' rookie users. This is in line with the research, which asserts that the SUS score merges into a single scale of perceived usability as users become more accustomed to the system.

7. CONCLUSION AND FUTURE RESEARCH

SUS has a recent history of being an efficient, affordable, and—most importantly—reliable instrument for gauging the usefulness of systems on a big scale. To reflect the crucial contextuality that the Arabic language and culture add, the translated A-SUS must be continually evaluated across a range of areas and stakeholders. The faculty viewpoint of A-SUS is included in this paper. Based on the topic matter, faculty age and qualifications, and how their scores are composited back into the usability and learnability as standalone constructs among the faculty, it yields a fascinating range of results. These findings will provide the A-SUS more theoretical heft to gather more precise information about variances in usability scores, particularly for educational apps like the Microsoft Teams program. The results have significant implications for usability practitioners, including the Public Authority for Applied Education and Training in Kuwait and other comparable organizations, who should carefully implement educational and training applications with special attention to customization for the age, the subjects, and the teacher qualifications.

8. REFERENCES

[1] Anie, S. O. (2011). The economic and social benefits of ICT policies in Nigeria. *Library Philosophy and Practice*, *1*, 125.

[2] Livingstone, S. (2012). Critical reflections on the benefits of ICT in education. *Oxford Review of Education*, *38*(1), 9–24.

[3] Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P., & Lam, S. (2020). COVID-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching*, *3*(1), 1–20.

[4] Nuere, S., & de Miguel, L. (2020). The Digital/Technological Connection with COVID-19: An Unprecedented Challenge in University Teaching. *Technology, Knowledge and Learning*, 1–13.

[5] Blair, I. (2020). *Mobile App Download and Usage Statistics (2020)*. https://buildfire.com/app-statistics/

[6] Gao, M., Kortum, P., & Oswald, F. L. (2020). Multi-Language Toolkit for the System Usability Scale. *International Journal of Human-Computer Interaction*, *00*(00), 1–19. https://doi.org/10.1080/10447318.2020.1801173

[7] Lewis, J. R., & Sauro, J. (2017). Revisiting the Factor Structure of the System Usability Scale. *Journal of Usability Studies*, *12*(4), 183–192.

[8] De Angeli, A., Athavankar, U., Joshi, A., Coventry, L., & Johnson, G. I. (2004). Introducing ATMs in India: a contextual inquiry. *Interacting with Computers*, *16*(1), 29–44.

[9] Frandsen-Thorlacius, O., Hornbæk, K., Hertzum, M., & Clemmensen, T. (2009). Non-universal usability? A survey of how usability is understood by Chinese and Danish users. *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems, 41–50.

[10] Hsieh, H. C. L. (2011). Exploring the impact of cultures on web usability test. *International Conference on Human Centered Design*, 47–54.

[11] Jonsson, O., Haak, M., Tomsone, S., Iwarsson, S., Schmidt, S. M., Mårtensson, K., Svensson, T., & Slaug, B. (2016). Cross-National Usability Study of a Housing Accessibility App: Findings From the European InnovAge Project. *Journal of Usability Studies*, *12*(1).

[12] Mol, M., Van Schaik, A., Dozeman, E., Ruwaard, J., Vis, C., Ebert, D. D., Etzelmueller, A., Mathiasen, K., Moles, B., Mora, T., Pedersen, C. D., Skjøth, M. M., Pensado, L. P., Piera-Jimenez, J., Gokcay, D., Ince, B. Ü., Russi, A., Sacco, Y., Zanalda, E., … Smit, J. H. (2020). Dimensionality of the system usability scale among professionals using internet-based interventions for depression: A confirmatory factor analysis. *BMC Psychiatry*, *20*(1), 1–10. https://doi.org/10.1186/s12888-020-02627-8

[13] Walsh, T., Nurkka, P., & Walsh, R. (2010). Cultural differences in smartphone user experience evaluation. *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, 1–9.

[14] AlGhannam, B. A. (2018). Perceived Usability Using Arabic System Usability Scale (A-SUS): Student Perspective of Smart PAAET App. *International Journal of Computer Science and InformationSecurity*, *16*(7), 66–78.

[15] AlGhannam, B. A., Albustan, S. A., Al-Hassan, A. A., & Albustan, L. A. (2018). Towards a Standard Arabic System Usability Scale: Psychometric Evaluation using Communication Disorder App. *International Journal of Human-Computer Interaction*, *34*(9), 799–804. https://doi.org/10.1080/10447318.2017.1388099

[16] Brooke, J. (1996). SUS: A "quick and dirty'usability. In *Usability evaluation in industry* (p. 189). CRC press.

[17] Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring Perceived Usability: The SUS, UMUX-LITE, and AltUsability. *International Journal of Human-Computer Interaction*, *31*(8), 496–505. https://doi.org/10.1080/10447318.2015.1064654

[18] Brajnik, G. (2000). Automatic web usability evaluation: what needs to be done. *Proc. Human Factors and the Web, 6th Conference*. http://users.dimi.uniud.it/~giorgio.brajnik/papers/hfweb00.html

[19] Brooke, J. (2018). System usability scale (SUS). *Iron and Steel Technology*. https://doi.org/10.5948/upo9781614440260.011.

[20] Al Ghurair, N., & Alnaqi, G. (2020). Schemes of eStories for Children with Social Communication Difficulties. Advances in Human-Computer Interaction, 2020, 1–9. https://doi.org/10.1155/2020/9530218

[21] Binyamin, S., Rutter, M., & Smith, S. (2016). the Utilization of System Usability Scale in Learning Management Systems: a Case Study of Jeddah Community College. ICERI2016 Proceedings, 1, 5314–5323. https://doi.org/10.21125/iceri.2016.2290

[22] Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 213–218.

[23] Drew, M. R., Falcone, B., & Baccus, W. L. (2018). What does the system usability scale (SUS) measure? International Conference of Design, User Experience, and Usability, 356–366.

[24] Finstad, K. (2006). The system usability scale and non-native English speakers. Journal of Usability Studies, 1(4), 185–188.