REVIEW OF RESEARCH
IMPACT FACTOR : 5.7631(UIF) UGC APPROVED JOURNAL NO. 48514 ISSN: 2249-894X
VOLUME - 7 | ISSUE - 11 | AUGUST - 2018

A STUDY ON THE RELIABILITY OF MARS SCALE AS AN ASSESSMENT TOOL TO GAUGE THE QUALITY OF FASHION APPLICATIONS

Vidhi Shah[1] and Dr. Rachna Gandhi[2]
[1]Research Scholar, Gujarat University.
[2]Assistant Professor, K. S. School of Business Management.

ABSTRACT

The usage of mobile shopping application for the purchase of fashion products has grown exponentially in recent years. However, there is currently no application quality assessment tool except for the star ratings to assess the apps. The object of this study is to customise a reliable, multidimensional measure for the purpose of rating the quality of mobile shopping applications. A literature search conducted revealed that there is currently no such scale available for mobile shopping applications, but a Mobile Application Rating scale with subscales, descriptors and anchors has been developed to judge the quality of health apps. Modification in the same developed scale has been undertaken and a survey through structured questionnaire of 522 respondents was conducted and the reliability of the scale has been checked for the same to be useful to researcher and application developer in future.

KEY WORD: *Mobile Shopping Applications, Objective Quality Dimensions, MARS, Rating Scale, Fashion Applications.*

INTRODUCTION:-

Mobile application usage for shopping has grown exponentially in recent years [1]. In the span of 2015-2017 the global use of smart phones increased by 906 million, reaching 2.82 billion devices (up 5% in a year), thereby increasing the internet use through mobile phones by 72% in a year [2]. 10.4 billion apps were downloaded in the first quarter of 2017 [3], with a projection of 111 billion for the whole year [4]. Because of the portability feature of smart phones, the application provides with a lot of ease and any time access. Given the rapid proliferation of mobile phone applications, especially smart phone applications, it is getting difficult for users as well as the companies, academicians and researchers to really identify the best quality applications and the reasons or factors indicating the quality assessment of the application. [5]. Barely any information is available on the quality of applications apart from the star ratings as published on the retailer's web pages. Also, the application reviews are quite subjective and may be coming from suspicious sources and not trustworthy [6]. Just by checking the popularity of the applications, it would yield barely any fruitful insights on the quality of the applications [7] The literature found so far has most of them focusing on technicality of websites, mostly presented in form of checklist which would not judge the quality of those features. [8-10]. Much analysis on application assessment has not been found, however website quality can be described as a function of content, appearance and multimedia, navigation, structure & design and uniqueness [11]. Kim et all [12] shortlisted 165 evaluation criteria while synthesising website evaluation criteria. Out of those 165, 33 were coded miscellaneous as not being able to be grouped and rest were clustered in 13 groups. Hence a reliable and objective yardstick is much needed to rate the quality of shopping application due to the scarce research available in that domain.

OBJECTIVES

The purpose of this paper is to modify and test the applicability of the MARS scale in order to get a reliable, multidimensional scale for rating the quality of mobile shopping apps for fashion products.

METHODOLOGY

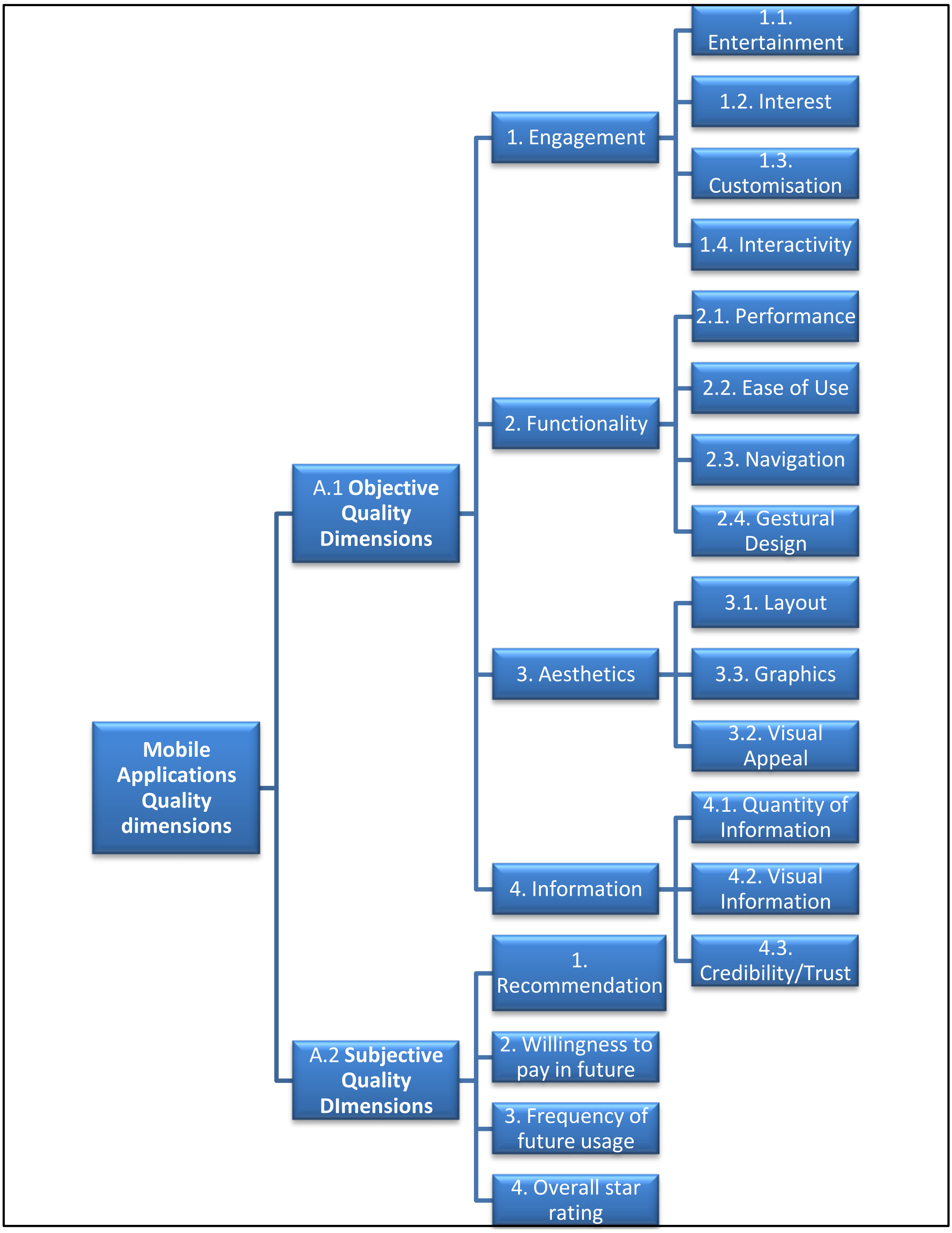
A detailed research was conducted to recognise articles, papers and publications explicitly containing web or application-based rating scale. Papers from January 2000 to January 2017 were taken into consideration, retrieved from EBSCO, ProQuest, IEEE Xplore, Science Direct and the online database of IIMA library. The terminologies used while searching included “mobile” “application” paired with “assessment criteria” “rating scale” “quality” “judge”.

Upon a comprehensive literature search, it was found that there has been no research in the direction of developing a uniform scale to access the quality of mobile shopping application specially to purchase fashion products. However, a journal article where in an assessment scale was developed for m-health applications was found. On diving deeper, the scale had been tried and tested for health applications but to be used in another domain. Hence as recommended by the authors for future scope of the research, the MARS (Mobile Application Rating Scale) by Stoyanov et. al [13] has been taken under scrutiny to check if it could be used as an assessment scale for shopping applications, specially fashion products.

The Mobile Application Rating Scale has quality criteria clustered within the engagement, functionality, aesthetics, information quality, and subjective quality categories, to develop 23 subcategories from which the 23 individual MARS items were developed. Each item on the scale uses a 5-pointer scale of 1 through 5 from inadequate to excellent. Out of these 23 subcategories, 18 subcategories which were found to be best suited for the application concerned with shopping of fashion products were retained. In order to check the reliability and internal consistency of the scale so obtained by deletion of certain subcategories is checked by the reliability statistics of SPSS. Cronbach Alpha and Intra-Correlated Correlation (ICC) have been calculated to check the same. The subjective quality dimensions have not been included in the calculation of ICC and Alpha owing to its subjective nature.

CONCEPTUAL FRAMEWORK

The conceptual framework of the four constructs of objective quality dimension having 14 subcategories and 4 subcategories of subjective quality dimension as modified by the researcher is shown below.



(Mobile App Rating Scale {MARS} for Mobile fashion shopping Application, Modified by the Researcher)

EMPIRICAL ANALYSIS & FINDINGS

Engagement Sub-Scale Inter-Rater Reliability

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.837	.837	4

Item Statistics

	Mean	Std. Deviation	N
Entertainment	3.80	.946	522
Interest	3.78	1.014	522
Customisation	3.80	1.041	522
Interaction	3.88	1.013	522

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.812	3.780	3.877	.098	1.026	.002	4
Inter-Item Correlations	.562	.524	.601	.076	1.145	.001	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Entertainment	11.45	6.751	.647	.424	.803
Interest	11.47	6.311	.685	.474	.786
Customisation	11.45	6.218	.679	.467	.789
Interaction	11.37	6.411	.662	.440	.796

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
15.25	10.828	3.291	4

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.562[a]	.520	.603	6.124	521	1563	.000
Average Measures	.837[c]	.813	.858	6.124	521	1563	.000

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

In order to understand interclass correlation co-efficient

Functionality Sub-Scale Inter-Rater Reliability

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.879	.880	4

Item Statistics

	Mean	Std. Deviation	N
Performance accuracy	3.99	.963	522
Ease of Use	4.08	.859	522
Navigability	4.15	.953	522
Gestural Designs	4.10	.883	522

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.081	3.994	4.148	.153	1.038	.004	4
Inter-Item Correlations	.647	.620	.683	.063	1.101	.001	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Performance accuracy	12.33	5.607	.723	.526	.851
Ease of Use	12.24	5.987	.739	.551	.845
Navigability	12.18	5.616	.732	.544	.847
Gestural Designs	12.23	5.807	.763	.587	.835

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
16.32	9.832	3.136	4

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.644[a]	.607	.680	8.238	521	1563	.000
Average Measures	.879[c]	.861	.895	8.238	521	1563	.000

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

Aesthetics Sub-Scale Inter-Rater Reliability

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.805	.809	3

Item Statistics			
	Mean	Std. Deviation	N
Layout	3.89	.942	522
Graphics	3.87	.953	522
Visual Appeal	4.05	1.026	522

Summary Item Statistics							
	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.937	3.874	4.050	.176	1.045	.010	3
Inter-Item Correlations	.585	.464	.753	.289	1.622	.018	3

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Layout	7.92	3.011	.689	.573	.697
Graphics	7.94	2.839	.748	.612	.633
Visual Appeal	7.76	3.150	.535	.296	.859

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
11.81	6.150	2.480	3

Intraclass Correlation Coefficient							
	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.579[a]	.533	.623	5.127	521	1042	.000
Average Measures	.805[c]	.774	.832	5.127	521	1042	.000
Two-way random effects model where both people effects and measures effects are random.							
a. The estimator is the same, whether the interaction effect is present or not.							
b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.							

Information Sub-Scale Inter-Rater Reliability

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.840	.842	3

Item Statistics

	Mean	Std. Deviation	N
Quantity of Information	4.12	.795	522
Visual information	4.09	.831	522
Trust	4.03	.901	522

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.080	4.029	4.121	.092	1.023	.002	3
Inter-Item Correlations	.639	.529	.695	.167	1.315	.007	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Quantity of Information	8.12	2.540	.661	.487	.818
Visual information	8.15	2.200	.794	.631	.688
Trust	8.21	2.241	.666	.486	.820

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
12.24	4.848	2.202	3

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.636[a]	.594	.676	6.243	521	1042	.000
Average Measures	.840[c]	.814	.862	6.243	521	1042	.000

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

Summarised Result of Interrater reliability and internal consistency of the MARS items and subscale scores, and corrected item-total correlations and descriptive statistics of items, based on independent ratings of 5 applications where fashion products can be purchased by 522 respondents.

Subscale/item	Corrected item Total Correlation	Mean	SD
Engagement alpha = 0.83, ICC = 0.83 (95% CI 0.81-0.85)			
1 Entertainment	.647	3.80	.946
2 Interest	.685	3.78	1.014
3 Customization	.679	3.80	1.041
4 Interactivity	.662	3.88	1.013
Functionality alpha = 0.87, ICC = 0.87 (95% CI 0.86-0.89)			
5 Performance	.723	3.99	.963
6 Ease of use	.739	4.08	.859
7 Navigation	.732	4.15	.953
8 Gestural design	.763	4.10	.883
Aesthetics alpha = 0.80, ICC = 0.80 (95% CI 0.77-0.83)			
9 Layout.	.689	3.89	.942
10 Graphics	.748	3.87	.953
11 Visual appeal:	.535	4.05	1.026
Information alpha = 0.84, ICC = 0.84 (95% CI 0.81-0.86)			
12 Quantity of information	.661	4.12	.795
13 Visual information	.794	4.09	.831
14 Trust	.666	4.03	.901

The descriptive statistics indicate the mean and standard deviation of the each of the objective quality dimensions for the responses of 522 respondents. To understand the Reliability of the scale, the Reliability statistics is calculated using SPSS. Further to get fair inter-rater reliability ICC is calculated considering two-way mixed effect.

The Cronbach Alpha for the Engagement Sub-scale consisting of four quality objective dimensions is 0.837 with an ICC of 0.837, having a 95% confidence to range between 0.81-0.85. These four quality objective dimensions include: Entertainment, Interest, Customisation and Interaction, each one having an item-total correlation of being higher than 0.60. For functionality Sub-scale, it is 0.87 which is even better with ICC 0.87, having a 95% confidence to range between 0.86-0.89. The four quality objective dimensions include: Performance accuracy, Ease of Use, Gestural design and navigation, each one having the item-total correlation of higher than 0.70. Aesthetics Cronbach's alpha is 0.80 with ICC = 0.80 (95% CI 0.77-0.83). The objective quality dimensions: Layout, graphics and visual appeal, have item-total correlation of 0.5 and above. Information Cronbach's Alpha is 0.84, ICC = 0.84 (95% CI 0.81-0.86). The three objective quality dimensions considered: Quantity of Information, Visual Information and Trust have item-total correlation of 0.66 or higher.

The general rule of thumb indicates a value of 0.7 or greater for Cronbach's alpha as good and above 0.80 to be even better. Thus, the scale has all four sub-scales having a very good value of reliability. Similarly, the ICC value shows a great level of internal consistency amongst the quality dimensions.

CONCLUSION

The multidimensional scale is consisting of four construct sub-scale with 14 objective quality dimensions and subjective quality dimensions without any changes in original MARS for the purpose of assessing the health and quality of mobile shopping applications for the purpose of fashion products is obtained. It is seen that the Cronbach alpha of each of the construct and its Intra-Corrected Correlation for checking Reliability and internal consistency seems moderately high. This scale provides a reliable app quality rating scale as per the current results. Hence, currently it can used for further researchers for pilot study and eventually can be used for other purposes.

REFERENCES

1. Riley, W. T., Rivera, D. E., Atienza, A. A., Nilsen, W., Allison, S. M., & Mermelstein, R. (2011). Health behavior models in the age of mobile interventions: are our theories up to the task?. *Translational behavioral medicine*, *1*(1), 53-71.

2. Jung, H. (2011). *Cisco visual networking index: global mobile data traffic forecast update 2010–2015*. Technical Report, Cisco Systems Inc. 2011. Available online: https://www. cisco. com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429. html (accessed on October 2017).

3.http://www.theguardian.com/technology/appsblog/2013/sep/19/gartner-mobile-apps-revenues-report.

5. Cummings, E., Borycki, E. M., & Roehrer, E. (2013). Consumers Using Mobile Applications. *Enabling Health and Healthcare Through ICT: Available, Tailored and Closer*, *183*, 227.

6. Kuehnhausen, Martin, and Victor S. Frost. "Trusting smartphone apps? To install or not to install, that is the question." In *2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pp. 30-37. IEEE, 2013.

7. Girardello, A., & Michahelles, F. (2010, September). AppAware: Which mobile applications are hot?. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 431-434). ACM.

8. Seethamraju, R. (2004). Measurement of user perceived web quality. *ECIS 2004 Proceedings*, 176.

9. Olsina, L., & Rossi, G. (2002). Measuring Web application quality with WebQEM. *Ieee Multimedia*, *9*(4), 20-29.

10. Aladwani, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information & management*, *39*(6), 467-476.

11. Moustakis, V., Litos, C., Dalivigas, A., & Tsironis, L. (2004, November). Website Quality Assessment Criteria. In *ICIQ* (pp. 59-73).

12. Kim, P., Eng, T. R., Deering, M. J., & Maxfield, A. (1999). Published criteria for evaluating health related web sites. *Bmj*, *318*(7184), 647-649.

13. Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth and uHealth, 3(1), e27.