Holistic User eXperience in Mobile Augmented Reality Using User eXperience Measurement Index

Fahad Ahmed Satti
Ubiquitous Computing Lab Department of Computer Engineering KyungHee University
South Korea
fahad.satti@oslab.khu.ac.kr

Jamil Hussain
Ubiquitous Computing Lab Department of Computer Engineering KyungHee University
South Korea
jamil@oslab.khu.ac.kr

Hafiz Syed Muhammad Bilal
Ubiquitous Computing Lab Department of Computer Engineering KyungHee University
South Korea
bilalrizvi@oslab.khu.ac.kr

Wajahat Ali Khan
Ubiquitous Computing Lab Department of Computer Engineering KyungHee University
South Korea
wajahat.alikhan@oslab.khu.ac.kr

Asad Masood Khattak
College of Technological Innovation Zayed University
Abu Dhabi, UAE
asad.khattak@zu.ac.ae

Ju Eun Yeon
Ubiquitous Computing Lab Department of Computer Engineering KyungHee University
South Korea
annyeon@oslab.khu.ac.kr

Sungyoung Lee
Ubiquitous Computing Lab Department of Computer Engineering Kyung Hee University,
South Korea.
sylee@oslab.khu.ac.kr

***Abstract*—User eXperience (UX) evaluation in the field of Mobile Augmented Reality (MAR) is a challenging task, which requires the application of many heterogeneous methods, producing a variety of raw signals and subjective data. This multi-method approach is essential for capturing the holistic UX of any product, service or system. In order to convert this data into information and subsequently knowledge, a comprehensive and scalable system is required which can not only quantify the individual UX metrics but also produce a concise result, which is interpretable by anyone. We call this result, the User Experience Measurement Index, and in this paper we present the results of adopting the mixed method UX evaluation approach for evaluating a prototype MAR application using various methods and sensors, applied before, during, and after its usage. Additionally, we present the methodology and results for calculating the UXMI.**

Keywords— Mobile Augmented Reality, Mixed Method, UX, UXMI

I. INTRODUCTION

Mobile Augment Reality (MAR) represents one of the most prominent technological advancement, that has led to a paradigm shift in the way information is presented to the users. Seamless, superimposition of virtual objects in the real-world environment, using smart phones, allows the users to experience immersive environments, within a visually comfortable environment. Additionally, with a good design and development process, which takes into account the user's familiarity[1], MAR systems and applications can lead to an improvement in usability and learnability[2] and lessen the effects of motion sickness, which are otherwise very prominent in Virtual Reality(VR)[3]–[5].

However, behind every usable and stable interactive system, is a thorough testing process. This process re-enforces the confidence of the stakeholders and allows the designers and developers to determine the extent to which, the users can achieve some pre-determined goals of the system, product, or service[6]. A key aspect of this process is to determine the User's experience, which pertains to the user's perception and responses, before, during, and after the use of an interactive system[7]. The UX evaluations measure subjective features of a user's interaction, using qualitative and quantitative methodologies[8]. As a result, UX evaluations are highly dependent on the context of the user (in terms of motivation and ability), system (in terms of its capabilities), and environment (in terms of comfortability and the existence of necessary triggers) of use[9]. UX evaluator biasedness and the subjectivity of UX evaluation results, is also a critical problem, which can be mitigated by utilizing the experience of many researchers working together to design the study, apply it, and interpret the results[10].

Therefore, in order to evaluate a holistic UX, multidimensional features are needed, which can not only reduce subjectivity of the evaluations but also re-enforce the results and provide both spatial and temporal interaction data[11]. While several initiatives have been taken for evaluating UX in interactive systems, such as Virtual Reality (VR) and Augmented Reality (AR), the domain of UX in MAR is relatively untapped[12]. This presents a good opportunity for the UX research community to apply state-of-the-art tools and techniques for evaluating holistic UX[13]. Additionally, the worldwide interest in augmented reality is also growing steadily, as shown in Figure 1. It is also interesting to note that the two peaks in the search trends, correspond with the release of Pokémon Go by Ninatic Inc. and iOS 11 by Apple Inc. both representing MAR systems, with a loyal fan base and marketing initiative. A mixed method UX evaluation approach for MAR systems is necessary to amplify their existing benefits and will lead to better acceptance by the users. This involves, application of a multitude of tools and techniques to gather a combination of observational, self-reported, and physiological measurements, by the UX expert before, during, and after usage of the MAR system[14].

978-1-7281-1460-6/ 19/$31.00 ©2019 IEEE

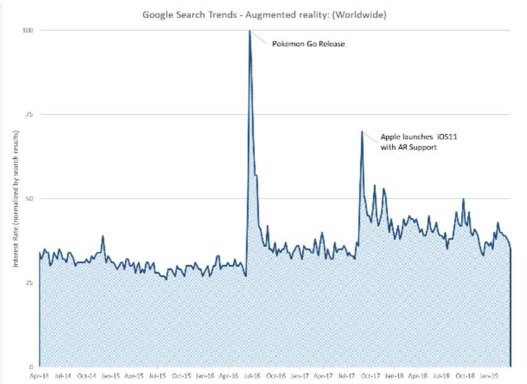


Fig. 1. Worldwide Google search trends in the last 5 years for the term "Augmented reality"

Figure 2, represents the UX evaluations ecosystem in MAR, which have to take into account the constraints, such as VR sickness, hardware and software performance, user's background, immersion requirements, and the interactive system, both in its virtual environment and the real environment. These factors when applied to the context of use, are used to evaluate the various aspects of UX. These correspond to the user's perception of non-instrumental qualities (such as hedonic qualities, fun, aesthetics, and others), user's perception of instrumental qualities (such as effectiveness, learnability, pragmatics, and others), and user's emotional responses (such as VR sickness, sadness, happiness, and others). These aspects, enable the UX researcher to identify the overall consequences in terms of anticipated user satisfaction, or in more general terms positive or negative UX.

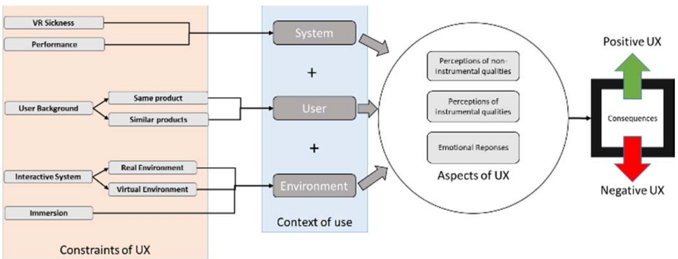


Fig. 2. UX evaluation ecosystem in MAR

In this paper, we present our approach towards solving this problem by the application of a mixed method UX evaluation approach, its application on a prototype MAR application, and a novel model for precise calculation of holistic UX.

II. Related Work

UX evaluations in the domain of AR/VR, have mainly focused on the utilization of individualistic approaches, such as utilizing questionnaires, observations, and physiological sensors which are used to evaluate user's perception of instrumental and non-instrumental qualities, and emotional responses. An abstract representation of some of these methods are shown in Table I.

UX metrics pertaining to the AR environment have been identified in the survey conducted by [15]. These include performance metrics, issue metrics, and behavioral and physiological metrics for each UX component (product, user, system, and service).

Effects of stereoscopic display on UX, in terms of task effectiveness, learning outcomes, user experience, flow, and immersion have indicated positive outcomes[16]. The authors utilized questionnaires for their evaluation and were able to collect qualitative data only.

In [5], the authors have explored the use of questionnaires and physiological sensors to evaluate the user's perception and physiological responses towards VR sickness. Utilizing NASA-TLX for evaluating physical and mental effort, Simulator Sickness Questionnaire (SSQ) and Subjective Units of Discomfort/Distress Scale (SUDs) for evaluating VR sickness, and physiological sensors for gaining implicit user data. The authors found a high correlation between the user's explicit responses to questionnaires and their implicit physiological responses. While the approach used by the authors was individualistic, it does indicate the effectiveness of a mixed method approach, which can enrich the UX results.

TABLE I. FACTORS IMPACTING UX EVALUATION IN AR/VR

UX evaluation methods in MAR/VR	UX evaluation techniques in AR/VR	Measureable factors	Aspects of UX
Questionnaires (Explicit UX evaluation)	NASA-TLX	Physical and Mental effort	Perception of instrumental qualities
	SSQ, SUDs	VR Sickness	Emotional responses
	SGUS	Learnability	Perception of instrumental qualities
	AtrikDiff, QUIS	Hedonic Quantities	Perception of non-instrumental qualities
		Pragmatic Quantities	Perception of instrumental qualities
Physiological Sensors (Implicit)	GSR, EMG, ECG	Interest, Stress, Focus, Relaxation, Excitement	Emotional responses
	EEG	Sadness, Happiness, Disgust, Fear, Angry, Neutral	Emotional responses
	Eye Tracking	Sadness, Happiness, Engagement, Interest	Emotional responses
Performance Evaluation (Observations)	Task completion	Effectivenss, Efficiency	Perception of instrumental qualities
	Error Rate	Effectiveness, Efficiency	Perception of instrumental qualities

Using the AtrikDiff questionnaire and NASA-TLX, authors in [17], evaluated user's cognitive load (perception of instrumental qualities), AttrakDiff for evaluating the pragmatic and hedonic qualities. The results obtained, were used to bridge the gap between subjective UX assessments and identify the effect of visual feedback and expertise on the user experience. However, no such correlation was found and the authors have suggested to include other techniques for identifying the holistic UX.

Similarly, in [18], the authors have utilized a custom questionnaire to evaluate 2 different MAR applications, in terms of emotional responses and perception of instrumental qualities.

Finally, in [19], the authors have used Smart Glasses User Satisfaction (SGUS) questionnaire, System Usability Scale (SUS), Questionnaire for User Interaction Satisfaction (QUIS), and interviews for evaluating learnability and user satisfaction levels using individual measures. While the results produced by each of these methods is very promising, the final result of the UX evaluation is based entirely on the evaluator's interpretation and not any empirical value.

UX evaluations also take into account the diversity of human attitude and learnability by using cross-sectional (difference in levels of expertise), longitudinal (pre and post usage evaluation), and retrospective reconstruction (recall) approaches[20], [21]. These approaches are important to not only identify the current user experience but also to capture a complete picture of how the user experience changes, with respect to timeliness and to a larger factor the ever evolving familiarity of users with similar systems[22]. Additionally, a long term evaluation also allows the evaluators to identify user preferences and usage contexts which can guide the general direction of product development and/or service delivery[23].

III. METHODOLOGY

In order to evaluate the mixed method approach to evaluating UX in MAR application, we developed a prototype application and applied multiple methods on it. The details are discussed in the following sub-sections.

A. *MAR App Development*

MAR Madness, has been developed on Android, using ARCore version 1.6.0. The compile time Android API version is 28, while the app supports minimum Android API version is 24. The application was tested using a Samsung Galaxy s7 with Android 8.0.0 and running API version 26.

The application produces up to 10000 apples, augmented on the real scene, captured from the main camera. New apples are generated, at random points on 3d view box around the camera, with a fixed rate of 2 seconds. As a player clicks on any apple, it is removed from the screen and the score is updated by 1. The game is played by each user individually and does not require any active internet connection.

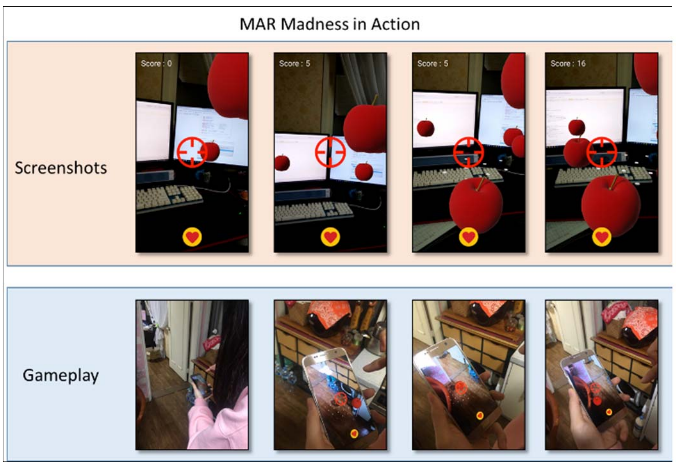


Fig. 3. Prototype MAR app used for UX evaluation process

B. *UX evaluation of the MAR App*

In order to holistic UX we gathered data, in 3 iterations, from 9 participants, before, during, and after usage. At these three stages we utilized different methods and techniques to capture a true picture of the overall UX. The methodology is shown in Fig. 4. Since the scope of this paper is to only evaluate the mechanism for applying multi-method UX evaluation for an MAR application, the context-of-use was simplified. In our case, the users were aged between 20-35, with 8 males and 1 female participant. The users also had ample computer science knowledge, as all of them are pursuing their BS, MS, or Ph.D. studies in Computer Engineering.

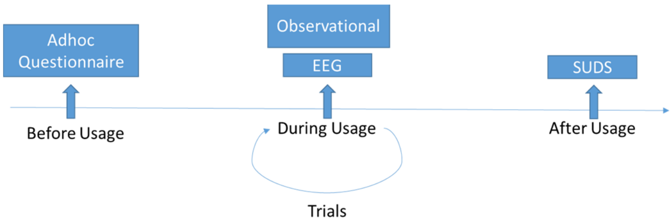


Fig. 4. Holistic UX evaluation methodology

The environment for testing was open, albeit with the constraint of sensor availability, due to which 8 participants utilized the lab environment, while 1 used it at their home. Finally, the system aspect was covered by the android phone, mentioned earlier, while the data was collected through a custom app and uploaded to a central CouchDB server.

At the first stage we used an adhoc questionnaire (shown in Fig. 6.), generated using Google Forms to collect general trends about the participants. This included, question such as their age, gender, familiarity to AR apps, familiarity with mobile AR apps, and motion sickness. The form was distributed before starting the experiment to all participants. This way we were able to collect the anticipated UX of the participants. The results are also shown in Fig.5.

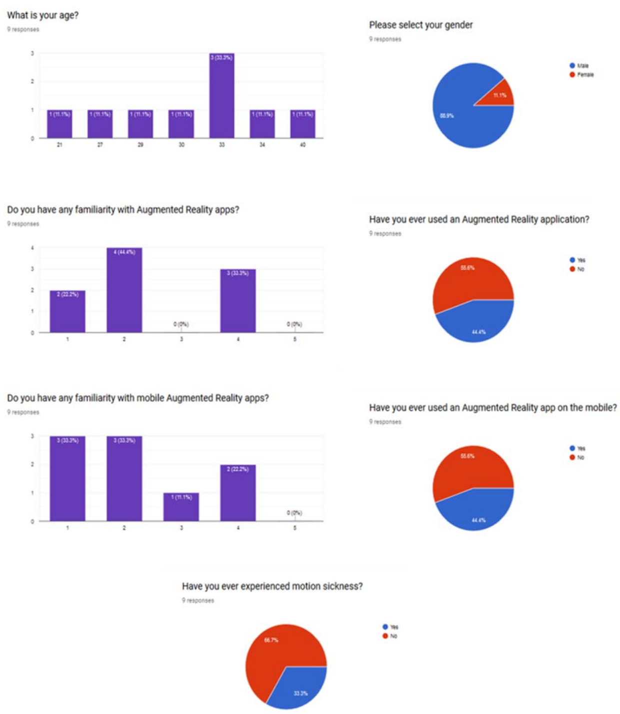


Fig. 5. Anticipated UX collection using adhoc questionnaire

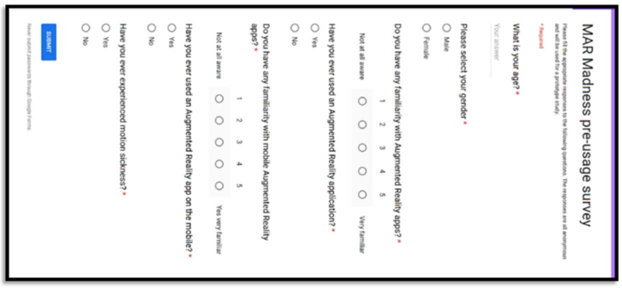
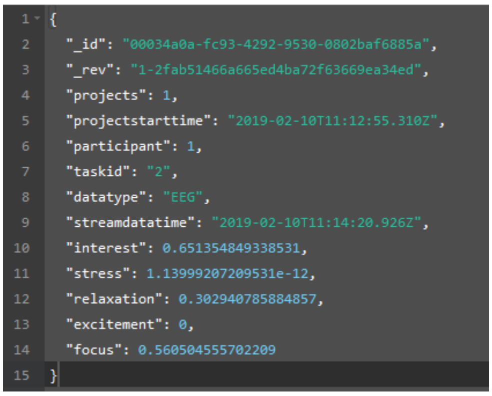


Fig. 6. Adhoc questionnaire for data collection

In the second stage we collected momentary UX data, for completing one task. In this task, the users were asked to hit the apples 10 times, without clicking anywhere else on the screen. Here we observed the users to determine the amount of time spent for task completion. We also collected the user's physiological data by using the EMOTIV EPOC+ EEG headset. The device utilized 14 channels for sensing emotions and cognition of the participants. The data collected through this device was loaded onto a local CouchDB server. A snapshot of sample data is shown in Fig. 6. Each participant passed through 3 trials, enriching the momentary UX. The results obtained after taking average of scaled (converting the values from 0-1 range to 0-100 range) metrics, are shown in Table II.

The excitement values were all found to be 0, most likely due to an incorrect sensor, and have been removed from the results. The remaining results are shown in Fig. 7. These indicate a general decreasing trend in time taken to complete the task and correspondingly a linear increase in learnability



```
{
  "_id": "00034a0a-fc93-4292-9530-0802baf6885a",
  "_rev": "1-2fab51466a665ed4ba72f63669ea34ed",
  "projects": 1,
  "projectstarttime": "2019-02-10T11:12:55.310Z",
  "participant": 1,
  "taskid": "2",
  "datatype": "EEG",
  "streamdatatime": "2019-02-10T11:14:20.926Z",
  "interest": 0.651354849338531,
  "stress": 1.13999207209531e-12,
  "relaxation": 0.302940785884857,
  "excitement": 0,
  "focus": 0.560504555702209
}
```


Fig. 7. EEG data stored in JSON form on a local CouchDB server

TABLE II. MOMENTARY UX RESULTS

Task Time (seconds)	Interest*	Stress*	Relaxation*	Focus*
5.25925926	64.50	1.14E-10	30.29	56.07

*Average of scaled values (0 to 100)

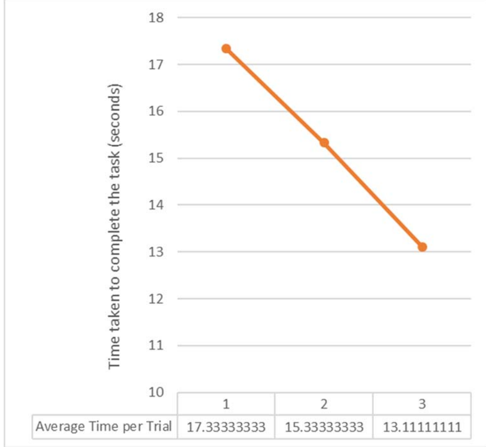
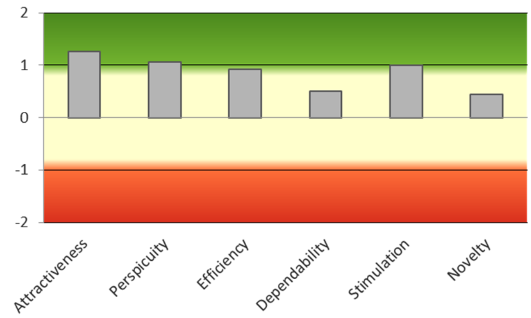


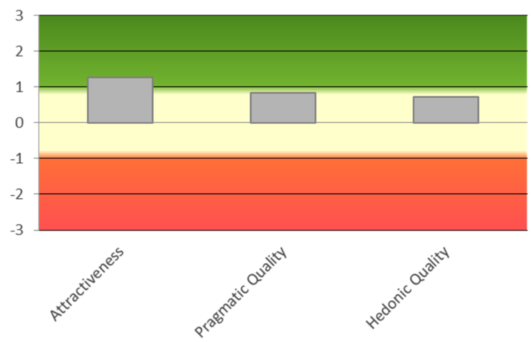
Fig. 8. Momentary UX for 9 participants in 3 Trials

For the third stage, we again utilized 2 questionnaires for evaluating instrumental and non-instrumental qualities. We used the User Experience Questionnaire (UEQ)[24] which contains 6 scales for measuring attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty of the interactive system (in this case our MAR app). The user can provide a score between 0 and 7, for 26 items which are then aggregated for each scale between -3 to +3, indicating negative or a positive response, correspondingly. The graph generated for this data using the UEQ version 4 is shown in Fig. 8. This data shows the results for the 6 scales were slightly above neutral. The same is true for the group-wise aggregated results.

Additionally, we also utilized the SUDs score to evaluate any motion sickness, experience during each trial. The results for this evaluation are shown in Fig. 9.



(a) Results for individual scales calculated by UEQ



(b) Results for aggregated data calculated by UEQ, providing valence, instrumental and non-instrumental qualities

Fig. 9. Results calculated by UEQ version 4

The average response for the 9 participants was 1.777 which indicate an overall comfortable environment with little unpleasantness (only 2 cases of unpleasantness above or equal to 5).

For calculation of Holistic UX, we started by scaling the average values for interest, stress, relaxation, and focus from 0 to 100, to _3 to +3. Similarly, we scaled the average value for motion sickness to also comply with _3 to +3 scale, instead of 0-10 scale. The scaling formula used is described in (1).

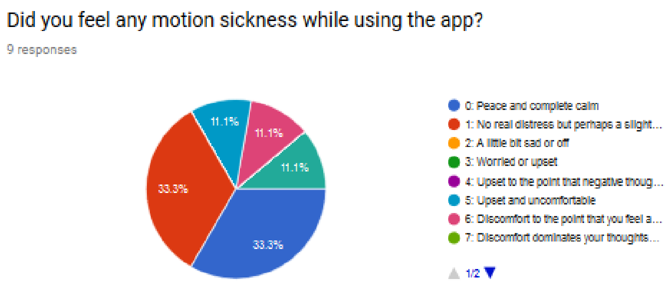


Fig. 10. SUDs result for evaluating motion sickness after usage of the MAR app

The scaled values and their relative weight, in presence of the other metrics is shown in Table III.

$$New_{val} = \left(\frac{((Old_{val} - Old_{min}) * (New_{max} - New_{min}))}{(Old_{max} - Old_{min})}\right) + New_{min} \quad (1)$$

The formula used for calculating relative weights is given in (2).

$$\text{Relative Value} = \frac{\text{Scaled Value}}{\sum \text{Scaled Values}} \quad (2)$$

TABLE III. SCALED AND WEIGHTED VALUES FOR UX METRICS

Scaled Values (-3 to +3)		Relative Values	
Attractiveness (A_i)	0.84	Attractiveness (Aw_i)	0.273295
Perspicuity (P_i)	1.22	Perspicuity (Pw_i)	0.396929
Efficiency (E_i)	2.5	Efficiency (Ew_i)	0.813378
Dependability (D_i)	1.06	Dependability(Dw_i)	0.344872
Stimulation (S_i)	1.34	Stimulation (Sw_i)	0.435971
Novelty (N_i)	1	Novelty (Nw_i)	0.325351
Motion Sickness (Ms_i)	-1.938	Motion Sickness (Ms_w_i)	-0.63053
Interest (In_i)	0.87	Interest (In_w_i)	0.283056
Stress (St_i)	-3.00E+00	Stress (St_w_i)	-9.76E-01
Relaxation (Re_i)	-1.1826	Relaxation (Re_w_i)	-0.38476
Focus (Fo_i)	0.3642	Focus (Fo_w_i)	0.118493

Next, we calculated, the four dimension of UX (Attractiveness, Hedonic Quality, Pragmatic Quality, and Cognition) adapted from [25]. The formulas used to calculate these are shown in (3) – (6).

$$A = Aw_i * A_i \quad (3)$$

$$HQ = (Sw_i * S_i) + (Nw_i * N_i) \quad (4)$$

$$PQ = (Pw_i * P_i) + (Ew_i * E_i) + (Dw_i * D_i) \quad (5)$$

$$C = (Ms\_w_i * Ms_i) + (In_{w_i} * In_i) + (St_{w_i} * St_i) + (Re_{w_i} * Re_i) + (Fo\_w_i * Fo_i) \quad (6)$$

This was followed by taking a sum of all values obtained from (3) – (6), as shown in Table IV. This result was then divided by the number of participants to give a final value between 0-1 of UX Measurement Index (UXMI). The results of these last two operations are shown in Table V.

TABLE IV. UX DIMENSIONS FOR MAR APP EVALUATION

Attractiveness	0.229567933
Hedonic Quality	0.909552316
Pragmatic Quality	2.883263925
Cognition	4.894562858

TABLE V. UX MEASUREMENT INDEX (UXMI) CALCULATION

Sum(A_i + HQ_i + PQ_i + C_i)	8.916947033
UXMI (0 to 1)	0.990771893

IV. CONCLUSION

These results show that our MAR app was able to receive a very high value of 0.99. While these results may be biased due to the small number of participants, they do prove the usefulness of using the UXMI for calculating Holistic UX. Additionally, the multi-method UX evaluation approach is the key for improving the long term acceptance and usage of any MAR application. Finally, the authors believe, this UX evaluation approach is generic enough for evaluating any interactive system, and in future we will look towards proving the same.

ACKNOWLEDGMENT

This research work was supported by Zayed University Cluster Research Fund # R18038. This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-0-01629) supervised by the IITP(Institute for Information & communications Technology Promotion) and NRF-2019R1G1A1011296

REFERENCES

[1] A. Dünser, R. Grasset, H. Seichter, and M. Billinghurst, “Applying HCI Principles to AR Systems Design,” Proc. 2nd Int. Work. Mix. Real. User Interfaces Specif. Authoring, Adapt. (MRUI ’07), no. March 11, pp. 37–42, 2007.

[2] B. Avery, W. Piekarski, J. R. Warren, and B. H. Thomas, “Evaluation of User Satisfaction and Learnability for Outdoor Augmented Reality Gaming BT - Seventh Australasian User Interface Conference (AUIC2006),” vol. 50, no. February 2014, pp. 17–24, 2006.

[3] A. Vovk, F. Wild, W. Guest, and T. Kuula, “Simulator Sickness in Augmented Reality Training Using the Microsoft HoloLens,” pp. 1–9, 2018.

[4] C. Lecon, “Motion Sickness in VR Learning Environments ATINER ’ s Conference Paper Series COM2018- 2514 Motion Sickness in VR Learning Environments,” 4th Int. Conf. Inf. Technol. Comput. Sci., no. August, 2018.

[5] J. Guna, G. Geršak, I. Humar, J. Song, J. Drnovšek, and M. Pogačnik, “Influence of video content type on users’ virtual reality sickness perception and physiological response,” Futur. Gener. Comput. Syst., vol. 91, pp. 263–276, 2019.

[6] I. 9241-11:2018, Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts, vol. 2018. ISO, Geneva, Switzerland, 2018.

[7] I. 9241-210:2010, “Ergonomics of human–system interaction — Part 210: Human-centred design for interactive systems,” vol. 2010, ISO, Geneva, Switzerland, 2010.

[8] E. L.-C. Law, P. van Schaik, and V. Roto, “Attitudes towards User Experience (UX) Measurement Effie,” Int. J. Hum. Comput. Stud., vol. 72, no. 6, pp. 526–541, 2014.

[9] B. Fogg, “A behavior model for persuasive design,” Proc. 4th Int. Conf. Persuas. Technol. - Persuas. ’09, p. 1, 2009.

[10] H. Xue and P. M. A. Desmet, “Researcher introspection for experience-driven design research,” Des. Stud., vol. 63, pp. 37–64, 2019.

[11] J. Hussain et al., “A multimodal deep log-based user experience (UX) platform for UX evaluation,” Sensors (Switzerland), vol. 18, no. 5, pp. 1–31, 2018.

[12] S. Irshad and D. R. B. A. Rambli, “User experience of mobile augmented reality: A review of studies,” Proc. - 2014 3rd Int. Conf. User Sci. Eng. Exp. Eng. Engag. i-USEr 2014, no. September, pp. 125–130, 2015.

[13] M. Pallot and K. Pawar, “A holistic model of user experience for living lab experiential design,” 2012 18th Int. Conf. Eng. Technol. Innov. ICE 2012 - Conf. Proc., 2012.

[14] D. Seminar and D. U. Experience, “Coffroth 1985-190.pdf,” 2011.

[15] Y. Arifin, T. G. Sastria, and E. Barlian, “User Experience Metric for Augmented Reality Application: A Review,” Procedia Comput. Sci., vol. 135, pp. 648–656, 2018.

[16] E. Loup-Escande, E. Jamet, M. Ragot, S. Erhel, and N. Michinov, “Effects of Stereoscopic Display on Learning and User Experience in an Educational Virtual Environment,” Int. J. Hum. Comput. Interact., vol. 33, no. 2, pp. 115–122, 2017.

[17] E. Loup-Escande, R. Frenoy, G. Poplimont, I. Thouvenin, O. Gapenne, and O. Megalakaki, “Contributions of mixed reality in a calligraphy learning task: Effects of supplementary visual feedback and expertise on cognitive load, user experience and gestural performance,” Comput. Human Behav., vol. 75, pp. 42–49, 2017.

[18] A. Dirin and T. Laine, “User Experience in Mobile Augmented Reality: Emotions, Challenges, Opportunities and Best Practices,” Computers, vol. 7, no. 2, p. 33, 2018.

[19] K. Helin, T. Kuula, C. Vizzi, J. Karjalainen, and A. Vovk, “User Experience of Augmented Reality System for Astronaut’s Manual Work Support,” Front. Robot. AI, vol. 5, no. September, pp. 1–10, 2018.

[20] E. Karapanos, J. B. Martens, and M. Hassenzahl, “Reconstructing experiences with iScale,” Int. J. Hum. Comput. Stud., vol. 70, no. 11, pp. 849–865, 2012.

[21] E. Karapanos, J.-B. Martens, and M. Hassenzahl, “On the retrospective assessment of users’ experiences over time,” p. 4075, 2010.

[22] E. Karapanos, Marc Hassenzahl, and J.-B. Martens, “User Experience Over Time,” p. 31000, 2013.

[23] B. Yang, Y. Liu, Y. Liang, and M. Tang, “Exploiting user experience from online customer reviews for product design,” Int. J. Inf. Manage., vol. 46, no. May 2018, pp. 173–186, 2019.

[24] U. E. Questionnaire, “User Experience Questionnaire,” pp. 3–5, 2013.

[25] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, and J. Thomaschewski, “Developing a UX KPI based on the user experience questionnaire,” Comput. Stand. Interfaces, no. January, 2019.