Human Factors and Systems Interaction, Vol. 154, 2024, 257–267
https://doi.org/10.54941/ahfe1005373

Usability Assessment of Extended Reality Applications. A Review

Mário Simões-Marques

CINAV, Escola Naval, Instituto Universitário Militar, Base Naval de Lisboa, 2810–001 Almada, Portugal

ABSTRACT

Extended Reality (a continuum that encompasses Virtual Reality, Mixed Reality, and Augmented Reality) is a recent concept that is gaining traction as new concepts of Cyber-Physical Systems are being researched and developed exploiting and integrating different modes of interaction (e.g., visual, language, audio, haptic). The new interface technologies find application in a host of fields, such as education and training, healthcare, security and defense, engineering and maintenance, and entertainment. Extended Reality usability assessment faces new challenges, considering the potential negative impacts (e.g., sickness, discomfort, and cognitive load) of using an immersive environment, and the need for strategies to avoid or, at least, mitigate such impacts. This paper reviews the state-of-the-art of Usability assessment methods applicable to the Extended Reality spectrum, categorizing them and identifying gaps to be bridged in the future.

Keywords: Cyber-physical interaction, User experience, Virtual reality, Augmented reality, Mixed reality

INTRODUCTION

Usability and User Experience (UX) are quite well-consolidated concepts that are being applied to Cyber-Physical Systems (CPS), which led to the development of a variety of Ergonomics/Human Factors focused design heuristics and frameworks and assessment methods. The technological evolution is turning real new and emergent forms of immersive and semi-immersive interaction with CPS. On one hand, this is allowing for the actual implementation of futuristic visions that were introduced in the context of Science-Fiction some decades ago, such as the "metaverse" (note that one of the key features of the Metaverse is being Immersive, in the sense of feeling real to the users (ART, 2022)). On the other hand, new forms of interaction require improved or even new ways of ensuring good and safe usage of CPS.

Extended Reality (XR) (i.e., a continuum that encompasses Virtual Reality (VR), Mixed Reality (MR), and Augmented Reality(AR)) is a recent concept (Cárdenas-Robledo et al., 2022; Rauschnabel et al., 2022) that is gaining traction as Gaming, Modelling and Simulation, Internet of Things, Industry 4.0, Robotics, Digital Twins or Serious Games domains, to name a few, are being researched and developed exploiting and integrating different interface device technologies (e.g., visual, language, audio, haptic). These interface technologies find application on a host of fields, such as education and training, medicine and healthcare, security and defense, engineering and maintenance, and entertainment. Extended Reality interaction places new challenges to researchers and designers, considering the growing complexity of the CPS' user environment. Among others is the impact of such forms of interaction on humans (e.g., the risk of creating discomfort, or physiological and cognitive load) and the methods for assessing them and the strategies for avoiding or, at least, mitigating potential negative impacts.

Therefore, the research questions addressed by this work are:

> *What are the Usability assessment methods being applied in the context of Extended Reality spectrum? and What features are they assessing?*

To answer this question, a review of papers published in recent years addressing Usability assessment studies performed to Extended Reality designed applications, identifying the methods used and the features assessed. The Preferred Reporting Items for Systematic Reviews (PRISMA) method (Liberati et al., 2009) is used as a formal tool and guideline for data collection of the literature review. The chosen data set consists of studies published in the last twenty years. The analyzed sample includes papers from journals searched on relevant scientific databases and meeting specific query constraints. For the retrieved papers, the methods used and the type of Usability assessment features they cover were identified. Previous published systematic reviews and surveys on Usability assessment methods were out of the scope of the current work.

This paper is structured as follows: the Method section presents the search performed, followed by the data collection and the analysis processes to extract relevant information; the protocol followed is instantiated in the Results section, including the chosen bibliographic repositories, the records' inclusion and exclusion criteria, as well as the search and analysis processes for collected papers. The Discussion section highlights the topics of Usability assessment methods applicable to the Extended Reality spectrum which are the target of the research questions, categorizing them. Finally, the Conclusion section provides a summary of the findings and presents proposals for further work.

METHOD

To follow the systematic review procedures of PRISMA (Page et al., 2021), a protocol setup is required in advance, detailing the chosen bibliographic repositories, the inclusion and exclusion criteria for the records, as well as the search and analysis processes of the collected papers.

For attaining relevant references, the main bibliographic repositories were searched (e.g., Complementary Index, Emerald Insight, IEEE Xplore Digital Library, Academic Search Complete, Directory of Open Access Journals, Business Source Complete, MEDLINE) using the EBSCO search engine. The protocol specified several retrieval constraints: i) peer-reviewed journal articles - excluding any other kind of documents (e.g., book chapters, conference proceedings, technical reports); ii) works published in English - excluding any other languages; and iii) time span: from 2004 to 2024.

A query string ("Extended Reality" AND "Usability OR User Experience OR UX" NOT "Survey" NOT "Review") was built with a composition of terms and Boolean operators chosen for finding (by title, abstract, or keywords) relevant English written peer-reviewed journal articles.

The process of extraction of the records was performed on 2024, January 4th, and a spreadsheet with the core fields for each of the retrieved records (title, abstract, keywords, authors' names and affiliations, journal name, and year of publication) was created, allowing a preliminary screening and assessing of records' titles and abstracts. Subsequently, a validation of records' compliance with the eligibility criteria was performed to articles referring to Usability assessment studies performed to Extended Reality designed applications, excluding articles not directly related with usability studies (e.g., reviews, surveys) or lacking information about the specific methods applied and demonstrated by study results.

Bibliographic details of the included studies were added to the spreadsheet as well as the essential items of the PRISMA checklist. However, given the exploratory nature of this work, PRISMA's items 12 to 27 were ignored. Finally, a pilot test on fifty randomly-selected papers was conducted in order to refine and code the extracted information, namely the assessment methods applied, the specific context of the Extended Reality spectrum analyzed, and the Usability features assessed by each method. This paper reflects the results of the review considering these stages. The review work is currently proceeding, encompassing the remaining records resulting from the bibliographic repositories search.

RESULTS

The search on the above-mentioned electronic databases allowed the retrieval, in total, of 1475 bibliographic records. By applying the constraints the number of records was reduced by considering the time span (1434), peer reviewed (650), English writing (638), scientific journals (524), and duplications (494). These filtered records corresponded to 777 records on the searched source repositories. Considering the journal editors, the most represented are: MDPI (91), IEEE (71), Taylor & Francis (52), Springer (41), and Wiley (26). The distribution of retrieved records per year is illustrated in Figure 1, which also offers the perspective that most publications meeting the constraints set for the review were published in the last five years. In fact, 75% of the records were published in the last three years: 90 records in 2021 (18%), 142 records in 2022 (29%), and 139 records in 2023 (28%). This is an evidence of the novelty of the topic.

The retrieved records, are currently being screened based on contents, to check if they meet the eligibility criteria of describing usability studies which contain a clear reference to the specific methods applied. Preliminary data, based on a random access to texts available online, points to an eligibility rate close to 40% (85 papers selected out of 208 papers screened). The selected papers follow a publication year distribution that reflects the distribution of the retrieved papers (i.e., 2019-2; 2020-8; 2021-13; 2022-21; 2023-38; 2024-3).

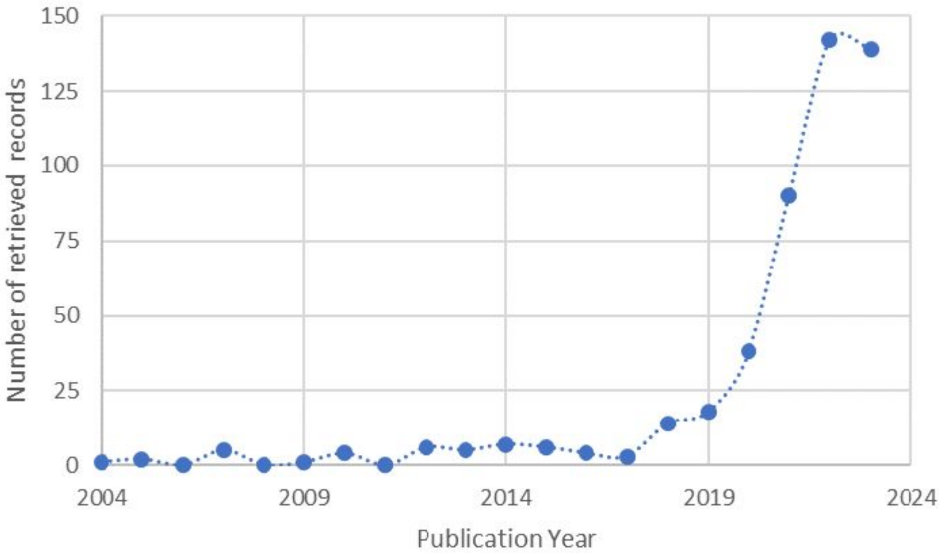

Figure 1: Distribution of the number of retrieved records per year.

The eligible papers were treated and its contents subject to a coding process following their full reading. The relevant articles were codified according to predefined categories: (i) type of Extended Reality application (e.g., VR, AR, MR, XR); (ii) type of interaction device (e.g., desktop, HMD (head-mounted display), handheld mobile device, CAVE (cave automatic virtual environment)); (iii) specific assessment methods adopted; (iv) usability feature addressed; (v) type of application domain.

The elicited data was compiled in a table, which is summarized in this article for reasons of space. The preliminary findings are presented in the next section.

DISCUSSION

Table 1 summarizes the findings of the study, regarding the assessment methods described in the sample of papers there were retrieved from the literature review. The methods are first sorted in descending by frequency of occurrence, and second by alphabetical order of those with identical frequency. The first column ('Type of assessed feature or Data Collection Method') presents no particular sorting criteria.

Note that most of the analyzed usability studies (which are not identified in this paper) combine multiple assessment methods; identically, some compare multiple platforms or several types of XR in the same study. This is the reason why the sum of the values presented in the column (or row) cells does not match the 'total' column (or row) values.

Regarding the XR type, it was found a prevalence of studies addressing VR apps (59), followed by AR (28), MR (7), and just four addressing XR in general. As for the type of interaction devices, the large majority of the apps assessed used HMD (77%), followed by handheld mobile (14%) and desktop (7%) devices; two studies addressed the use of CAVEs. Regarding the application domain, the usability studies addressing Medicine and Healthcare (including the education and training of professionals) were the most frequent in this sample (36%), followed by Education and Training (28%), and Interaction (23%) usability studies.

Regarding the type of data collection methods, the usability studies analyzed tend to be mainly based on qualitative (e.g., questionnaires) rather than on quantitative (e.g., measurements) types of assessment methods. Regarding the assessed features, a significant portion of the assessment method types identified address features that go beyond the most commonly assessed in Usability/UX (i.e., learnability, efficiency, memorability, errors, satisfaction); these features tend to relate to major issues that may affect the use of XR apps, such as Sickness, Presence, Anxiety and Simulation Effectiveness. Observation and Perceived Workload methods were also used in some of the analyzed studies.

The following subsections identify the specific methods used in the above mentioned publications.

Qualitative Assessment Methods

Virtually all papers analyzed used one or a combination of qualitative assessment methods to evaluate the Usability or User Experience of the application(s) under study. These methods are usually based on questionnaires, often using a Likert Scale (Likert, 1932) to classify the answers. Some methods are holistic in assessing the multiple dimensions of Usability, while others focus on a particular dimension.

General Usability/UX assessment – the methods used more frequently in the studies were the System Usability Scale (SUS) (Brooke, 2013) and the User Experience Questionnaire (UEQ) (Laugwitz et al., 2008), followed by the Post-Study System Usability Questionnaire (PSSUQ) and its non-laboratory version the Computer System Usability Questionnaire (CSUQ) (Lewis, 2002); the Usability Scale for Handheld Augmented Reality (HARUS) (Santos et al., 2014) was used to assessed handheld AR applications; while the Game Engagement Questionnaire (GEQ) (Brockmyer et al., 2009) and the Game User Experience Satisfaction Scale (GUESS) (Phan et al., 2016) were used to assessed [serious] game applications; Nielsen Heuristics (Nielsen & Molich, 1990) and the USE Questionnaire (Lund, 2001) were also used in more than one studies. Finally, the following methods were applied in only one of the analysed studies: Expectation Measure and Ease of Use (Davis, 1989); Learner-User eXperience Questionnaire QLUX (Assaf & Morán-Mirabal, 2023); Media and Technology Usage and Attitudes Scale (MTUAS) (Rosen et al., 2013); Extended System Usability Scale (SUS-E) (Harper & Dorton, 2021); Instructional Usability Scale (SUsI) (Assaf & Morán-Mirabal, 2023); and Technology Usage Inventory (TUI) (Kothgassner et al., 2013).

Sickness assessment – sickness is an undesired effect of the exposure to simulators and XR environments. The following three methods, sorted by decreasing frequency, were found in the analyzed papers: Simulator Sickness

Questionnaire (SSQ) (Kennedy et al., 1993); Virtual Reality Sickness Questionnaire (VRSQ) (Kim et al., 2018); and Visually Induced Motion Sickness Susceptibility Questionnaire (VIMSSQ) (Keshavarz et al., 2019).

Presence assessment – the level of presence felt by users immersed in an XR environment is an important feature for the success of an app. Some studies assessed this feature using the following methods: Presence Questionnaire 2.0 (Witmer & Singer, 1998); I-group Presence Questionnaire (IPQ) (Berkman & Çatak, 2021); Slater-Usoh-Steed presence questionnaire (SUSpq) (Usoh et al., 2000); Augmented Reality Immersion Questionnaire (ARI) (Georgiou & Kyza, 2017); Presence in Augmented Reality Questionnaire (pAR) (Regenbrecht & Schubert, 2002); and Self-Presence Questionnaire (SPQ) (Ratan & Hasler, 2009).

Acceptability assessment – another important feature for any application is the degree of users' acceptance. To assess this, the following methods were used in the studies: Technology Acceptance Model 3 (TAM-3) (Venkatesh & Bala, 2008); Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003); and Service User Technology Acceptability Questionnaire (SUTAQ) (Hirani et al., 2017).

Anxiety assessment – the motivation to use an app can be affected by the level of anxiety its use may generate. To assess the Flow Short Scale (FSS) (Rheinberg et al., 2003) and the State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1970) two methods were applied in the studies analyzed.

Satisfaction assessment – users' satisfaction is an important Usability feature of an application. The specific assessment of this was done using the User Satisfaction Evaluation Questionnaire (USEQ) (Gil-Gómez et al., 2017) and the Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST) (Demers et al., 2000) methods.

Emotions assessment – one particular study used the Achievement Emotions Questionnaire (AEQ) (Paoloni et al., 2014) to assess the emotions experienced by students using an AR app.

Desirability assessment – a couple of papers assessed the desirability of AR apps by using the Product Reaction Card (Benedek & Miner, 2002) method.

Consumer Loyalty assessment – the Net Promoter Score (Keiningham et al., 2008) was used in three studies addressing the use of VR and AR apps.

Simulation Effectiveness assessment – The Simulation Effectiveness Tool - Modified (SET-M) (Leighton et al., 2015) was used in two different studies to assess AR apps.

Ease of task completion assessment – it was done in three studies using the After-Scenario Questionnaire (ASQ) (Lewis, 1991).

Interviews – this is frequently used to assess the Usability of apps and find areas for improvement. Interviews can be **semi-structured**, usually involving open ended questions; or **structured**, often using Likert Scale answers, and processed statistically. Some studies tailor a customized or adapted assessment method based on this type of interviews/questionnaires.

Observation – Usability tests usually are monitored and can be recorded. The information gathered in the test sessions is very helpful for the assessment of the application. The analyzed studies that embraced this approach used the following well-known methods: Think-aloud (Lewis, 1982) and Cognitive Walkthrough (Wharton et al., 1994). The Wizard of Oz (Kelley, 1983) method was also used to simulate the expected behavior of VR applications, assessing them in an early stage of the design.

Perceived Workload – the perceived workload is a very important feature of applications meant to be used to perform demanding or complex tasks. The methods more frequently used were the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988), and the Raw Task Load Index (RTLX) (Hart, 2006). The usability study of a MR application designed for use in the surgical domain applied the SURG-TLX (Wilson et al., 2011) method.

Quantitative Assessment Methods

A much more reduced fraction of the studies (16 in total), applied quantitative usability assessment methods in combination with the above mentioned qualitative ones. In the analyzed papers the data collection was supported by:

Sensors – that monitored the users while they were performing the tasks. Namely, two studies used **Eye tracking** to locate the gaze of the users, identifying the areas where attention was focused; and one study monitored the **Heart Rate** of patients involved in training activities supported by a MR application.

Measurements – for evaluating the **Time** spent, the number of **Errors**, and the **Movement Patterns** while using an app. A study also assessed the 3D Intersection Over Union (**IoU**) for comparing AR and VR in the representation of real objects.

CONCLUSION

The analysis of the eligible papers retrieved contributed to answering the research questions addressed in this work, identifying a number of Usability assessment methods that are being applied in the context of Extended Reality, as well as the features they are assessing. The results were summarized in Table 1 and the methods were addressed in the Discussion section. The distribution of publications per year identified in the bibliographic search reveals that about 75% of the records concentrated in the last three years, being an evidence of the traction the discussion of XR usability assessment is gaining. Apparently there is still a large prevalence of studies addressing VR apps, compared with AR and MR apps. The interaction devices targeted by the studies were mainly HMD. The application domains were mainly related with Medicine & Healthcare, Education & Training, and Interaction usability studies. Regarding the Usability/UX assessment there is a prevalence of Qualitative methods when compared with Quantitative methods. Some traditional methods (e.g., SUS, UEQ) are intensively used in the XR context, but new methods are emerging to assess specific types of interaction (e.g., HARUS), or to assess particular types of features or impacts that are more specific to XR applications (e.g., Sickness, Presence, Anxiety, Simulation Effectiveness). Regarding the emerging methods it seems to be important, in the near future, to clarify their limitations and advantages in order to evolve to a reduced set of robust and comprehensive XR usability assessment tools that can be applied in a standardized and comparable way, contributing to improve the design of XR applications and also to allow the comparison of their Usability.

Table 1. Summary of assessment methods used in the random sample of the literature review retrieved studies.

Type of Assessed Feature or Data Collection Method	Specific Method	Number of studies	Type of XR				Interaction device type				App Domain						
			VR	AR	MR	XR	Head Mounted Display	Handheld Mobile	Desktop	CAVE	Medicine/Healthcare	Education/Training	Interaction	Entertainment/Gaming	Industry	Tourism & Exploration	Analytics
Total selected after screening		85	58	28	7	4	74	14	7	2	31	24	20	3	3	2	2
General Usability/ UX	SUS	50	29	18	3	2	39	11	5	1	16	13	15	2	2	2	
	UEQ	12	8	3	1		10	2	2	1	6	4	2				
	PSSUQ/CSUQ	4	3	1			3	1				3			1		
	HARUS	3		2	1		1	2				1	1			1	
	GEQ	2	1	1			2		1	1		1		1			
	GUESS	2	2				2				1	1					
	Nielsen Heur.	2		2			2					2					
	USE Questionnaire	2	2				2					2					
	Expect. Measure and Ease of Use	1	1				1						1				
	QLUXQ	1	1						1			1					
	MTUAS	1	1				1				1						
	SUS-E	1	1				1					1					
	SUsI	1	1						1			1					
	TUI	1	1				1				1						
Sickness	SSQ	10	10	1			1				4	2	4				
	VRSQ	3	3				3				1	1	1				
	VIMSSQ	1	1	1		1	1	1				1					
Presence	Presence Quest. 2.0	6	6				6				4		2				
	IPQ	4	4				3		1	1	1	2	1				
	SUSpq	4	4				4		1		2	2					
	ARI	1			1		1									1	
	pAR	1		1			1				1						
	SPQ	1	1							1		1					
Acceptability	TAM-3	3	3				3				2	1					
	UTAUT	2	2				2				1		1				
	SUTAQ	1	1				1				1						
Anxiety	FSS	1		1			1				1						
	STAI	1			1		1										1
Satisfaction	USEQ	2	2				2				1			1			
	QUEST	1		1			1				1						
Emotions	AEQ	1		1				1				1					
Desirability	Prod.Reaction Card	2		2			1	2					1			1	
Consumer Loyalty	Net Promoter Score	3	1	2			2	2	1				2			1	
Simul. Effectiveness	SET-M	2		2			2				1	1					
Ease task completion	ASQ	3	3				3		1		1		2				
Interviews	Semi-structured	22	12	10	2	1	18	5	3		7	7	5	1		1	1
	Structured	10	9	4	2	2	10	2			5	2	2				1
Observation	Think-aloud	6	3	2	1		4	2	2		1	2	2			1	
	Cognit.Walkthrough	2	1		1						1	1					
	Wizard of Oz	2	2				2		1		1		1				
Perceived Workload	NASA TLX	12	8	5			12		1		2	3	4		3		
	RTLX	2	1	1			2	1			2						
	SURG-TLX	1			1		1				1						
Sensors	Eye tracking	2		1	1		1	1					1			1	
	Heart Rate	1	1				1				1						
Measurements	Time	14	11	3	3		13	2	2		4	2	7				1
	Errors	9	7	1	1		8	2	2		3	2	3				1
	Movement Patterns	4	3	1	1		4					1	2				1
	IoU	1	1	1	1		1						1				

ACKNOWLEDGMENT

The author would like to acknowledge the Portuguese Navy for funding this work.

REFERENCES

ART (2022). Metaverse - Virtual World, Real Challenges. Brussels: Council of European Union. General Secretariat. Analysis and Research Team.

Assaf, N. and Morán-Mirabal, L. F. (2023). Instructional Usability and Learner-User eXperience Assessment in a Virtual Reality Educational Milieu: A Deductive Tech-Instructionality Model from EdTech, 2023 Future of Educational Innovation-Workshop Series Data in Action, Monterrey, Mexico, 1–8.

Benedek, J. and Miner, T. (2002). Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting. Proceedings of Usability Professionals Association, 8–12.

Berkman, M. I. and Çatak G. (2021). I-group Presence Questionnaire: Psychometrically Revised English Version, Mugla Journal of Science and Technology, 7(0), 1–10.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. J. Exp. Soc. Psychol., 45, 624–634.

Brooke, J. (2013). SUS: A Retrospective. J. Usability Stud., 8, 29–40.

Cárdenas-Robledo, L. A., Hernández-Uribe, O., Reta, C. and Cantoral-Ceballos, J. A. (2022). Extended reality applications in Industry 4.0. – A systematic literature review. Telematics and Informatics, 73(101863), 1–39.

Demers, L., Weiss-Lambrou, R. and Ska, B. (2000). Item analysis of the Quebec user evaluation of satisfaction with Assistive Technology (QUEST). Assist Technol., 12(2), 96–105

Georgiou, Y. and Kyza, E. A. (2017). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. Int J Hum Comput Stud, 98, 24–37.

Gil-Gómez J. A., Manzano-Hernández P., Albiol-Pérez S., Aula-Valero C., Gil-Gómez H., and Lozano-Quilis J. A. (2017). USEQ: A short questionnaire for satisfaction evaluation of virtual rehabilitation systems, Sensors (Switzerland), vol. 17, no. 7, pp. 1–12.

Harper, S. B. and Dorton, S. L. (2021). A Pilot Study on Extending the SUS Survey: Early Results. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65(1), 447–451.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. Proc Hum Factors Ergon Soc Annual Meeting 50(9), 904–908.

Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Hancock, Peter A.; Meshkati, Najmedin (eds.). Human Mental Workload. Advances in Psychology. Vol. 52. Amsterdam: North Holland. 139–183.

Hirani, S. P., Rixon, L., Beynon, M., Cartwright, M., Cleanthous, S., Selva, A., Sanders, C. and Newman, S. P. (2017). Quantifying beliefs regarding telehealth: Development of the Whole Systems Demonstrator Service User Technology Acceptability Questionnaire. J. Telemed. Telecare, 23, 460–469.

Keiningham, T., Aksoy, L., Cooil, B., Andreassen, T. and Williams, L. (2008). A Holistic Examination of Net Promoter. J. Database Mark. Cust. Strategy Manag., 15, 79–90.

Kelley, J. F. (1983). An empirical methodology for writing user-friendly natural language computer applications. Proceedings of ACM SIG-CHI '83 Human Factors in Computing systems, New York, ACM, 193–196.

Kennedy, R. S., Lane, N. E., Berbaum, K. S. and Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. The International Journal of Aviation Psychology, 3(3), 203–220.

Keshavarz, B., Saryazdi, R., Campos, J. L. and Golding, J. F. (2019). Introducing the VIMSSQ: Measuring susceptibility to visually induced motion sickness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2019, SAGE Publications: Los Angeles, CA, USA, 63, 2267–2271.

Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual Reality Sickness Questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Applied Ergonomics, 69, 66–73.

Kothgassner, O., Felnhofer, A., and Hauk, N. (2013). Technology Usage Inventory (TUI): Manual.

Laugwitz, B., Schrepp, M. and Held, T. (2008). Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (Ed.): USAB 2008, LNCS 5298, 63–76.

Leighton, K., Ravert, P., Mudra, V. and Macintosh, C. (2015). Updating the Simulation Effectiveness Tool: Item Modifications and reevaluation of Psychometric Properties. Nurs Educ Perspect., 36(5), 317–323.

Lewis, C. H. (1982). Using the "Thinking Aloud" Method In Cognitive Interface Design (Technical report). IBM. RC-9265.

Lewis, J. (1991). Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. SIGCHI Bull., 23, 78–81.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. International Journal of Human-Computer Interaction, 14(3–4), 463–488.

Liberati, A. et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration, PLoS Medicine.

Likert, Rensis (1932). A Technique for the Measurement of Attitudes. Archives of Psychology, 140, 1–55.

Lund, A. M. (2001). Measuring usability with the use questionnaire. Usability Interface, 8(2), 3–6.

Nielsen, J.; Molich, R. (1990) Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in computing Systems Empowering People—CHI '90, Seattle, WA, USA.

Page, M. J. et al. (2021) PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews, BMJ. BMJ Publishing Group, 372.

Paoloni P., Vaja A., and Muñoz V. (2014). Reliability and validity of the "achievement emotions questionnaire": A study of argentinean university students, electronic, J. Res. Educ. Psychol. 12, 671–692.

Phan, M. H., Keebler, J. R., and Chaparro, B. S. (2016). The development and validation of the game user experience satisfaction scale (GUESS). Human factors, 58(8), 1217–1247.

Ratan, R. A. and Hasler, B. S. (2009). Self-Presence Standardized: Introducing the Self-Presence Questionnaire (SPQ). Proceedings of the 12th annual international workshop on presence (Vol. 81).

Rauschnabel, P. A. et al. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality. In: Computers in Human Behavior. Vol. 133: Elsevier Ltd, pp. 1–18.

Regenbrecht H, Schubert T. (2002). Measuring Presence in Augmented Reality Environments: Design and a First Test of a Questionnaire. Proceedings of the Fifth Annual International Workshop Presence, 138–144.

Rheinberg, F., Vollmeyer, R. and Engeser, S. (2003). Die Erfassung des Flow-Erlebens [The Assessment of Flow Experience]. In J. Stiensmeier-Pelster, & F. Rheinberg (Eds.), Diagnostik von Selbstkonzept, Lernmotivation und Selbstregulation [Diagnosis of Motivation and Self-Concept]. Göttingen: Hogrefe.

Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A. and Rokkum, J. (2013). The Media and Technology Usage and Attitudes Scale: An Empirical Investigation. Comput. Hum. Behav., 29, 2501–2511.

Santos, M. E. C., Taketomi, T., Sandor, C., Polvi, J., Yamamoto, G. and Kato H. (2014). A usability scale for handheld augmented reality. In Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (VRST '14). Association for Computing Machinery, New York, NY, USA, 167–176.

Spielberger, C. D., Gorsuch R. L. and Lushene R. E. (1970). "Stai" Manual for the State-Trait Anxiety Inventory (Self Evaluation Questionnaire). Palo Alto California: Consulting Psychologist, vol. 22, 1–24.

Usoh, M., Catena, E., Arman, S. and Slater, M. (2000). Using presence questionnaires in reality. Presence, 9(5), 497–503.

Venkatesh, V. and Bala H. (2008). Technology acceptance model 3 and a research agenda on interventions. Decis Sci, 39, 273–315.

Venkatesh, V., Morris, M. G., Davis, G. B. and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS quarterly (2003), 425–478.

Wharton, C., Riemann, J., Lewis, C. and Poison, P. (1994). "The cognitive walkthrough method: a practitioner's guide". In Nielsen, Jakob; Mack, Robert L. (eds.). Usability inspection methods. John Wiley & Sons, 105–140.

Wilson M. R., Poolton J. M., Malhotra N., Ngo K., Bright E. and Masters R. S. (2011) Development and validation of a surgical workload measure: The surgery task load index (SURG-TLX). World J Surg., 35(9), 1961–1969.

Witmer B. G. and Singer M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence. Presence, 7(3), 225–240.