MO-XR: Method for Observing User Interactions with XR Applications.

Natalia Adamczewska*
Bournemouth University,
Bournemouth, UK
nbadamczewska@
bournemouth.ac.uk
*Corresponding author

Wen Tang
Bournemouth University,
Bournemouth, UK
wtang@bournemouth.ac.uk

Zequn Li
Bournemouth University,
Bournemouth, UK
zli1@bournemouth.ac.uk

Ethan Southall
Bournemouth University,
Bournemouth, UK
southalle@bournemouth.ac.uk

Timothy Develin
Bournemouth University,
Bournemouth, UK
tdevlin@bournemouth.ac.uk

Alberto De Caro
Bournemouth University,
Bournemouth, UK
adecaro@bournemouth.ac.uk

***Abstract*— One of the essential stages of an extended reality (XR) product development is evaluation. It allows researchers assess whether the product meets the requirements of the user and whether it behaves as it was expected by its creators.**

Many different evaluation methods for XR product development are available to record, describe, quantify, and analyze user experience (UX), with observations being one of them. Observation methods are a flexible tool to perform evaluations, yet they are often poorly applied, lacking description of data collection and data analysis approaches. This paper addresses that issue and proposes a novel approach to performing observation study for extended reality (XR) product development. The proposed method is clearly structured which allows researchers to perform observations in a systematic way to evaluate XR systems. An observation tool is developed and applied to the evaluation for a meditation VR app that we have developed as part of the ASPIRE project. In this paper, the findings of our study in which the tool was applied and explored with survey data are provided. The results show that insights gained the tool enrich understandings of the survey data, which allows for more in-depth evaluation of an XR application.

Keywords*— *user experience (UX), extended reality (XR), virtual reality, evaluation, observations.

This project is funded by the European Development Fund. The authors wish to thank the participants and the Adding to Social capital and individual Potential In disadvantaged REgions (Project)..

I. Introduction

Accurate, consistent and reliable data collection and analysis procedures are essential for enabling researchers to make informed decisions in the process of XR product design, development, and evaluation. A common evaluation method in XR user experience (UX) field is observation [1]. Recording of individual's behavior by independent observers is the hallmark of behavioral observations [2]. Observational techniques are methods by which researchers gather firsthand data on a phenomenon under study [3]. They provide evaluators with an opportunity to collect data on a wide range of behaviors, by capturing a great variety of interactions, and openly exploring the evaluation topic [4].

Direct observation is a method frequently applied in usability testing in which a trained observer records how users perform a scenario of tasks [5]. Observations can be performed at various stages of an XR product development process, such as during early phase design, or to evaluate a fully functioning product. Regardless the stage of the product development that observation techniques are applied, the technique allows researchers to make informed decisions on the development of their XR systems, or whether its design meets their objectives.

Observations are highly flexible and are influenced by several factors, such as the aim of the study and the purpose for making observations. Despite the value of the observational method and the frequency of its application, observations are often poorly executed. Edwards, Hubert & Kramarova [6] previously pointed out that even though UX researchers choose a right method, they do not necessarily have the skills to execute it correctly. Manually observing user behaviors is often related to several difficulties in data collection, including biased and poorly trained observers [7]. Furthermore, unstructured observations with limited descriptive information on the process of data collection and analysis may compromise the validity and reliability of research findings. Since researchers are subject to confirmation bias in their research, they seek for the evidence that supports their claims [8], and may pay less attention to data that is contradictory to their assumptions, and therefore omit it [9]. Most certainly, creators of an XR product would like to view it in a positive way which may result in overlooking, intentionally or not, negative aspects of its design and implementation details. This issue is even more severe when a study lacks a systematic approach to the product evaluation methodologies, with compromised quality and transparency of data collection, analysis and interpretation.

This paper aims to address the above issues and proposes a tool that would allow researchers to evaluate XR products by performing structured observations using a set of predefined codes and interval sampling. To the best of our knowledge, this is the first attempt to apply interval sampling method of observation to capture array of interactions of users of an XR product. Since surveys are frequently utilised to assess UX [1], because they allow for the users to give feedback from their point of view in a quick and cost effective manner to administer to the target population and score their results, the instrument would enable researchers to enrich their survey data with observational data that could be analysed in a quantitative manner, which may be an alternative for researchers with less experience in qualitative research methods.

II. Instrument Development

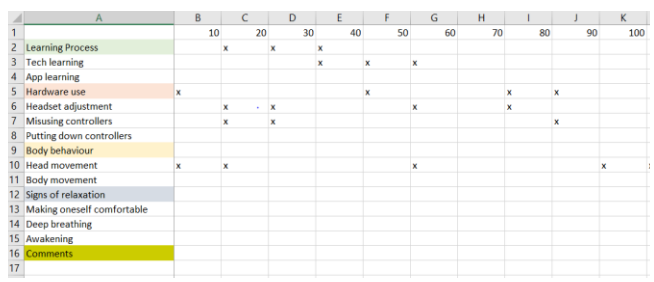
There are different ways of recording user behaviors during observation. Within an observation period, behaviors may be recorded continuously, or at specific intervals. Continuous recording is considered the gold standard for behavioral sampling, as this method records all occurrences of behaviors and their durations [10, 11]. However, such recording is often troublesome and challenging for researchers, since it requires vast amount of work and time to perform due to the large volume of data that it generates, or direct comparisons of response frequencies and durations are made. As such there is a need for alternative methods.

For the purpose of conducting observations, we applied noncontinuous sampling to capture target behaviors over fixed intervals [12]. With this approach, the coder records behaviors that occur at any time during the sampling interval using zero-one coding (behavior present or not present). With one-zero sampling, an interval is counted if the target event, despite its duration, is observed in any portion of the interval. Duration of observations may be defined by the length of specific experience, predetermined length of observations, or duration of a specific behavior. It is less intensive than continuous sampling, and therefore more feasible for researchers to conduct [13]. Furthermore, the versatility of this method allows researchers to make decisions as to how long intervals should be spaced. Longer observations may require longer intervals, and shorter intervals tend to result in values that match more closely the continuous pattern scores, but with more recording effort. There is a danger that too long intervals could overestimate behaviors that occurs for only a part of an interval and underestimate behaviors if multiple instances occur within one interval (and thus would be counted only once). Consequently, in our current study, a 10-minutes virtual reality experience was performed. It was decided that 10-seconds intervals would be efficient enough to capture user behaviors without causing significant workload.

TABLE I. Category of Codes

Category	***Code***	***Explanation***
Learning process	Learning how to use VR equipment	User learns how to use the VR headset or controllers.
	Learning how to use the application	User learns how to use the application.
Hardware use	Headset adjustment	VR headset adjustments – either by a user or a researcher.
	Misusing controllers	Controllers used in a way not predicted by the application.
	Putting down controllers	User puts down controllers and rests their hands.
Body behavior	Body movements	User makes body movements (i.e. rotation)
	Head movements	User makes head movements (i.e. rotation, looking up, down, sides).
Signs of relaxation	Making oneself comfortable	Adjusting or relaxing the body.
	Deep breathing	User breaths deeply.
	Awakening	Facial and bodily expressions related to awakening, verbal feedback.

Our observational tool defines a variety of behavioral codes relating to a user behavior while interacting with a VR system. The tool provides a means of sampling behaviors throughout a session. Such technique allows for quantification of observation data, and therefore, data achieved through this method can be used in conjunction with the quantitative findings from other methods. Furthermore, the structured nature of the tool ensures study replication.



	A	B	C	D	E	F	G	H	I	J	K
1		10	20	30	40	50	60	70	80	90	100
2	Learning Process		x	x	x						
3	Tech learning				x	x	x				
4	App learning										
5	Hardware use	x				x			x	x	
6	Headset adjustment		x	x			x		x		
7	Misusing controllers		x	x						x	
8	Putting down controllers										
9	Body behaviour										
10	Head movement	x	x				x				x
11	Body movement										
12	Signs of relaxation										
13	Making oneself comfortable										
14	Deep breathing										
15	Awakening										
16	Comments										
17											

Fig. 1. Example of the MO-XR tool use.

Given that it is the first attempt to propose such tool to code behavior of XR users, coding scheme needed to be developed. NA and ES developed 28 initial codes. This was followed by a collaborative process of aggregation and differentiation which was performed by all co-authors. The codes were then unified since they had similar semantics; and differentiated to identify where concepts had multiple meanings, thus, required splitting; and clarified to ensure that there was an agreed meaning to each code. The final list of codes was then refined and cleared for redundancies and repetitions. There are four categories of codes which are presented in Table 1. Figure 1 shows how the tool is utilized to evaluate our ASPIRE VR - Cabin Relaxation.

A. *Learning process*

The first category refers to learning how to use the hardware and the application. It involves verbal and nonverbal communication with researchers: learning how to use hardware relates to any help and guidance needed to be able to use the headset and/or controllers; learning how to use the application, relates to any help or guidance needed to be able to use the application, such as teleporting, navigating, information when the experience starts and ends, and so on.

B. *Hardware use*

The category refers to the time after the initial learning. Headset adjustment relates to any changes made to the headset during the experience, such as correcting the alignment to be able to better see the VR environment, or tightening the headset. Misusing controllers refers to using the controllers in a way not predicted by the application, such as pressing wrong buttons, or nervously seeking the correct buttons (i.e. pressing many buttons at the same time).

C. *Body movements*

The nature of the VR application evaluated in the study is static, therefore any movements of users were rather sporadic and subtle. There are two codes in this category – head movements that relate to e.g. the exploration of the VR environment; and body movements that are not directly predicted by the application, thus, are rather unusual and for this reason coded as abnormality of the application use.

D. *Signs of relaxation*

ASPIRE VR - Cabin Relaxation was designed to have minim user interactions, since the objective of the app is to immerse users in a VR environment to feel relaxed. Therefore, the aim of our observation study was to evaluate the VR application to assess whether or not it promotes relaxation. The tool also involves items that allows for the assessment of body signals of relaxation. For example, making oneself comfortable is a behavioral related to body language signal that is coded in the tool, such body language could be user's limbs hang loosely, they do not twitch and seldom cross one another, the user addresses any discomfort and finds a more comfortable position, such as leaning back on the armchair, user's legs sit gently on the floor. Any signs of tension, such as crossing legs, rigid movements, holding limbs close to the body indicate lack of relaxation. Deep breathing is one of the objectives of the application in which it is practiced, thus the behavior is coded separately. Signals of long, slow and deep breath suggest a state of calm [14]. On the other hand, short breath may signal not only stress, anxiety, but also excitement, therefore such instances were not coded as a signal of relaxation.

Post-experience awakening is a code which was developed during the formation of the tool, and it relates to the moment of exit from VR, which have received limited attention in the UX field [15]. It was observed that some participants experienced a process or transition similar to the process of awakening from sleep. That is, it was a transition from the meditative state to waking which implies physiological processes that lead to a new behavioral state[16]. The code relates to verbal and nonverbal cues. However, it has to be taken with caution, since such transition ought to be electrophysiologically defined for the objective assessment of the state [16]. However, due to the commonality of the occurrence of the signals of the state transition, we decided to include the code, as it directly relates to the application aim of relaxation..

III. METHODS

The Bournemouth University Research Board Ethics approved all the procedures and granted the study ethical approval (ID# 46007). The study was conducted in February and March 2023 in the UK and France. The study documentation, as well as the VR application, was available in both English and French.

A. *Participants*

Participants were recruited from the ASPIRE project participants across the UK and France, and from the general population using flyers distributed around the Bournemouth University campus. The inclusion criteria required participants to be over the age of 18 years and being able to give informed consent. A sample of 77 participants trialled ASPIRE VR - Cabin Relaxation and 51 agreed to be videorecorded (28 males, 23 females), and constituted the sample of this study. Of these, 5 were ASPIRE participants from the UK, 10 from France and 36 participants were recruited at the University. The average age of participants was 30.12 (SD = 13.29, 18 – 70).

B. *Cabin Relaxation Application*

ASPIRE VR – Cabin Relaxation (Figure 2) is a virtual reality app developed in Unity (2019.4.28f1) and the Oculus plugin for Unity. The aim of the application is to provide the user with between 8-10 minutes of guided meditation, where they sit and listen to the guide to relax in a warm and snowy cabin environment. The virtual environment included ambient audio (i.e., sound of wind and fire) and a guided mindfulness voiceover. The user is encouraged to observe the environment around them, imagining themselves within the cabin and to finally focus on their breathing for 3 to 5 minutes. ASPIRE VR - Cabin Relaxation incorporated a guided mindfulness voiceover delivered by a female narrator, who was fluent in English and French, thus both language versions involved the same voice. The voiceover delivered a focused-attention mindfulness practice which guided participants' attention to different parts of the virtual cabin as well as different physical sensations from the body.

Fig. 2. View inside the virtual cabin.

The application has been uploaded to the Meta Quest Store via the Applab platform, and can be seen and downloaded here: https://www.oculus.com/experiences/quest/8985434064860511/. For the user studies, the application was run natively on a Meta Quest Pro with Meta Quest Pro full light blockers for most of the participants and, for a small number of participants, the Meta Quest Pro VR earphones were also used. Meta Quest Pro is more comfortable to wear and less intrusive for the relaxation experience intended to be delivered and provides LCD panels at resolutions of 1800x1920 pixels for each eye, giving users an improved visual experience. The added light blockers further the immersive experience by blocking out excess light and glare from the real world, thus improving the overall user experience. The earphones used provide high fidelity audio and minimise background noise, allowing the user another method to remove themselves from the outside world. Combining these three pieces of technology enables the user to truly remove themselves from the outside world and take full advantage of the relaxation experience that the ASPIRE VR – Cabin Relaxation application aims to provide.

C. *Measures*

The survey questions were based on the measurement applied by Saginer et al. [17], which were adapted from validated questionnaires. Pragmatic and hedonic quality items were adapted from AttakDiff 2 [18]. Questions on ease of use and usefulness were based on scales from Davis [19]. Presence was assessed by using the Witmer & Singer [20] Presence Questionnaire. Intention to use items were developed by Fishbein and Ajzen [21]. Furthermore, participants were asked 11 questions related to their attitudes towards new technologies and VR. All items were 7-point Likert scale questions.

D. *Procedure*

The study was conducted at Bournemouth University campus, and at the ASPIRE sites across France and in the UK. The research took place in a quiet room with a comfortable armchair. Prior to the session, the researcher set up the VR Oculus Quest Pro equipment and disinfected the VR headset, hand controllers, and laptop computer. On their arrival participants were given information about the research project and the objectives of the experiment. They were then asked for their written consent for participating in the study and informed that they were free to stop participating at any time. Participants began by completing questionnaires to provide data about their demographics, technology literacy, and current stress level.

Next, they were instructed on how to put on the headset. Additional verbal instructions, and at times physical assistance, were provided while the participant placed the headset onto their head. Participants could wear glasses if needed in the headset. They then performed the VR task. This consisted, first, of a training phase in which they had to follow instructions to learn how to use the controllers. Next, participants performed VR meditation which lasted around 10 minutes, which was video recorded. After the session, they filled out a second series of questionnaires. The experiment lasted approximately 40 minutes, including filling out the questionnaires. Participants were given £20 voucher for their time.

IV. RESULTS

IBM SPSS Statistics version 28 was used for data analyses. Descriptive data analysis allowed for presentation of the proportion of the occurrence of each pattern. Correlational analyses were then performed to establish relationships between each pattern occurrence and the survey data.

Kolmogorov-Smirnov test was applied to each variable determine if data was normally distributed. The results indicate that only five variables were normally distributed: survey items including stress (p = .171), tension (p = .070), exhaustion (p = .187), usefulness (p = .200), and pragmatic quality (p = .062). Thus, Spearman's rank-order correlation was applied to data analysis.

A. *Intercoder reliability*

For the current study, it was decided that fully crossed design in which all cases are coded by all coders was not feasible to the large volume of data, thus each coder coded different cases. Coding was performed by NA, WT, ZL, TD and ES. NA and ES were responsible for the coding frame development, and WT, TD and ZL were not involved in the process. Such approach was decided since some researchers recommend coders to be individuals external to the research team who had no role in designing the coding frame [22]. Thus, the team constituted of researchers involved and uninvolved in the coding frame development to address the issue of the coding frame's external objectivity. Furthermore, such strategy provoked dialogue between researchers to identify issues with the coding frame, how and why interpretations conflict.

Prior to coding, training of coders was performed by NA and ES, since it was an in-depth coding of multiple patterns that were precisely defined, and given the nature of the application chosen – nuanced. Some of the patterns required greater degree of interpretation and therefore needed to be discussed with the team. Next, each member of the team coded two same full datasets to establish intercoder reliability for each code.

Interrater reliability was performed to assess the rigour and transparency of the coding frame and its application to the data [23, 24]. Intercoder reliability assessment ensures that the coding frame is sufficiently specified, and data is consistently coded [23]. It allows for its communicability across persons by showing that the basic analytic structure has meaning that extends beyond an individual researcher [23]. That is, performing interrater reliability guarantees that multiple individuals can understand and contribute to the analytic process, thus, the analysis transcends the imagination of a single person [25].

To maintain high coding standard, interrater reliability was determined by Cohen's Kappa statistics for randomly selected videos. Cohen suggested the Kappa result be interpreted as follows: values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement [26]. It was therefore decided that the minimum value of the interrater reliability over 0.61 would be satisfactory. The interrater reliability between the coders ranged from .696 to .871 which indicates substantial to almost perfect agreement.

B. Video analysis

The total amount of occurrences was calculated for each devoted pattern, which ranged from 1% to 100% of the observation period (Table 1). All patterns were categorised by frequency as follows: patterns observed in less than 10% of the total time of the experience, were considered rare, 11-40% were considered moderately frequent, and 41-60% were considered common, and more than 61% were considered very common. Since the duration of the VR experience varied, the percentage of each pattern within that experience was calculated and used to estimate of the percentage of time spend exhibiting each pattern.

Furthermore, it was found that participants who spent more time using the VR application, were less likely to show interest in new technologies (r = -.425, p = .002) and experimenting we new technologies (r = -.479, p < .001). Strikingly, there was a weak, however positive and significant relationship between the time spent in VR and the perception that VR is dangerous (r = .298, p = .034). Moreover, participants who spent more time using the application, were more likely to rate the interface quality more poorly (r = -.388, p = .005).

TABLE II. TOTAL AMOUNT OF OCCURENCES OF EACH CODE.

Code	***<10%***	***11-40%***	***41-60%***	***>61%***
Learning how to use VR equipment	92.2%	7.8%	0%	0%
Learning how to use the application	68.8%	29.4%	0%	2%
Headset adjustment	100%	0%	0%	0%
Misusing controllers	94.1%	5.9%	0%	0%
Putting down controllers	100%	0%	0%	0%
Body movements	90.2%	2%	0%	7.8%
Head movements	21.6%	39.2%	17.6%	21.6%
Making oneself comfortable	78.4%	17.6%	3.9%	0%
Deep breathing	27.5%	31.4%	17.6%	23.5%
Awakening	98%	2%	0%	0%

C. Learning

The analysis revealed that participants who spent more time learning how to use the VR equipment, showed less interest in new technologies (r = -.441, p = .001), and were less likely to experiment with new technologies (r = -.581, p < .001). On the other hand, participants who spent less time learning, showed higher increase in happiness (r = -.433, p = .001), and concentration (r = -.378, p = .007). It was found individuals who spent more time learning how to use the application, rated interface quality more poorly (r = -.349, p = .012), and were less able to examine the VR environment (r = -.379, p = .007).

D. Hardware

Participants who misused controllers in a way that was not predicted by the app, needed more time to learn the application (r = .321, p = .023), showed higher levels of exhaustion after the VR activity (r = .357, p = .011), and tension (r = .285, p = .045). Furthermore, participants who put down controllers during the VR meditation, showed bigger decrease in stress (r = .308, p = .030). Interestingly, participants who viewed VR technology as risky, were also more likely to spend more time adjusting their headsets (r = .355, p = .011).

E. Body movements

In terms of head movements, there were no significant correlations. However, participants who spent more time on moving their bodies, scored more highly on performance assessment (r = .315, p = .024).

F. Relaxation indicatives

The analysis revealed that longer periods of deep breathing were related to higher increase in happiness (r = .451, p = .001) and concentration (r = .461, p < .001). On the other hand, shorter periods of deep breathing, were related to lower ability to examine (r = -.380, p = .006), fear of using VR (r = -.426, p < .001) and the perception that VR is dangerous (r = -.314, p = .025). Spending more time on making themselves comfortable during the VR meditation, was related to higher feelings of presence (r = .448, p = .003), enjoyment (r = .363, p = .010), and ability to examine (r = .331, p = .018). Lastly, the indicatives of awakening, were related to higher increase in happiness (r = .376, p = .007) and concentration (r = .358, p = .011), and were negatively related to exhaustion (r = -.376, p = .015).

V. Discussion

The analysis of the observational data in conjunction with the survey data revealed that usability aspects were related to participants' perceptions on VR technology. Participants who are interested in new technologies, may potentially be more familiar with the VR equipment, needed less time to learn how to use the equipment. This is congruent with Fox et al. [27] who claimed that adoption to VR is slower than adoption for other technologies, which relates to high cost of entry, limited software library, cybersickness or frequent hardware updates, which may be serious barriers to many people and therefore did not have a chance to use it.

Furthermore, users who viewed VR technology as risky, spent more time adjusting their headsets. The uniqueness and importance of the code is related to very limited attention that the moments of donning or doffing the headset receive [15]. It might be speculated that one of the reasons for such behavior may be related to users' concerns over presence and immersion in VR overshadowing their awareness of reality. By strapping on a VR headset, a user is transported to a new, immersive world, which may be an intense experience for some individuals. Full engagement in VR is associated with a diminished sense of users' real environment [28], and potentially, users' negative perceptions of VR technology precluded them from psychological involvement in the scenario and fully engaging with the VR experience. In fact, users who expressed their fear of using VR and perceived it as dangerous, were less likely to follow the VR meditation guidance and showed shorter moments of deep breathing, and ultimately, showed lower increase or even decrease in their self-stated level of happiness.

The study revealed that learnability was related to presence components [17]. That is, participants who needed more time to learn how to use the application, rated interface quality more poorly and were less able to examine the VR environment. It may be speculated that some participants did not master using the application. Therefore, they were not able to fully engage with the VR experience, hence their lower self-stated ability to examine. Furthermore, learnability of the application was associated with enhanced mood after the experience.

Users who followed the meditation guidance and showed longer periods for deep breathing, were happier and more concentrated. Interestingly, the moments of exit from VR were also associated with a higher increase in happiness and concentration. The findings may relate to feelings of presence, since there was a significant positive correlation between self-rated happiness and presence, which is congruent with previous research [29]. Furthermore, participants who spent more time on making themselves comfortable during the VR meditation, showed higher levels of presence. Presence is perceived as a necessary mediator that allows real emotions to be activated by VR [30]. In fact, more immersive experiences are related to higher increase in emotional responses [31]. Thus, it appears that feelings of presence are a key to enhancing mood in VR.

A. *Strengths of the tool*

The novel tool for coding user behaviors provides a structured and descriptive method for observation study while users engage with an XR product. Maintaining high interrater reliability was an important consideration when developing the instrument. There is an excellent agreement between observers which suggests that different observers can use the tool and still produce very similar data, thus allowing reliable evaluation of XR products. Half day of training was sufficient for performing the coding, however, the exact amount of training and practice required will depend on many factors, such as previous experience of observers, or the nature of the application evaluated.

The outcomes of the survey supported the findings derived from the observational tool. That is, the Cabin Relaxation Application enhanced users' mood by increasing relaxation, which was reported by participants in their questionnaire responses, and it was observed when they interacted with the application. That is, following the VR meditation guidance which was observed and immersing oneself in the VR experience, which was reported in the survey, were associated with greater level of happiness.

Moreover, survey responses allowed for the exploration of some patterns of behavior which was observed when participants used the application. That is, users who were less likely to follow the VR meditation guidance, were less happy after the experience, which was associated with their negative attitudes towards VR. Furthermore, users who were less pleased with the application, spent more time learning how to use it, which implies that more attention should be paid to users' learning process since it may compromise their attitude towards the product.

Further psychometric testing of the instrument is required to increase evidence of validity and reliability. For example, researchers could apply it to XR applications that are more dynamic. The tool is very flexible and can be accommodated towards the objectives of the evaluation. That is, the intervals can be longer or shorter, as well as new codes can be added depending on the nature of the product. The current study evaluated a meditation application that was aimed at enhancing mood, hence the codes were designed around that purpose, yet the tool allows for codes modification. However, researchers using the tool should focus on maintaining high interrater reliability when developing new codes and make sure that all observers share similar perspectives on the codes.

VI. CONCLUSION

This study contributes to the user experience literature by providing an observational tool which offers a structured way of approaching videorecorded data. Using the tool, quantitative behavioral data is recorded during observation based on the activity observed by coders. Survey data can be further utilised to validate and expand on the observations. It is an inexpensive and easy-to-use method of evaluating XR products, which is highly flexible and can be tailored towards the research objectives and can easily be applieded by observers with different backgrounds and experience levels.

ACKNOWLEDGMENT

This work is funded by grants from the EU Interreg European Regional Development Fund (ASPIRE 191). The authors thank the participants and all the team members of the Adding to Social capital and individual Potential In disadvantaged Regions (ASPIRE) project.

REFERENCES

[1] Kim, Y.M., I. Rhiu, and M.H. Yun, A Systematic Review of a Virtual Reality System from the Perspective of User Experience. International Journal of Human–Computer Interaction, 2020. 36(10): p. 893-910.

[2] Floyd, F.J., et al., 3.01 - Observational Methods, in Comprehensive Clinical Psychology, A.S. Bellack and M. Hersen, Editors. 1998, Pergamon: Oxford. p. 1-21.

[3] Baker, L., Observation: A Complex Research Method. Library Trends, 2006. 55(1): p. 171-189.

[4] Moser, A. and I. Korstjens, Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. European Journal of General Practice, 2018. 24(1): p. 9-18.

[5] Susan Farrell, T.F. Field Studies 2024 1.04.2024]; Available from: https://www.nngroup.com/articles/field-studies/#:~:text=Direct%20observation%20is%20a%20purely,help%20create%20natural%20task%20flows.

[6] Edwards, R., B. Huber, and O. Kramarova, Observations on the Field of UX Research. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2020. 64(1): p. 1571-1575.

[7] Ross, L., et al., Validation of alternative behavioral observation methods in young broiler chickens. Poult Sci, 2019. 98(12): p. 6225-6231.

[8] White, C., et al., Writing up qualitative research. Qualitative research practice: A guide for social science students and researchers, 2014: p. 367-400.

[9] Fugelsang, J.A., et al., Theory and data interactions of the scientific mind: evidence from the molecular and the cognitive laboratory. Can J Exp Psychol, 2004. 58(2): p. 86-95.

[10] Lehner, P.N., Handbook of ethological methods. 1998: Cambridge University Press.

[11] Hämäläinen, W., et al., Measuring behaviour accurately with instantaneous sampling: A new tool for selecting appropriate sampling intervals. Applied Animal Behaviour Science, 2016. 180: p. 166-173.

[12] 12. Alessi, G.J., Behavioral observation for the school psychologist: Responsive-discrepancy model. School Psychology Review, 1980. 9(1): p. 31-45.

[13] Martin, P. and P. Bateson, Recording methods. Measuring behaviour: an introductory guide. 2007, Cambridge University Press.

[14] Russo, M.A., D.M. Santarelli, and D. O'Rourke, The physiological effects of slow breathing in the healthy human. Breathe (Sheff), 2017. 13(4): p. 298-309.

[15] Knibbe, J., et al., The Dream is Collapsing: The Experience of Exiting VR, in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018, Association for Computing Machinery: 2018. p. 1-13 483.

[16] Åkerstedt, T., et al., Awakening from sleep. Sleep Medicine Reviews, 2002. 6(4): p. 267-286.

[17] Sagnier, C., et al., User Acceptance of Virtual Reality: An Extended Technology Acceptance Model. International Journal of Human–Computer Interaction, 2020. 36(11): p. 993-1007.

[18] Hassenzahl, M., M. Burmester, and F. Koller, AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität, in Mensch & Computer 2003: Interaktion in Bewegung, G. Szwillus and J. Ziegler, Editors. 2003, Vieweg+Teubner Verlag: Wiesbaden. p. 187-196.

[19] Davis, F.D., Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 1989. 13(3): p. 319-340.

[20] Witmer, B.G. and M.J. Singer, Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments, 1998. 7(3): p. 225-240.

[21] Fishbein, M. and I. Ajzen, Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. 1975: Addison-Wesley Publishing Company.

[22] Kolbe, R.H. and M.S. Burnett, Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. Journal of Consumer Research, 1991. 18(2): p. 243-250.

[23] Joffe, H. and L. Yardley, Content and thematic analysis. 2003.

[24] Mays, N. and C. Pope, Qualitative research: rigour and qualitative research. Bmj, 1995. 311(6997): p. 109-112.

[25] Kurasaki, K.S., Intercoder reliability for validating conclusions drawn from open-ended interview data. Field methods, 2000. 12(3): p. 179-194.

[26] McHugh, M.L., Interrater reliability: the kappa statistic. Biochem Med (Zagreb), 2012. 22(3): p. 276-82.

[27] Fox, J., D. Arena, and J.N. Bailenson, Virtual reality: A survival guide for the social scientist. Journal of Media Psychology, 2009. 21(3): p. 95-113.

[28] Slater, M. and A. Steed, A Virtual Presence Counter. Presence: Teleoper. Virtual Environ., 2000. 9(5): p. 413–434.

[29] Felnhofer, A., et al., Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. International Journal of Human-Computer Studies, 2015. 82: p. 48-56.

[30] Price, M., et al., Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. Journal of anxiety disorders, 2011. 25(6): p. 763-770.

[31] Visch, V.T., E.S. Tan, and D. Molenaar, The emotional and cognitive effect of immersion in film viewing. Cognition and Emotion, 2010. 24(8): p. 1439-1445.