EXPLORATIONS IN GAME EXPERIENCE: A CASE STUDY OF 'HORIZON ZERO DAWN'

Barbaros Bostan
Game Design Department
Bahcesehir University
Istanbul, Turkey
barbaros.bostan@comm.bau.edu.tr

Mehmet İlker Berkman
Communication Design
Bahcesehir University
Istanbul, Turkey
ilker.berkman@comm.bau.edu.tr

***Abstract*— The aim of this study is to find the relationship between the specific gameplay elements of a chosen game and the gameplay experience it provides which will be measured by a selected survey from the literature, the Game Experience Questionnaire (GEQ). Before illustrating the relation between the game elements and the player experience measured via GEQ, a confirmatory factor analysis was conducted to validate the factor structure of GEQ and GEQ is also evaluated for its unidimensionality, internal reliability, convergent validity and discriminant validity. The selected game for the study is Horizon Zero Dawn which is an open-world action role-playing video game developed by Guerrilla Games. Selling more than 2.6 million units worldwide in two weeks, the game has become the best-selling new first party IP launch on the PlayStation 4. Our study aims to answer the following questions: (1) Can we measure a successful and popular game's experience with GEQ? (2) What are the gameplay elements that make this game so popular? (3) Are there any correlations with the subscales of GEQ and the gameplay elements in question?**

Keywords*— *game experience; gameplay elements; game experience questionnaire; GEQ; role playing games

I. INTRODUCTION

The concept of user experience is the subjective relationship between the user and the application [1] and player experience in computer games is usually associated with four terms: Engagement, Immersion, Flow and Presence. Engagement is the state of involvement with a video game and is closely related with how enjoyable the experience is. Immersion is a quantifiable description of a technology and includes the extent to which the computer displays are extensive, surrounding, inclusive, vivid and matching [2]. Flow is the enjoyable state experienced when engaged in a challenging activity [3] and presence is the sense of being inside a virtual world [4]. Discussing the relationship between these terms and the game experience is beyond the scope of this article but it may be argued that these terms are the outcome of extreme gaming experiences rather than the core elements of the interactive gameplay process itself [5].

Core elements of the interactive gameplay can be identified as mechanics, dynamics and aesthetics, based on the MDA framework [6]. The experience of gaming is a personal experience and is also affected by the genre or the game in question because each genre, even each game in a specific genre, uses different mechanics. Even if the mechanics seem similar, how these mechanics act on player inputs and with each other over time, i.e. dynamics of the game, changes. Aesthetics, the emotional responses evoked in the player also differ, even between the quests or levels of a single game. The aim of this study is to find the relationship between the specific gameplay elements of a chosen game and the gameplay experience it provides which will be measured by a selected survey from the literature. The gameplay elements in question can be conceived as the mechanics, dynamics and aesthetics of the MDA framework [6].

II. RELATED STUDIES

Modeling player experience is a diverse field of research. Wiemeyer, Nacke, Moser and Mueller [7] analyzed the following general and domain-specific psychological models of player experience: the Player Experience of Need Satisfaction (PENS) [8] based on Self-Determination Theory [9]; Keller's [10, 11] Attention, Relevance, Confidence, Satisfaction (ARCS); Sweetser and Wyeth's [12] GameFlow based on Flow theory [3]; the Presence-Involvement-Flow Framework (PIFF2) [13]; the Fun of Gaming (FUGA) [14]; Core Elements of the Gaming Experience (CEGE) [5]; and the Play Heuristics of Desurvire and Wiberg [15]. The researchers stated that the models either apply general concepts from other fields to gaming or address the process of player experience from different perspectives (players, developers, researchers). Some of the elements defined by these models are hardly separable and there are complex interactions between them. At the end, when they combine the results, the following (social-)psychological elements of player experience are reported: competence; autonomy and control; immersion, (spatial and social) presence, flow and GameFlow; involvement and (enduring) engagement; social relatedness and social interaction; challenge; tension; curiosity; fantasy; positive and negative emotions; intrinsic goals; feedback and evaluation.

Assessment of the player experience with surveys is also a research area with different approaches. One of the earliest experience scale is the Presence Questionnaire of Witmer and Singer [16] commonly used in VR (virtual reality) and VEs (virtual environments). The MEC spatial presence questionnaire (MEC-SPQ) [17], the Spatial Presence Experience Scale (SPES) [18] and the Social Presence in Gaming Questionnaire (SPGQ)

Bostan, B. & Berkman, M.I. (2017) Explorations in Game Experience: A Case Study of 'Horizon Zero Dawn'. Proceedings of the Eurasia Graphics 2017, Istanbul, Turkey, November 4-5.

[19] are three presence related questionnaires used to assess player experience. In terms of domain specific surveys three of them deserve special attention. Core Elements of the Gaming Experience Questionnaire, CEGEQ, consists of 38 items for 10 dimensions that measure enjoyment, frustration, CEGE, puppetry, video-game, control, facilitators, ownership, gameplay, and environment [5]. The Cronbach alpha for the whole questionnaire is .794. The Game Engagement Questionnaire (GEnQ) consists of 19 items that measure absorption, flow, presence, and immersion [20]. The Cronbach alpha for the whole questionnaire is .85. The Game Experience Questionnaire (GEQ) consists of 36 items representing 7 latent variables that measure competence, immersion, flow, tension, challenge, positive and negative affect [21]. The subscales of GEQ have an average Cronbach's alpha score of .81.

Considering the diversity of game experience surveys mentioned above, our survey selection criteria are: (1) choosing a domain specific survey rather than a general one applied to gaming and (2) selecting a survey which is tested with sufficient number of participants that play different types of games. Among the three domain specific surveys, CEGEQ [5] was tested with 15 participants who played Tetris. GEnQ [20] is tested twice, first with 153 junior high school students and then with 107 undergraduate students who played a first-person shooter named 'S.T.A.L.K.E.R.: Shadow of Chernobyl'. The typical game playing time per week based on six predetermined time categories were taken but which genres or games they usually played were not published. GEQ [21] is tested with 380 participants with diverse gaming backgrounds: First Person Shooter games (22%), Role Playing games (14%), Sport games (13%), Puzzle/board/card games (11%), Action adventure games (10%), Strategy games (9%), and other genres (e.g., simulation games, fight games, children's games, music games) (11%).

Based on our selection criteria, GEQ seemed the most suitable choice. It is designed to assess the games in terms of playing experience. While it also have a Social Presence Module and a Post-Game module, the 36-item core covers a broader range of dimensions that are associated with gaming experience, compared to other questionnaires. The original development efforts depend on more participants that play games within a wider variety of genres. On the other hand, the psychometric qualities of GEQ were not completely established [22, 23], especially related with its factor structure that represents its construct validity. Brühlmann and Schmid [24] explored the factor structure of GEQ using principal axis exploratory factor analysis with oblimin rotation. Their results did not match the factor structure suggested in scoring guidelines published by Ijselstein et al. [21] and the factors evaluated through two different games were not similar. However, a confirmatory approach has not been employed yet, to validate the factor structure of GEQ. For this reason, we decided to run a confirmatory factor analysis as a part of this study before illustrating the relation between the game elements and the player experience measured via GEQ.

III. METHODOLOGY

The selected game for the study is Horizon Zero Dawn which is an open-world action role-playing video game developed by Guerrilla Games and published by Sony Interactive Entertainment for the PlayStation 4 on 28 February 2017 in North America, on 1 March in Europe and 2 March in Asia. Selling more than 2.6 million units worldwide in two weeks, including units sold at retail and digitally through the PlayStation Store, the game has become the best-selling new first party IP launch on the PlayStation 4. The game has a Metascore of 89 based on 115 critics and a user score of 8.3 based on 4584 ratings[1]. Game critics referred the gameplay design of Horizon Zero Dawn as the best of everything with the minimum number of flaws. Choosing a commercial game with such positive reviews for a player experience study has the risk of having extreme negative skewness within the collected data. However, our study aims to answer the following questions: (1) Can we measure such a successful title's game experience with GEQ? (2) What are the gameplay elements that make this game so popular? (3) Are there any correlations with the subscales of GEQ and the gameplay elements in question?

After an analysis of game reviews for Horizon Zero Dawn, 36 gameplay elements drew our attention. Seven of them are narrative elements: quality of the main quest, length of the main quest, quality of side quests, number of side quests, antagonists of the story, side characters of the story, and lore/fiction of the world. Four elements focus on the protagonist of the game, Aloy: Aloy's physical appearance, Aloy's dialogue options, Aloy's voice acting, and Aloy's facial expressions and animations. Six elements are related with the customization power or the interactivity range provided by the game: armor variety, weapon variety, machine variety, trap variety, ammo variety, and skill variety. There are 3 audio-visual elements: environment (visual) design, machine (visual) design and game soundtrack. 10 are core gameplay elements: enemy AI, crafting system, game controls, combat mechanics, machine overriding, game UI, save system, travel options, inventory system, and game feedback. Besides the side quests, the game offers several activities and collectibles, which are summarized by 6 elements: Bandit Camps, Hunting Grounds, Tallnecks, Cauldrons, Corrupted Zones and Collectibles. Participants were asked to rate their satisfaction levels about these 36 gameplay elements on a five-point Likert scale with a rating ranging from 1 for 'not at all satisfied' to 5 for 'extremely satisfied'. Participants were then asked to answer the GEQ questionnaire, followed by demographic questions.

Three hundred and thirty-seven participants (44 female; 276 male; 17 without gender declaration) took part in an online study, which was announced on social media platforms and several forums of the selected game. Ages ranged from 13 to 66 (M = 29.05, SD = 8.651). In terms of gaming habits, 6 participants without any declaration, the remaining 331 participants are classified into five groups: 0-1 h/week (.6%), 2-5 h/week (12.2%), 6-10 h/week (31.5%), 11-15 h/week (26.1%)

[1] Metacritic website, accessed on the 5th of June 2017, http://www.metacritic.com/game/playstation-4/horizon-zero-dawn

and more than 15 h/week (27.9%), indicating that the majority of the participants are "frequent" game players.

IV. RESULTS

A. Psychometric Evaluation of GEQ

We evaluated each GEQ item for their skewness and identified that data on 14 of 36 items were skewed, when the skewness statistic divided by its standard error and the kurtosis statistic divided by its standard error are greater than z = ±3.29, suggesting that the distribution is severely skewed. [25]. Visual inspection of histograms also supports this finding. For this reason, we decided to embrace the partial least squares confirmatory factor analysis (PLS-CFA) approach instead of a covariance-based method, as PLS-CFA is not sensitive to skewed data [26, 27]. The skewness is due to the positive bias of participants, who have voluntarily evaluated what they have been playing. Besides, the sensory-imagery immersion construct of GEQ drives independent items that make it a formative variable, which can be inspected via PLS based methods. Using PLS-CFA, we evaluated GEQ for its unidimensionality, internal reliability, convergent validity and discriminant validity using construct development and validation guidelines of Lewis et al. [28] and Straub et al. [29], based on the predetermined structure of 7 latent variables described by Ijselstein et al. [21].

As an indicator of unidimensionality of a latent variable, each item should load with a high coefficient on one factor only, and this factor should consistently be the same for all items that are supposed to measure it, with an eigenvalue exceeding 1. A predefined latent variable with items that load on two factors exceeding eigenvalue 1 suggests that it is not, unidimensional. Thus, those items may refer to a latent variable in addition to the intended construct. The items intended for competence, immersion, flow, tension, positive and negative affect lead to a single construct with an eigenvalue exceeding 1, but the items intended for challenge construct revealed two factor with eigenvalue greater than 1. The first factor has an eigenvalue 1.94 while the second has 1.13. The results imply that items intended for challenge construct are also driven by some other latent variable. Eigenvalues of all latent variables can be followed on Table II.

The item CHA 23 has a higher load of .623 on a secondary dimension rather than its intended dimension, exceeding .6. Items CHA 32 and CHA 26 also have a medium load exceeding .4 on the same secondary factor, while the items CHA 11 and CHA 26 also have medium level load on a second factor. So, the resulting factor structure is violating the unidimensionality of the latent variable Challenge. These problematic loads are given on Table I.

TABLE I. VARIABLES / FACTORS CORRELATIONS OF GEQ

CHALLENGE					
	F1	F2	F3	F4	F5
CHA_11	0.663	-0.316	**0.568**	-0.249	-0.275
CHA_23	0.573	**0.623**	0.197	-0.143	**0.473**
CHA_32	0.603	**0.564**	-0.188	0.302	**-0.437**
CHA_33	0.658	-0.278	**-0.558**	-0.421	0.027
CHA_26	0.613	**-0.497**	-0.015	0.557	0.256
COMPETENCE					
	F1	F2	F3	F4	F5
COM_02	0.761	-0.072	-0.324	-0.045	0.555
COM_10	0.555	-0.524	0.632	-0.135	0.019
COM_15	0.733	0.109	-0.262	-0.500	-0.364
COM_21	0.509	0.745	0.415	0.097	0.063
COM_17	0.726	-0.156	-0.170	0.587	-0.274
POSITIVE AFFECT					
	F1	F2	F3	F4	F5
PE_01	0.705	-0.390	-0.291	0.515	-0.013
PE_04	0.643	0.424	0.562	0.216	-0.209
PE_14	0.773	-0.284	-0.111	-0.408	-0.378
PE_06	0.799	-0.164	0.228	-0.198	0.493
PE_20	0.509	0.696	-0.497	-0.056	0.082
NEGATIVE AFFECT					
	F1	F2	F3	F4	
NE_07	0.586	**-0.611**	-0.505	-0.171	
NE_08	0.681	-0.201	0.653	-0.266	
NE_09	0.803	0.102	0.009	0.587	
NE_16	0.600	**0.688**	-0.259	-0.317	
SENSORY IMAGERY IMMERSION					
	F1	F2	F3	F4	F5
SII_03	0.461	**-0.646**	0.533	0.160	-0.244
SII_12	0.438	**0.720**	0.472	-0.021	-0.260
SII_18	0.710	-0.130	-0.394	-0.515	-0.239
SII_19	0.674	0.098	-0.492	0.536	-0.086
SII_30	0.788	0.011	0.202	-0.076	0.576
FLOW					
	F1	F2	F3	F4	F5
FLO_05	0.571	**0.697**	-0.415	0.126	-0.006
FLO_25	0.773	-0.187	0.208	0.500	-0.271
FLO_28	0.699	0.358	0.554	-0.243	0.134
FLO_13	0.767	-0.267	-0.215	-0.446	-0.308
FLO_31	0.775	-0.385	-0.189	0.068	0.459
TENSION / ATTENTION					
	F1	F2	F3		
TA_22	0.846	-0.023	0.533		
TA_24	0.764	0.584	-0.274		
TA_29	0.773	-0.552	-0.312		

These results imply that some variables assigned to constructs can be related with some other dimension that is considered or not considered within the scale.

The reliability indicator for each latent variable is Cronbach's alpha value, as well as Dillon-Goldstein (D-G) rho value. D-G rho is a better indicator of reliability since it does not assume that each manifest variable is equally important in defining the latent variable, while Cronbach's alpha makes this assumption [26]. Latent variable Negative Effect and Sensory-Imagery Immersion have Cronbach alpha values of .59, which

are quite close to .6 that determine a reliable construct, while others exceed this limit, as given below. All D-G rho values are above the acceptable limit of .7, indicating that predetermined latent variables of GEQ have internal reliability. Reliability indicators can be followed on Table II.

TABLE II. RELIABILITY INDICATORS AND EIGENVALUES OF GEQ DIMENSIONS

Latent variable	Cronbach's alpha	D.G. rho (PCA)	Eigenvalues
CHALLENGE	0.605	0.760	1.942
			1.131
			0.709
			0.662
			0.557
COMPETENCE	0.676	0.795	2.211
			0.871
			0.773
			0.625
			0.520
POSITIVE AFFECT	0.723	0.819	2.406
			0.924
			0.712
			0.521
			0.437
NEGATIVE AFFECT	0.589	0.765	1.812
			0.896
			0.748
			0.545
SENSORY IMAGERY IMMERSION	0.599	0.758	1.984
			0.962
			0.945
			0.585
			0.524
FLOW	0.766	0.843	2.601
			0.869
			0.604
			0.529
			0.397
TENSION / ATTENTION	0.708	0.837	1.897
			0.646
			0.456

The average variance extracted (AVE) index should exceed .5 to verify convergent validity, suggesting that a construct's items should assemble at a higher degree compared to items measuring other latent variables. The Flow and Tension-Attention constructs met this criterion with a value of .63, while others did not. This result suggests that most of the constructs of GEQ are not able to explain more than half of the variance of their indicators [27]. In other words, the low AVE's indicate that items do not represent their latent variables clearly. Results can be followed on Table III.

If the shared variance between two constructs is lower than the AVE for each individual construct, then each construct has different measures from each other, indicating the discriminant validity of the scale [30], which is called Fornell-Larcker criterion. When we inspected the shared variances between constructs and AVE's, we observed that shared variance of Positive Effects and Sensory-Imagery Immersion is .42, which is higher than the .38 AVE value for Sensory-Imagery Immersion, as seen on Table III. This finding violates the Fornell-Larcker criterion, providing evidence that disproves discriminant validity of the scale.

TABLE III. CONVERGENT AND DISCRIMINANT VALIDITY INDICATORS OF GEQ

	CHA	COM	PE	NE	SII	FLO	TA	(AVE)
CHA		0.002	0.016	0.040	0.023	0.107	0.087	0.375
COM	0.002		0.258	0.014	0.217	0.140	0.032	0.436
PE	0.016	0.258		0.119	**0.422**	0.327	0.105	0.481
NE	0.040	0.014	0.119		0.081	0.064	0.282	0.450
SII	0.023	0.217	**0.422**	0.081		0.321	0.065	0.379
FLO	0.107	0.140	0.327	0.064	0.321		0.022	0.516
TA	0.087	0.032	0.105	0.282	0.065	0.022		0.630
(AVE)	0.375	0.436	0.481	0.450	0.379	0.516	0.630	

Another detailed indicator of the discriminant validity of a scale is that each item has the highest loading on their designated latent variable [27]. When we inspected the cross-loadings of GEQ items, we observed that each item has the highest load on its designated construct. However, some items that strongly load on Positive Effects have also moderately loaded on Sensory-Imagery Immersion.

Besides the EFA factor loads indicating unidimensionality, we also utilized cross-loadings to identify discriminant validity. As suggested by Chin [27], we used the squares of cross-loadings because this representation provides "more intuitive interpretation since it represents the percentage overlap between an item and any construct". As it can be investigated on Table IV, Items have loaded strongly on their intended construct at the first hand, providing evidence for unidimensionality of GEQ.

TABLE IV. ITEM CROSS-LOADINGS OF GEQ ON LATENT VARIABLES

	CHA	COM	PE	NE	SII	FLO	TA
CHA_11	**0.254**	0.010	0.000	0.009	0.001	0.018	0.012
CHA_23	**0.368**	0.002	0.004	0.185	0.008	0.000	0.246
CHA_32	**0.392**	0.000	0.000	0.051	0.000	0.022	0.070
CHA_33	**0.383**	0.001	0.003	0.002	0.012	0.074	0.009
CHA_26	**0.480**	0.015	0.084	0.013	0.092	0.145	0.001
COM_02	0.001	**0.592**	0.148	0.016	0.124	0.056	0.021
COM_10	0.006	**0.305**	0.075	0.000	0.052	0.031	0.009

COM_15	0.006	**0.470**	0.103	0.011	0.059	0.062	0.012
COM_21	0.001	**0.173**	0.043	0.002	0.022	0.018	0.000
COM_17	0.005	**0.641**	0.187	0.012	0.203	0.135	0.029
PE_01	0.001	0.186	**0.475**	0.043	0.167	0.107	0.034
PE_04	0.016	0.066	**0.409**	0.047	0.162	0.122	0.054
PE_14	0.021	0.233	**0.621**	0.049	0.290	0.260	0.055
PE_06	0.004	0.111	**0.645**	0.131	0.280	0.218	0.084
PE_20	0.004	0.048	**0.253**	0.029	0.113	0.079	0.029
NE_07	0.045	0.002	0.013	**0.267**	0.000	0.000	0.114
NE_08	0.027	0.001	0.051	**0.438**	0.055	0.039	0.043
NE_09	0.040	0.026	0.092	**0.723**	0.071	0.039	0.258
NE_16	0.002	0.001	0.061	**0.372**	0.033	0.061	0.115
SII_03	0.013	0.020	0.079	0.005	**0.154**	0.060	0.006
SII_12	0.000	0.014	0.059	0.032	**0.164**	0.063	0.030
SII_18	0.011	0.173	0.154	0.025	**0.492**	0.185	0.025
SII_19	0.010	0.063	0.127	0.023	**0.287**	0.086	0.014
SII_30	0.019	0.150	0.381	0.073	**0.800**	0.230	0.054
FLO_05	0.025	0.043	0.223	0.074	0.172	**0.408**	0.032
FLO_25	0.089	0.078	0.113	0.030	0.131	**0.550**	0.001
FLO_28	0.032	0.105	0.211	0.034	0.247	**0.543**	0.030
FLO_13	0.079	0.081	0.179	0.021	0.132	**0.550**	0.004
FLO_31	0.078	0.058	0.110	0.015	0.137	**0.530**	0.003
TA_22	0.052	0.042	0.097	0.210	0.093	0.038	**0.776**
TA_24	0.038	0.005	0.076	0.192	0.025	0.011	**0.576**
TA_29	0.090	0.018	0.027	0.131	0.010	0.000	**0.537**

To investigate formative variables, we examined each formative indicator's weight (relative importance) and loading (absolute importance), using bootstrapping with a sample of 5000 at the 5% significance level [32]. For formative variables, the recommended standardized path coefficients should be greater than .100 [31] or .200 [26]. Sensory – Imagery Immersion is the only latent variable considered as a formative construct. We followed Lohmöller's recommendation with a more liberal approach since many of our variables did not meet the criteria of Chin, as seen follows.

TABLE V. FORMATIVE VARIBLE WEIGHTS

Items	Weight
SII_03	0.112
SII_12	0.153
SII_18	0.372
SII_19	0.139
SII_30	0.625

Based on the whole psychometric quality indicators, it is possible to say that GEQ is valid and reliable tool to investigate the player experience. Although our data revealed some flaws of unidimensionality in GEQ, it should be considered that the data set evaluated in this study is collected from the players of only one game.

B. Factor Analysis of Gameplay Elements

An exploratory approach is required to identify the constructs since our Gameplay Elements are developed only for evaluation of the selected game. The items were generated based on the elements within the game, rather than aiming to measure a set of theoretical constructs.

Prior to the factor analysis of gameplay elements, the individual responses to the items were screened to determine if there was substantial skewness or kurtosis, as well as inspected for outliers. The data were negatively skewed and this is not surprising for a satisfaction scale measured with a game with very positive reviews but concern arises when the skewness statistic divided by its standard error and the kurtosis statistic divided by its standard error are greater than $z = \pm 3.29$ [25]. As suggested by Tabachnick and Fidell [25] and Howell [33], Logarithmic (Log 10) transformation is applied to Gameplay Elements. The transformation was suggested for substantially negative skewness and is represented by $NEWX = LG10(K - X)$ where K is a constant from which each score is subtracted so that the smallest score is 1; usually equal to the largest score + 1. As a result of this transformation method, item scores are reversed as 0 for "Satisfactory" and 1 for "Unsatisfactory".

A principal components analysis (PCA) was run on 36 transformed gameplay elements that measure satisfaction levels on 337 participants. The suitability of PCA was assessed prior to analysis. Inspection of the correlation matrix showed that all variables had at least one correlation coefficient greater than 0.3. The overall Kaiser-Meyer-Olkin (KMO) measure was 0.917 which is classified as 'marvelous' according to Kaiser [34]. Bartlett's test of sphericity was statistically significant ($p < .0005$), indicating that the data was likely factorizable.

The initial PCA with oblimin rotation revealed nine components that had eigenvalues greater than one and but the 8th component had an eigenvalue of 1.085 and the 9th component had an eigenvalue of 1.007. To remove these two components with eigenvalues barely above 1, the extraction criteria were set to a fixed number of seven factors and the remaining components explained a cumulative variance of 38.975%. The interpretation of the data was consistent with the gameplay attributes the questionnaire was designed to measure with strong loadings of core gameplay elements on Component 1, story protagonist items on Component 2, audio-visual items on Component 3, narrative items on Component 4, customization items on Component 5, a group of mixed items on Component 6, two prominent core gameplay elements, "game controls" and "combat mechanics", on Component 7. Component loadings are presented in Table VI. Component 1 is named as "Core Gameplay Elements", Component 2 is named as "Avatar Elements", Component 3 is named as "Audio-Visual Elements", Component 4 is named as "Narrative Elements", Component 5 is named as "Customization Elements", Component 6 is named as "Secondary Gameplay Elements", and Component 7 is named as "Playability Elements".

TABLE VI. ROTATED PCA SOLUTION

Structure Matrix

	Factor						
	1	2	3	4	5	6	7
Inventory system	,658	,240	,369	,303	-,411	,286	-,240
Crafting system	,625	,255	,314	,321	-,400	,425	-,271
Activities - Bandit Camps	,597	,320	,093	,348	-,363	,416	-,204
Skill variety	,547	,336	,254	,396	-,456	,339	-,202
Enemy AI	,511	,343	,246	,342	-,419	,237	-,251
Game UI	,497	,306	,313	,277	-,327	,292	-,388
Trap variety	,491	,159	,276	,380	-,487	,329	-,184
Game feedback	,484	,396	,392	,202	-,389	,386	-,231
Collectibles	,402	,237	,347	,366	-,386	,331	-,031
Save system	,395	,135	,361	,339	-,318	,374	-,253
Aloy`s dialogue options	,471	,624	,162	,185	-,412	,587	-,214
Aloy`s voice acting	,215	,590	,285	,233	-,231	,157	-,173
Aloy`s facial expressions and animations	,124	,585	,186	,205	-,132	,152	-,174
Aloy`s physical appearance	,266	,524	,314	,228	-,303	,345	-,109
Quality of the main quest	,153	,418	,366	,386	-,268	,276	-,281
Machine (visual) design	,225	,233	,602	,207	-,273	,229	-,256
Game soundtrack and sound effects	,293	,362	,554	,126	-,272	,299	-,204
Activities - Tallnecks	,303	,393	,409	,257	-,376	,327	-,049
Environment (visual) design	,123	,178	,328	,149	-,130	,144	-,244
Number of side quests	,347	,189	,164	,623	-,393	,284	-,207
Quality of side quests	,402	,352	,071	,575	-,417	,329	-,087
Length of the main quest	,128	,295	,234	,512	-,321	,132	-,220
Side characters of the story	,343	,493	,295	,501	-,315	,320	-,338
Activities - Cauldrons	,396	,184	,316	,404	-,260	,179	-,316
Activities - Hunting Grounds	,335	,156	,180	,381	-,240	,212	-,183
Weapon variety	,275	,135	,149	,300	-,710	,215	-,218
Armor variety	,453	,253	,203	,278	-,664	,444	-,087
Ammo variety	,324	,281	,341	,328	-,601	,272	-,242
Machine variety	,249	,216	,366	,471	-,564	,155	-,267
Machine overriding	,419	,333	,378	,403	-,473	,372	-,241
Travel options	,315	,214	,332	,251	-,336	,688	-,175
Antagonists of the story	,371	,436	,191	,416	-,330	,520	-,366
Activities - Corrupted Zones	,491	,180	,238	,461	-,390	,511	-,241
Lore/fiction of the world	,079	,367	,276	,143	-,235	,374	-,308
Game controls	,352	,193	,290	,282	-,339	,319	-,635
Combat mechanics	,299	,248	,367	,320	-,452	,124	-,566

Six of the proposed core gameplay elements are grouped together in Component 1 with the addition of an activity element, "Bandit Camps" and the "Collectibles" element. "Skill Variety" and "Trap Variety", which are proposed customization elements, fall under Component 1 but they also have high loadings on Component 5 which is composed of the remaining customization elements. Four elements that focus on the protagonist of the game, Aloy, are grouped together in Component 2 with the addition of a narrative element, "Quality of the Main Quest". The three proposed audio-visual elements are grouped together in Component 3 with the addition of an activity, "Tallnecks". Four of the proposed narrative elements are grouped together in Component 4 with the addition of two activities, "Cauldrons" and "Hunting Grounds". Four of the proposed customization elements are grouped together in Component 5 with the addition of "Machine Overriding". Component 6 is a mix of different variables, possibly highlighting the problematic elements in the questionnaire. Component 7 is comprised of two elements, "Game Controls" and "Combat Mechanics", which distinguish themselves from other core gameplay elements in a significant way. The overall factor structure is consistent with the proposed gameplay elements structure with the exception of activities. Although the developers named these elements as activities, they are not regarded as a group of variables and they fall under different components in the PCA.

Reliability of the Gameplay Elements items and explored constructs are evaluated based on their Cronbach's alpha value. A Cronbach's alpha of .923 yielded for the whole 36 items. This result suggests that there is redundancy within the items.

For the latent variables explored through the factor analysis, Cronbach's alpha values are as follows. Reliability of the "Avatar", "Secondary Gameplay" and "Playability" dimensions are below the threshold of .7, but Nunnaly [35] suggests that a Cronbach's alpha value above .6 is reliable for a freshly developed instrument. Although the items of "Audio Visual Elements" construct does not met this criterion, we decided to investigate these constructs' relation with GEQ constructs, whether to see it is possible measure the quality of a game through the player's evaluation of its gameplay elements.

TABLE VII. RELIABILITY INDICATORS FOR GAMEPLAY ELEMENTS DIMENSIONS

Factor	**Cronbach's Alpha**
Core Gameplay Elements	.833
Avatar Elements	.66
Audio Visual Elements	.547
Narrative Elements	.705
Customization Elements	.739
Secondary Gameplay Elements	.644
Playability Elements	.611

C. GEQ and Gameplay Elements

We employed a Pearson's correlation to assess the concurrent validity of our approach based on assessment of Gameplay Play Elements with the GEQ based evaluation of player engagement. Results, given in Table VII shows that correlations between Gameplay Elements dimensions and GEQ dimensions are significant in most cases at .01 level marked with an "*". However, these correlations are low or medium.

TABLE VIII. CORRELATIONS OF GEQ AND GAMEPLAY ELEMENTS

	GEQ Negative Effects	GEQ Positive Effects	GEQ Competence	GEQ Challenge	GEQ Immersion	GEQ Tension Attention	GEQ Flow
Core Gameplay Elements	.167*	-.352*	-.241*	-.148*	-.346*	.137*	-.317*
Avatar Elements	.119*	-.307*	-.224*	-.062	-.358*	.127*	-.265*
Audio Visual Elements	.153*	-.282*	-.211*	-.074	-.319*	.137*	-.241*
Narrative Elements	.156*	-.281*	-.208*	-.106*	-.245*	.109*	-.210*
Customization Elements	.130*	-.252*	-.196*	-.130*	-.252*	.103*	-.229*
Secondary Gameplay Elements	.166*	-.337*	-.283*	-.094*	-.328*	.167*	-.283*
Playability Elements	.194*	-.350*	-.221*	-.039	-.286*	.073	-.173*

"Negative Effects" dimension of GEQ correlates with all Gameplay Elements significantly, but the level of the correlations is quite low. "Positive Effects" dimension of GEQ correlates with "Playability Elements", "Core Gameplay Elements", "Secondary Gameplay Elements" at a medium level, while a little bit lower with others. "Competence" measures in GEQ are significantly related with all Gameplay Elements dimensions, however with a low level. "Challenge" does not significantly correlate with "Playability Elements", "Audio-Visual Elements" and "Avatar Elements", but correlate with others at a low level. Immersion measures of GEQ has the strongest correlation with Gameplay Elements dimensions, compared to other dimensions of GEQ. However, these correlations are mostly at a medium level. Tension/Attention has low correlations with Gameplay Elements, except the "Playability Elements". These correlations are positive, suggesting that satisfaction from dimensions of Gameplay Elements reduces the irritation, annoyance or frustration of user. Flow correlates with "Core Gameplay Elements" at a medium

level, and its correlation with other dimensions are slightly below medium, except "Playability Elements".

V. DISCUSSION

Results given above suggest that it is possible to uncover the underlying structure of gameplay elements with an EFA but the correlations between the gameplay elements and latent variables of GEQ are still low. This outcome is not surprising because it may not be possible to associate a single variable of GEQ, in example Flow, with a single gameplay factor such as the "Core Gameplay Elements". Flow is defined as an optimal state of consciousness where we feel our best and perform our best, so it is more meaningful to associate it with a group of gameplay elements, rather than trying to explain it with a single factor.

Our item set for evaluating the Gameplay Elements does not have an acceptable level of reliability on its dimensions. Some of the constructs emerged through EFA were not clear, while some items have loaded on more than one construct. This may be depending on the limitations of the study or measurement tools employed within the study. Also, the methods used for normalizing the skewed data influences results.

Our study could not conclusively determine that a game evaluation method involving the players can be based on assessment of gameplay elements. On the other hand, our results provided some evidence that players can assess game play elements which are also correlated with their experience with the game. Future work relying on gameplay elements oriented evaluation approach may benefit our factor analysis results to develop a better set of items.

As a measurement tool employed in this study, GEQ was not clearly evaluated for its psychometric quality before. However, our study contributed to the development of GEQ, providing evidence for its reliability and validity, while exposing some flaws in its unidimensionality. Developers of GEQ or any future scale for evaluation of gaming experience may consider our findings for their work. Future studies may also focus on other variables that affects gaming experience, in addition to gameplay elements.

REFERENCES

[1] P. Wright and J. McCarthy, Technology as experience, MIT Press, 2004.

[2] M. Slater, M. Usoh and A. Steed, "Taking Steps: The Influence of a Walking Metaphor on Presence in Virtual Reality", ACM Transactions on Computer-Human Interaction (TOCHI) 2(3) September, 1995, 201-219.

[3] M. Csikszentmihalyi, Flow: The psychology of optimal experience, New York: Harper & Row, 1990.

[4] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments", Presence: Teleoperators and Virtual Environments, 6 (6), 1997, pp. 603-616.

[5] E. H. Calvillo-Gámez, P. Cairns, A. L. Cox, "Assessing the core elements of the gaming experience", in R. Bernhaupt (ed), Evaluating user experience in games, Springer, London, UK, 2010, pp 47–71.

[6] R. Hunicke, M. LeBlanc and R. Zubek, "MDA: A Formal Approach to Game Design and Game Research", Proceedings of AAAI04 WS on Challenges in Game AI, 2004, pp. 1-5.

[7] J. Wiemeyer, L. Nacke, C. Moser and F. `Floyd' Mueller, "Player Experience", in Serious Games: Foundations, Concepts and Practice, Springer International Publishing, 2016, pp. 243–271.

[8] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach", Motivation and Emotion 30(4), 2006, pp. 344–360. http://doi.org/10.1007/s11031-006-9051-8.

[9] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being", Am Psychol 55(1), 2000, pp. 68–78.

[10] J. M. Keller, "Development and use of the ARCS model of instructional design", JID 10(3), 1987, pp. 2–10.

[11] J. M. Keller, Motivational design for learning and performance: the ARCS model approach, Springer Science & Business Media, New York, 2009.

[12] P. Sweetser and P. Wyeth, "Game flow: a model for evaluating player enjoyment in games", ACM CIE 3(3), 2005.

[13] J. Takatalo, J. Hakkinen, J. Kaistinen and G. Nyman, "User experience in digital games: differences between laboratory and home", Simulation and Gaming 42(5), 2011, pp. 656–673.

[14] K. Poels, Y. A. W. de Kort, W. A. IJsselsteijn, "FUGA—The fun of gaming: measuring the human experience of media enjoyment", Deliverable D3.3: game experience questionnaire, TU Eindhoven, Eindhoven, The Netherlands, 2008.

[15] H. Desurvire and C. Wiberg, "Game usability heuristics (PLAY) for evaluating and designing better games: the next iteration", in A. A. Ozok and P. Zaphiris (eds), Online communities and social computing, LNCS 5621, Springer, Berlin, 2009, pp. 557–566.

[16] B. G. Witmer and M. J. Singer, " Measuring presence in virtual environments: a presence questionnaire", Presence: Teleoperators and Virtual Environments, 7(3), 1998, pp. 225-240.

[17] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, F. Jäncke, S. Böcking, H. Schramm, A. Gysberg, T. Hartmann, C. Klimmt, J. Laarn, N. Ravaja, A. Sacau, T. Baumgartner, P. Jäncke, "MEC spatial presence questionnaire (MEC-SPQ): short documentation and instructions for application", Report to the European community, project presence: MEC (IST-2001-37661), 2004.

[18] T. Hartmann, W. Wirth, H. Schramm, C. Klimmt, P. Vorderer, A. Gysbers, S. Böcking, N. Ravaja, J. Laarni, T. Saari, F. Gouveia, A. M. Sacau, "The spatial presence experience scale (SPES)", JMP, 2015, doi:10.1027/1864-1105/a000137.

[19] Y. A. de Kort, W. A. IJsselsteijn, K. Poels, "Digital games as social presence technology: Development of the social presence in gaming questionnaire (SPGQ)", in Proceedings of PRESENCE, 2007.

[20] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, J. N. Pidruzny, "The development of the game engagement questionnaire: a measure of engagement in video game-playing", Journal of Experimental Social Psychology, 45, 2009, pp. 624–634.

[21] IJsselsteijn, W., van den Hoogen, W.H.M., Klimmt, C., de Kort, Y.A.W., Lindley, C., Mathiak, K., Poels, K., Ravaja, N., Turpeinen, M. and Vorderer, P. "Measuring the experience of digital game enjoyment", In Proc. of Measuring Behavior, Maastricht Netherlands, 2008, pp. 88–89.

[22] Johnson, D., Watling, C., Gardner, J., and Nacke, L. E. "The edge of glory: The relationship between metacritic scores and player experience", In Proc. of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play, CHI PLAY '14, ACM (New York, NY, USA, 2014), 141–150.

[23] Norman, K. L. "GEQ(game engagement/experience questionnaire): a review of two papers", Interacting with Computers 25, 4, 2013, pp. 278–283.

[24] Brühlmann, F. and Schmid, G. "How to Measure the Game Experience? Analysis of the Factor Structure of Two Questionnaires", CHI'15 Extended Abstracts, April 18–23, 2015, Seoul, Republic of Korea. ACM 978-1-4503-3146-3/15/04.

[25] Tabachnick, B.G. and Fidell, L.S. Using multivariate statistics (5th ed.). Boston: Allyn and Bacon, 2007.

[26] Chin, W.W. “The partial least squares approach to structural equation modeling”, In Marcoulides G.A. (Ed.), Modern methods for business research (pp. 295–336). Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

[27] Chin, W.W. “ How to write up and report PLS analyses”, In Vinzi, V.E., Chin, W.W., Henseler, J. and Wang, H. (Eds.), Handbook of partial least squares (pp. 655–690). Berlin Heidelberg: Springer, 2010.

[28] Lewis, B.R., Templeton, G.F. and Byrd, T.A. “A methodology for construct development in MIS research”, European Journal of Information Systems, 14 (4), 2005, pp. 388–400.

[29] Straub, D., Boudreau, M., & Gefen, D. “Validation guidelines for IS positivist research”, Communications of the AIS, 13, 2004, pp. 380–427.

[30] Fornell, C. and Larcker, D. F. 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18, 1981.

[31] Lohmöller, J.-B. Latent variable path modeling with partial least squares. Heidelberg, Germany: Physica, 1989.

[32] Hair, J. F., Jr., Ringle, C. M., & Sarstedt, M. “PLS-SEM: Indeed a silver bullet”, Journal of Marketing Theory and Practice, 19(2), 139–151, 2011.

[33] Howell, D.C. Statistical methods for psychology (6th ed.). Belmont, CA: Thomson Wadsworth, 2007.

[34] Kaiser, H.F. “An index of factorial simplicity”, Psychometrika, 39, 1974, pp. 31-36.

[35] Nunnally, J. and Bernstein, H. Psychometric Theory. McGraw Hill. 1978