International Journal of Human-Computer Studies 118 (2018) 38–46

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ)

Daniel Johnson[a,*], M. John Gardner[b], Ryan Perry[c]

[a] Queensland University of Technology (QUT), GPO Box 2434, Brisbane, QLD 4001, Australia
[b] CSIRO Land and Water, Ecosciences Precinct, GPO Box 2583 Brisbane QLD 4001, Australia
[c] Department of Marketing and Management, Faculty of Business and Economics, The University of Melbourne, Parkville, VIC 3010, Australia

ARTICLE INFO

Keywords:
Videogame player experience
Scale validation
Confirmatory factor analysis

ABSTRACT

Accurate measurement of the player experience in videogames is key to understanding the impacts of videogame play, designing and developing engaging videogames, and effectively applying game design principles in other fields. A large number of player experience questionnaires are available, but in most cases empirical validation of the scales is limited or absent. Two of the most commonly used scales are the Player Experience of Need Satisfaction (PENS) and the Game Experience Questionnaire (GEQ). Both scales were developed using a rational-theoretical approach, but neither scale has had formal factor-analytic studies published, limiting our capacity to judge the empirical validity of the scales. We present detailed exploratory and confirmatory factor analyses of both scales based on responses from a sample (n = 571) of videogame players. The GEQ is partially supported (using a revised factor structure); the PENS is largely supported (with a more minor revision of the factor structure). We provide suggestions for the most effective use of both scales in future research.

1. Introduction

The ability to accurately assess the experience of players during gameplay has implications for building more enjoyable (and successful) videogames, understanding the impact of videogame play, and effectively applying game design principles in other fields. As a result, being able to measure player experience is an increasing area of focus among both videogame researchers and developers (Cairns et al., 2006; Denisova et al., 2016; Drachen and Canossa, 2009; Nacke and Drachen, 2011). While the field is steadily advancing, several challenges remain in terms of valid and reliable measurement of the player experience (PX).

Common approaches to measuring PX range from more objective approaches such as physiological measurements and in-game behaviour analytics to more subjective techniques including interviews, focus groups, in-game probes and questionnaires. Of these, questionnaires offer a low-cost means of inquiring about players' experience of videogame play. Although subjective, questionnaires aim to provide a consistent assessment of aspects of the PX across individuals. However, at present a large variety of questionnaires are being employed, often in the absence of a clear indication of their relative empirical validity and reliability. Moreover, the use of a large variety of questionnaires decreases the ability to compare outcomes across studies.

Following Denisova et al. (2016) our aim is not to suggest that one single questionnaire could assess all aspects of PX. However, we do propose there is value in identifying the more effective questionnaires as well as evaluating the extent to which more commonly employed questionnaires are accurately measuring the aspects of the PX they purport to assess. The current study presents an analysis of two of the more commonly used PX questionnaires; the Player Experience of Need Satisfaction (Ryan et al., 2006) and the Game Experience Questionnaire (IJsselsteijn et al., 2007, 2008a,b).

1.1. Measuring player experience

The range of approaches to measuring PX are associated with differing strengths and weaknesses. Physiological measures of the PX offer greater objectivity but are generally more costly (temporally and financially) and can be harder to interpret (Lankoski and Bjork, 2015). Commonly employed physiological assessments of PX include heart-rate, respiration rate, electromyography (muscle activation), electro-encephalography (cortical activity) and electrodermal activity (skin conductivity) (Nacke, 2013). In-game behaviour analytics (or telemetry) similarly offer objectivity but as emerging area of focus, challenges remain regarding how best to reduce complexity via profiling and how to link in-game behaviour to subjective experience (Sifa et al.,

* Corresponding author.
E-mail addresses: dm.johnson@qut.edu.au (D. Johnson), john.gardner@csiro.au (M.J. Gardner), ryan.perry@unimelb.edu.au (R. Perry).

https://doi.org/10.1016/j.ijhcs.2018.05.003
Received 2 August 2017; Received in revised form 22 February 2018; Accepted 11 May 2018
Available online 17 May 2018
1071-5819/

2017).

More subjective approaches to assessment of PX have centred around interviews, focus groups, in-game probes and questionnaires (Lankoski and Bjork, 2015). These techniques are relatively low-cost alternatives with less challenges around interpretation than physiological measures of telemetry. Interviews, focus groups and in-game probes offer relatively in-depth insights but are difficult to deploy at larger scale. In contrast, questionnaires can be easily deployed to very large groups and while providing less depth of insight than other subjective approaches, allow for a focus on specific aspects of PX (Lankoski and Bjork, 2015).

At present, in the PX field there is a large variety of options for assessing PX via questionnaires (Denisova et al., 2016). Existing reviews (Denisova et al., 2016; Mekler et al., 2014) identify questionnaires designed to measure a range of PX related constructs including presence, immersion, engagement, need satisfaction, enjoyment, fun, frustration, challenge, cognitive absorption and flow. This in part reflects the multi-faceted nature of PX as well as differing conceptual theories regarding motivations for play (e.g., Ryan et al., 2006) or the most relevant components of the PX (IJsselsteijn et al., 2007, 2008a,b). However, in some cases, multiple options exist for measuring the same (or very similar) constructs (e.g., flow questionnaire vs. flow in the GEQ). Moreover, in part as a result of the relative youth of the field, very few of the existing questionnaires have been fully empirically validated. Scales that have not been formally validated rely on presumptions that they accurately reflect the constructs they purport to measure, and that subscales measure constructs that are truly distinct. Until these presumptions are tested, studies that use these scales may be drawing inaccurate conclusions due to unaccounted measurement error. If the scale is used in a series of studies then any error may be compounded.

1.2. PENS and GEQ

Two of the more commonly used questionnaires in the PX field are the Player Experience of Need Satisfaction (PENS - Ryan et al., 2006) and the Game Experience Questionnaire[1] (GEQ - IJsselsteijn et al., 2007, 2008a,b; Poels et al., 2007). The PENS draws on the well-established theory of motivation – Self-Determination Theory (Deci et al., 1991; Ryan and Deci, 2000) – which describes the manner in which experiences satisfy universal needs (competence, autonomy and relatedness) in people. In short, experiences that satisfy these needs are more likely to result in intrinsic motivation to engage in that activity as well as positive wellbeing outcomes. The PENS measures these three universal needs as well as two additional constructs: presence/immersion and intuitive controls (see method section for further detail). In contrast, the GEQ does not draw on a specific theory. Rather, it is based on conceptual accounts of the PX and focus-groups conducted with videogame players. The full questionnaire includes scales designed to measure positive affect, negative affect, frustration, flow, challenge, immersion and competence (see method section for further detail). Both scales were developed using what is termed a rational-theoretical approach; one in which scale developers write items that are subjectively consistent with their theoretical understanding of the intended scale construct(s) (Simms and Watson, 2007).

1.2.1. GEQ

The GEQ has been used in a variety of research settings. Much of the research has assessed the impact of whether the other 'player(s)' in the game are computer-controlled or human-controlled (Gajadhar et al., 2009, 2010, 2008a,b, Johnson et al., 2015b) as well as whether they are co-located (Gajadhar et al., 2010) and act cooperatively or competitively (Souders et al., 2016). Other studies have focused on specific influences on PX including the context of play (at home compared to on public transport (Engl and Nacke, 2013); the genre of the game (Johnson et al., 2015a, 2012); the personality of the player (Johnson et al., 2012); the degree of challenge in the level (Nacke et al., 2010b; Nacke and Lindley, 2008); the type of controller (Nacke, 2010) or button layout (Ellick et al., 2013); sound effects and music (Nacke et al., 2010a); and network latency (Hohlfeld et al., 2016). Outside of factors that influence PX in recreational digital games, research has also assessed how the PX (measured via the GEQ) relates to review scores (Johnson et al., 2014). In non-digital settings research has explored the PX for older adults in a tabletop game (Al Mahmud et al., 2008, 2010)

Research using the GEQ has also explored PX in non-recreational games. This has extended to comparing PX in serious games to recreational games (Rotoly De Lima et al., 2015) as well as exploring the PX of older adults in a game designed for cognitive training and screening (Boletsis and McCallum, 2016a,b) as well as exergames (Gerling et al., 2011; Liukkonen et al., 2015). Comparative studies within the domain of serious games have assessed the impact of type of avatar (Kao and Harrell, 2015); colour of avatar (Kao, 2016); collaboration (Oksanen, 2013); and type of feedback (Kao and Harrell, 2016).

1.2.2. PENS

The PENS has been used in many of the same settings as the GEQ. Within recreational game play, the PENS has been used to assess the impact of whether the other 'player(s)' in the game are computer-controlled or human-controlled (Johnson et al., 2015b; Tamborini et al., 2010; Vella et al., 2015b); the genre of the game (Johnson and Gardner, 2010; Johnson et al., 2015a); the players' personality (Johnson and Gardner, 2010); and sound effects (Robb et al., 2017). Outside of factors that influence the PX, in line with the GEQ, research has also assessed how the PX (measured via the PENS) relates to review scores (Johnson et al., 2014). In non-recreational settings, the PENS has been used (as has the GEQ) to assess the PX in exergames (Peng et al., 2012). Outside of topics also explored using the GEQ, the PENS has been used to assess the impact of emotionally rewarding play experiences (Bopp et al., 2016); storytelling (Bormann and Greitemeyer, 2015); graphic fidelity (Gerling et al., 2013); cognitive action (Inchamnan and Wyeth, 2013); and the impact of in-game violence (Przybylski et al., 2009).

Given the difference in focus between the GEQ and PENS, and the PENS roots in self-determination theory, it is unsurprising the PENS has been used in various setting related to motivation, identity and wellbeing. With respect to motivation the PENS has been used to explore whether need satisfaction mediates the relationship between self-esteem and post-game motivation (Birk et al., 2015). In terms of player identity, research has used the PENS to explore the impacts of the players sense of identify with their in-game avatar (Birk et al., 2016) as well as the players concept of their ideal self (Przybylski et al., 2012). Research has also explored how the experience of need satisfaction (measured via the PENS) relates to emotional, psychological and social wellbeing (Vella et al., 2013, 2015a).

1.2.3. Scale validation

In light of the extent to which both the GEQ and PENS have been use in PX research, it is somewhat surprising that full empirical validation for either scale is yet to be published. That said, support for the structure and effectiveness of the PENS is provided by the scale authors across four studies employing the scale to assess need satisfaction with different games and audiences (Ryan et al., 2006). However, full formal validation of the scale has not been published. Similarly, while the authors of the GEQ provide detailed information on the scale development process and exploratory evidence of criterion validity

[1] The original validation paper for the PENS has been cited 1351 times (according to Google scholar accessed 4th December 2017) and the collection of papers and manuscripts outlining the details of the GEQ have been cited 594 times.

(Poels et al., 2007), the factorial structure of the scale has not yet been fully, formally validated. Notably, a work-in-progress publication by Brühlmann and Schmid (2015) describes an examination of the underlying factorial structure of the PENS (with the relatedness subscale excluded) and GEQ, also assessed via exploratory Factor Analysis (EFA). The authors explored the use of these scales with two games (played for 3 minutes each) and found support for the proposed structure of the PENS but were unable to verify the proposed structure of the GEQ.

The current study is designed to provide further insight regarding the empirical validity of both scales. The widespread use of both scales alongside a steady increase in interest in understanding the player experience mean that greater depth of insight regarding how both scales perform is timely. Moreover, as noted above, both scales were initially developed using a rational-theoretical approach. In the current study we assess the scales using exploratory and confirmatory factor analysis, since this approach has been argued to offer greater evidence for discriminant validity (Simms and Watson, 2007), and many psychometricians recommend an integrative approach encompassing both rational-theoretical item generation by field experts, and factor analysis to ensure internal scale consistency (e.g., Clark and Watson, 1995; Simms and Watson, 2007). An example of the application of this approach in the player experience domain is the construction of the Game-Specific Attribution Questionnaire (Depping and Mandryk, 2017). The current study builds on the work conducted by Brühlmann and Schmid (2015) by assessing the full PENS (including the relatedness subscale), analysing responses across a wide variety of games played for extended periods and employing both exploratory and confirmatory factor analysis with an appropriately large sample size (Van Hooris and Morgan, 2007).

2. Study 1

Study 1 examined the factor structure of the GEQ.

2.1. Method

2.1.1. Participants and procedure

Participants were 573 university students (468 male, 102 female and 3 unspecified), with a mean age of 20.85 ($SD = 5.87$). Participants were recruited on the basis of their interest in videogames, and most (82%) were university students. Participants were recruited via advertisement in videogame related university courses, email lists of people who has previous participated in videogame related research, Facebook, online forums (related to gaming) and snowball sampling. As part of a larger survey, participants were asked to complete the PENS and the GEQ (core module). Twenty-one participants from the sample did not complete the GEQ, leaving 552 respondents for analysis of the scale.

On average, the participants played videogames 16.6 h per week (SD = 12.5; range = 1–100 h) and played their favourite game for an average of 9.5 h per week (SD = 9.6; range = 1–100 h). Participants' favourite games came from a range of videogame genres. The most common genre of participants' current favourite game was first-person shooter games (24.9%), followed by action role-playing games (13.6%), action adventure games (11.9%), role-playing games (9.6%), massively multiplayer online role-playing games (8.7%), multiplayer online battle arenas (5.8%), real-time strategy (5.6%), and various other genres (19.9%).

Participants were asked to complete an online survey which asked them to report on their player experience with respect to their "current favourite game" (considering only games they had played in the last six months). Because participants were involved via an online survey completed at a time of their choosing, there was no way to control for the time since they had last played the game. To deal with this issue, a guided-recall process was used to prime respondents before they answered questions about their gaming experiences. The accuracy of episodic memory can be substantially improved by guiding respondents to recall various aspects or elements of an experience (including setting, sequencing and recall of various senses), rather than just asking them to "recall what happened". This approach of systematically guiding recall, initially developed by Fisher and Geiselman (1992), has been demonstrated via meta-analysis to be an effective mechanism of improving recall (Memon et al., 2010), and has been used to generate a number of specific memory-enhancement inductions (e.g., Madore et al., 2015). In the present study, participants were instructed to "Think back to the last time you played 'name of current favourite game'. Try to remember where you were, what was happening in the game, and how you felt at the time. In the box below please explain (in about 30 words or less) which part of the game you were playing and what was happening."

2.1.2. Measures

The GEQ is a 33-item scale which is designed to measure game players' experience across seven dimensions: Immersion, Flow, Competence, Positive and Negative Affect, Tension, and Challenge. Items in the questionnaire are presented as statements, which are rated by respondents to indicate their experience while playing the game. Ratings are made on a 5-point Likert scale with anchors as follows: "0 - not at all", "1 - slightly", "2 - moderately", "3 - fairly", and "4 - extremely".

- Immersion (also referred to as Sensory and Imaginative Immersion) is assessed with six items reflecting aspects of how strongly connected with the game players felt.
- Flow is assessed with five items, which indicate whether players lost track of their own effort and or the passage of time during the game.
- Competence is assessed with five items relating to how well players judged their own performance against the game's goals.
- Positive affect (5 items) related to positive emotional experiences.
- Negative affect (4 items) related to negative emotional experiences.
- Tension (also referred to as Tension/Annoyance) is assessed with 3 items related to these specific negative emotions.
- Challenge is assessed with five items, which indicate the degree to which players found the game to be difficult or challenging.

2.2. Results and discussion

To validate findings across exploratory and confirmatory factor analysis procedures, the sample was randomly assigned to the EFA group ($N = 226$), and the CFA group ($N = 326$). Models were estimated using maximum likelihood parameter estimates with standard errors robust to non-normality (using a sandwich estimator) and an oblique geomin rotation. Missing data was estimated using full information maximum likelihood.

Descriptive statistics for the originally proposed seven-factor GEQ are presented in Table 1 below. In the exploratory analysis, the seven-factor structure proposed by the GEQ scale authors was not fully supported. The eigenvalue scree plot suggested a five-factor solution provided the best fit to the data. This solution was supported by parallel analysis using 500 random datasets with the same number of observations and variables, which indicated that only the first five

Table 1
Descriptive Statistics for the GEQ sub-scales (as originally proposed).

Subscale	Min.	Max.	Mean.	SD.	Skew.	Kurt.
Competence	1.80	5.00	3.937	0.619	−0.478	0.075
Sensory & imag. imm.	1.17	5.00	3.779	0.772	−0.538	−0.193
Flow	1.00	5.00	3.499	0.896	−0.380	−0.302
Tension	1.00	5.00	1.994	0.892	1.117	0.968
Challenge	1.00	5.00	3.002	0.776	0.051	−0.308
Neg. affect	1.00	4.00	1.764	0.572	0.949	0.931
Pos. affect	2.40	5.00	4.255	0.564	−0.826	0.394

Table 2
Item content and factor loadings for the GEQ exploratory factor analysis in Study 1.

Items		Factors				
Code	Label	1	2	3	4	5
P01	I felt content	**0.576**	−0.008	0.060	0.022	0.141
P06	I felt happy	**0.595**	0.036	0.149	0.108	0.010
P14	It felt good	**0.692**	−0.039	0.047	0.117	−0.014
P04	I thought it was fun	**0.738**	−0.166	−0.048	−0.065	−0.007
P20	I enjoyed it	**0.645**	−0.200	0.012	−0.033	0.107
N07	It gave me a bad mood	−0.088	**0.742**	0.097	−0.101	−0.006
T22	I felt annoyed	−0.068	**0.796**	0.008	−0.004	0.009
T29	I felt frustrated	0.085	**0.703**	−0.035	0.052	−0.012
T24	I felt irritable	−0.047	**0.790**	−0.032	−0.049	0.024
H23	I felt pressured	0.063	**0.408**	−0.077	0.085	−0.059
H32	I felt time pressure	0.028	**0.360**	−0.035	0.275	−0.126
H33	I had to put a lot of effort into it	0.174	**0.306**	0.027	0.179	0.150
C02	I felt skilful	0.235	0.097	**0.485**	0.064	−0.056
C15	I was good at it	−0.008	−0.005	**0.883**	−0.101	0.028
C10	I felt competent	0.158	−0.017	**0.671**	0.011	0.134
C21	I was fast at reaching the game's targets	−0.021	−0.091	**0.482**	0.163	−0.067
F05	I was fully occupied with the game	0.170	0.038	0.056	**0.647**	−0.042
F25	I lost track of time	0.046	0.104	−0.116	**0.639**	0.077
F13	I forgot about everything around me	−0.085	−0.029	−0.017	**0.782**	0.015
F28	I was deeply concentrated in the game	0.164	0.079	0.134	**0.543**	−0.029
F31	I lost connection with the outside world	−0.269	−0.089	0.017	**0.858**	0.047
I03	I was interested in the game's story	−0.006	−0.060	−0.029	−0.030	**0.810**
I18	I felt imaginative	0.188	0.016	−0.003	0.162	**0.419**
I19	I felt that I could explore things	−0.048	0.081	0.057	0.030	**0.571**
I30	It felt like a rich experience	0.201	−0.015	−0.003	0.263	**0.430**

Notes: The code number corresponds to the items number from the original GEQ scale. The code letters represent the originally hypothesized factor labels: P = positive affect; N = negative affect; T = Tension; H = Challenge, C = competence; F = flow; I = immersion. Bolded coefficients highlight factor loadings greater than ± 0.3 (i.e., our cutoff level for retaining an indicator).

eigenvalues from the sample correlation matrix were larger than eigenvalues generated via chance. From the five-factor EFA solution, we inferred that *negative affect, tension/annoyance*, and three of the five *challenge* items loaded onto a single factor – a construct herein referred to as *negativity*. Next, items that exhibited the weakest loading on a given factor, or that cross-loaded on two or more factors, were sequentially removed. Items were removed one at a time, and the factor matrix was recalculated for the remaining items at each step until all items loaded strongly (i.e., >0.3) and solely one of the five retained factors. Deleted items tended to cross-load on both their theoretically-intended factor along with *positive affect*. The exception was the item labelled "I thought about other things", which overlapped (negatively) with *flow*. In total, eight items were removed, leaving 25 items in the analysis. The five factor EFA solution is presented in Table 2.

Initially (as a point of comparison), a confirmatory factor analysis (CFA) was conducted on the full sample assuming the original proposed factor structure of the GEQ. When assessing model fit, Hu and Bentler (1999) provide widely accepted cut-off criteria for determining acceptable fit: Root Mean Square Error of Approximation (RMSEA) should be less than 0.06, Standardized Root Mean Residual (SRMR) should be less than 0.08, and Comparative Fit Index (CFI) should be greater than 0.95. The CFA assuming the original proposed factor structure of the GEQ revealed poor model fit (fit indices as follows: $\chi 2 = 1061.87$ (df = 474), RMSEA = 0.062, SRMR = 0.076,

Table 3
Model fit indices for the revised GEQ scale (GEQ-R) confirmatory factor analysis in Study 1.

Model	χ^2	df	RMSEA	RMSEA 90% CI	SRMR	CFI
Null	2571.67	300	0.152	0.147, 0.158	0.216	–
GEQ-R	504.27	265	0.053	0.046, 0.060	0.068	0.895
GEQ-R with covariances	378.21	261	0.037	0.029, 0.045	0.062	0.948

CFI = 0.834). Following this, CFA was conducted (using responses from the second half of the sample) with the modified 5-factor structure identified by the EFA. In the CFA conducted on the revised 5-factor structure, the RMSEA and SRMR indicated good fit, however the CFI was below the generally accepted cutoff. As Kenny (2015) has noted however, an RMSEA value less than 0.158 in the null model will always result in a CFI that is too small in specified models. Furthermore, Mplus modification indices suggested freeing a total of four parameters by allowing covariances among the error terms of the observed variables. Fit was improved as a result of these changes, which suggests that these indicators are substantially related to one another beyond their common association with the latent variables. Fit indices for each model are presented in Table 3 and the final GEQ-R model is presented in Fig. 1.

We provide an updated factor structure for the GEQ in which (a) negative affect, tension annoyance and challenge reflect a single *negativity* factor, and (b) a number of items that cross-loaded with positive affect are removed. The present factor analyses suggest that the remaining five factors of the revised scale (GEQ-R) are distinct and valid constructs.

Although the model fit indices were acceptable, these were close to generally accepted rules of thumb (Hu and Bentler, 1999) with apparent room for improvement. Most critically, the RMSEA for the null model was unusually low, resulting in a small CFI value for the hypothesized model (Kenny, 2015). CFI is an index of how well an observed model fits the data compared to an alternative model – in this case the null model (in which the observed variables all co-vary with one another but otherwise exhibit no higher-order factor structure). A low RMSEA score for the null model indicates that the items are not particularly good indicators of any higher-order factor structure, and the resulting low CFI score for the GEQ-R indicates that although the specified factors are distinct and valid constructs, the items are not especially good indicators of their hypothesized latent construct. Hu and Bentler (1999) noted that RMSEA and CFI are most sensitive to models with misspecified factor loadings (whereas SRMR is most sensitive to misspecified covariances between the factors). Indeed, a number of scale items appear to be strongly related to one another beyond a common association with their respective factor. Model fit improved considerably when these covariances were included.

3. Study 2

Study 2 examined the factor structure of the PENS.[2]

3.1. Method

3.1.1. Participants and procedure

Here we used the same sample as described in Study 1, this time

[2] While we are not able publish the details of the individual PENS items in the current paper (as permission to use the scale in academic settings does not extend to publishing individual items), the numbering we have used throughout matches that provided in the PENS manual. Thus, readers with access to the PENS scale are able to identify the items being referred to in this paper. Readers interested in obtaining a copy of the PENS should contact the scale authors directly.

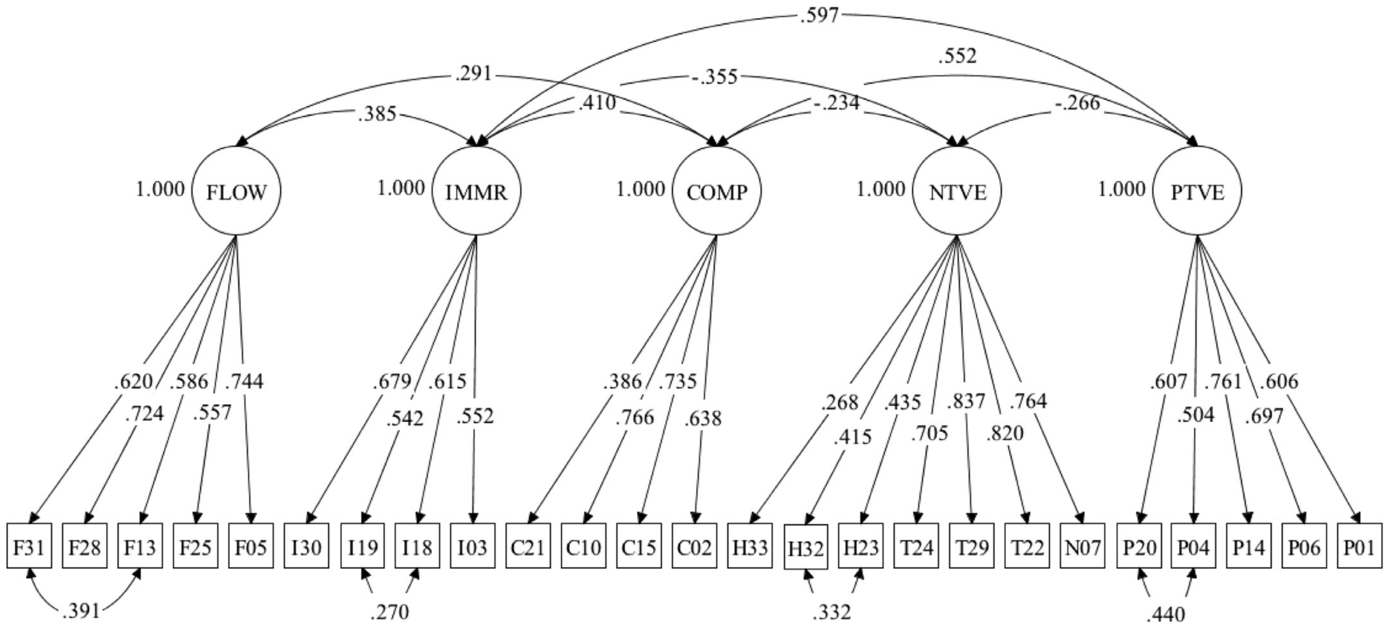

Fig. 1. Standardized parameter estimates for the factor structure of the revised GEQ (GEQ-R). Squares indicate the 26 manifest variables (scale items retained from the original GEQ). Ellipses indicate the five latent factors that correspond to each independent subscale of the GEQ-R. IMMR = immersion, COMP = competence, NTVE = negativity, and PTVE = positive affect. All factor loadings, correlations among the latent factors, and residual variances (not shown in the figure) were significant to $p < 0.01$.

Table 4
Descriptive Statistics for the PENS sub-scales (as originally proposed).

Subscale	Min.	Max.	Mean.	SD.	Skew.	Kurt.
Competence	2.00	7.00	5.796	0.944	−0.948	1.137
Autonomy	1.00	7.00	5.522	1.098	−0.811	0.692
Relatedness	1.00	7.00	3.954	1.489	0.056	−0.552
Presence & immersion	1.11	7.00	4.296	1.313	−0.083	−0.695
Intuitive controls	2.00	7.00	5.865	0.901	−0.895	1.001

using the full sample ($N = 573$). The presentation order of the two scales was counterbalanced, so that half of the participants completed the GEQ first, and half completed the PENS first. Also as in Study 1, the sample was randomly split across EFA ($N = 282$) and CFA ($N = 291$) groups. We also used the sample estimator and missing data procedure from Study 1.

3.1.2. Measures

The PENS is a 21-item scale designed to measure game players' experience across five dimensions: Competence, Autonomy, Relatedness, Presence/Immersion, and Intuitive Controls. Items in the PENS are presented to respondents as statements about their game experience, which are rated by respondents on a 7-point Likert scale ranging from "1 – do not agree" to "7 – strongly agree".

- Competence is assessed with three items reflecting aspects of how capable the players felt. There's strong conceptual overlap between this scale and the GEQ competence scale.
- Autonomy is assessed with three items, indicating the extent to which players experienced freedom and choice in the game. There's no clear overlap between this subscale and any of the PENS constructs.
- Relatedness is assessed with three items assessing the extent to which players feel connected to "other players in the game". This construct does not overlap with any constructs in the GEQ; it's also notable that this construct is meaningless for games which do not feature inter-player interaction via online or local competitive or cooperative play.
- Presence/Immersion (9 items) is related to emotional engagement in the game. There is conceptual overlap between this construct and the Flow and Challenge subscales of the GEQ.
- Intuitive Controls is assessed with three items, which indicate the degree to which players found they could translate their choices into in-game actions. This construct overlaps with the Competence subscale of the GEQ.

3.2. Results and discussion

Descriptive statistics for the originally proposed 5-factor PENS are presented in Table 4 below. Consistent with the EFA process described for Study 1, our interpretation of the eigenvalues plot for the PENS items suggested a four factor model would provide the best fit to the data. Parallel analysis also supported a four factor solution. The factor loadings indicated that, of the five factors originally hypothesized by the scale's authors, (a) *competence* and *intuitive controls* items load on a single factor, and (b) three of the *presence* items load on a common factor with the three *autonomy* items. Items that exhibited the weakest loading on a given factor, or that cross-loaded on two or more factors, were sequentially removed as in Study 1. After removing two cross-loading items (P4 and P8), all remaining presence and autonomy items loaded solely on their hypothesized factors. At this stage, we also removed item P6 for cross-loading on both the *presence* and *relatedness* factors. The final four-factor solution is presented in Table 5.

Initially (as a point of comparison), a CFA conducted on the full sample assuming the original proposed factor structure of the PENS revealed poor fit (fit indices as follows: $\chi 2 = 418.75$ (df = 179), RMSEA = 0.070, SRMR = 0.068, CFI = 0.886). Following this, CFA was conducted (using the second half of the sample) with the modified 4-factor structure identified by the EFA. The resulting fit indices RMSEA and SRMR indicated reasonably good model fit, and the CFI was slightly below the cutoff criteria (Hu and Bentler, 1999). However a CFI > 0.90 is also considered acceptable by some (see Hooper et al., 2008), and this index may be less reliable than RMSEA in confirmatory contexts (Rigdon, 1996; Sobel and Bohrnstedt, 1985). The RMSEA for the null model was above Kenny's (2015) cutoff value of 0.158, suggesting that further model improvements could reliably produce a more acceptable

Table 5
Item codes and factor loadings for the PENS exploratory factor analysis in Study 2.

Codes	Factors			
	1	2	3	4
C01	**0.765**	−0.016	0.053	−0.061
C02	**0.755**	0.014	0.093	0.089
C03	**0.494**	0.029	−0.037	0.198
I01	**0.539**	−0.005	0.023	−0.178
I02	**0.428**	0.160	−0.070	0.061
I03	**0.636**	−0.072	−0.072	−0.003
P01	0.035	**0.778**	−0.009	0.046
P02	−0.051	**0.773**	−0.013	0.046
P03	0.014	**0.867**	−0.016	0.007
P05	0.046	**0.515**	0.198	−0.003
P07	0.009	**0.629**	0.205	−0.049
P09	−0.056	**0.672**	0.189	0.01
R01	0.047	0.065	**0.817**	−0.013
R02	0.015	0.023	**0.845**	0.068
R03	0.054	0.166	**−0.560**	−0.015
A01	−0.025	−0.104	0.037	**0.696**
A02	0.083	0.091	0.006	**0.597**
A03	0.012	0.068	0.019	**0.602**

Notes: The code number refers to the items number from the original PENS scale. The code letters represent the originally hypothesized factor labels: C = Competence, I = Intuitive Controls, P = Presence, R = Relatedness, A = Autonomy. Bolded coefficients highlight factor loadings greater than ± 0.3 (i.e., our cutoff level for retaining an indicator). As the PENS is a scale used for commercial purposes we are not able to provide specific items details, however the numbering used is consistent with the PENS scale manual v1.6.

Table 6
Model fit indices for the PENS-R confirmatory factor analysis in Study 2.

Model	χ^2	df	MLR scaling	RMSEA	RMSEA 90% CI	SRMR	CFI
Null	1585.73	153	1.223	0.179	0.171, 0.187	0.264	–
PENS-R	252.45	129	1.171	0.057	0.047, 0.068	0.065	0.914
PENS-R with covariances	203.93	127	1.167	0.046	0.034, 0.057	0.063	0.946
PENS-R with covariances and cross-loadings	173.60	125	1.162	0.037	0.022, 0.049	0.051	0.966

CFI. As presented in Table 6, model fit was substantially improved when allowing covariances between the error terms that were suggested by the MPLUS modification indices.[3] The modification indices also suggested including additional cross-loading pathways of item R03 on the presence factor and item I02 on the autonomy factor. Freeing these parameters also substantially improved model fit. In the CFA sample (but not the EFA sample), item I02 was observed loading on both the competence/intuitive and the autonomy factors. Furthermore, in the CFA, item R03 was not significantly associated with its latent relatedness variable ($\beta = -0.177$, SE = 0.091, $p = 0.051$). However, this pathway became significant when the additional association with presence was included (see Fig. 2). Like the GEQ, the final PENS model showed some covariation between items.

As in Study 1, tests of discriminant validity (Henseler et al., 2015) indicated that the AVE for each latent construct in the PENS was greater than its maximum latent factor correlation.

[3] We were unable to model a further suggested covariance between items seven and eight as the model failed to converge; probably as this would have left too few indicators on the relatedness factor.

4. General discussion

In the present studies, we examined discriminant validity of two commonly used measures of video game PX for the first time. Our analysis largely confirmed the structure of the PENS as originally stipulated by the scale developers (Ryan et al., 2006) with the exception that competence and intuitive controls loaded on the same factor, and three items required removal. The factor structure for the GEQ, in contrast, deviated somewhat from the seven-dimension structure that it was developed to assess (IJsselsteijn et al., 2007, 2008a,b) with a reduction to five factors and with eight items removed. Our findings broadly align with those of Brühlmann and Schmid (2015) in largely supporting the proposed factor structure of the PENS. The primary difference between their findings and ours (regarding the discriminant validity of competence and intuitive controls) may relate to our study incorporating player responses to their current favourite game and Brühlmann and Schmid's requiring players to engage with two specific games (with which they had varying degrees of familiarity). As discussed further below, competence and intuitive controls may be more distinct factors when players are less familiar with a game. Our findings with respect to the GEQ also largely align with those of Brühlmann and Schmid. Neither study provides clear support for the proposed factor structure of the GEQ.

Acceptable absolute fit indices (RMSEA and SRMR) in both studies suggest that the PX dimensions of both scales as revised in the present studies are reasonably valid and independent. The low relative fit index (CFI) suggests that, although the latent factor structures for both scales are plausible, the items may be less than ideal indicators of these factors (see Hu and Bentler, 1999), and it was in fact the case that relative model fit improved to an acceptable level when item covariances were included in the model. Nevertheless, CFI may be more appropriate for assessing fit in exploratory rather than confirmatory contexts (Rigdon, 1996).

The four-factor PENS-R showed acceptable fit to the data. Model fit was also substantially improved when cross-factor pathways and covariances between some items were included. We therefore conclude that the PENS is a reasonable measure of its dimensions as originally stipulated. However, researchers should confirm their findings hold when dropping the cross-loading items R03 and I02 where it is appropriate to ensure independence between the presence and relatedness dimensions, and between the autonomy and competence/intuitiveness dimensions.

Our present analyses contribute to a commonly recommended integrated approach for psychometric scale validation (Clark and Watson, 1995; Simms and Watson, 2007). However, it is more ideal to take into consideration both discriminant validity and construct validity simultaneously and at all stages of scale development. In particular, the possible issue of mis-specified factor loadings in our results illustrates the importance of considering discriminant validity at the scale development stage of selecting appropriate scale items. Loevinger (1957) offered an early theoretical framework for scale development that emphasized the importance of couching the item selection phase in a theoretically predicted structure of hypothesized scale dimensions. Of relevance to improving PX scales, she recommended starting with formal construct definitions to guide the writing of relevant and representative items, as well as an over-inclusive initial item pool to allow all possible content of the construct(s) to be covered.

4.1. Caveats

Several elements of the research which impact the generalizability of our findings are noteworthy. Firstly, this research was conducted with a snowball sampling method that was seeded with university students in computing courses. As snowball sampling tends to tap populations that are similar to the seed sample (Wylie and Jolly, 2013), we cannot claim that our sample reflects the general population at

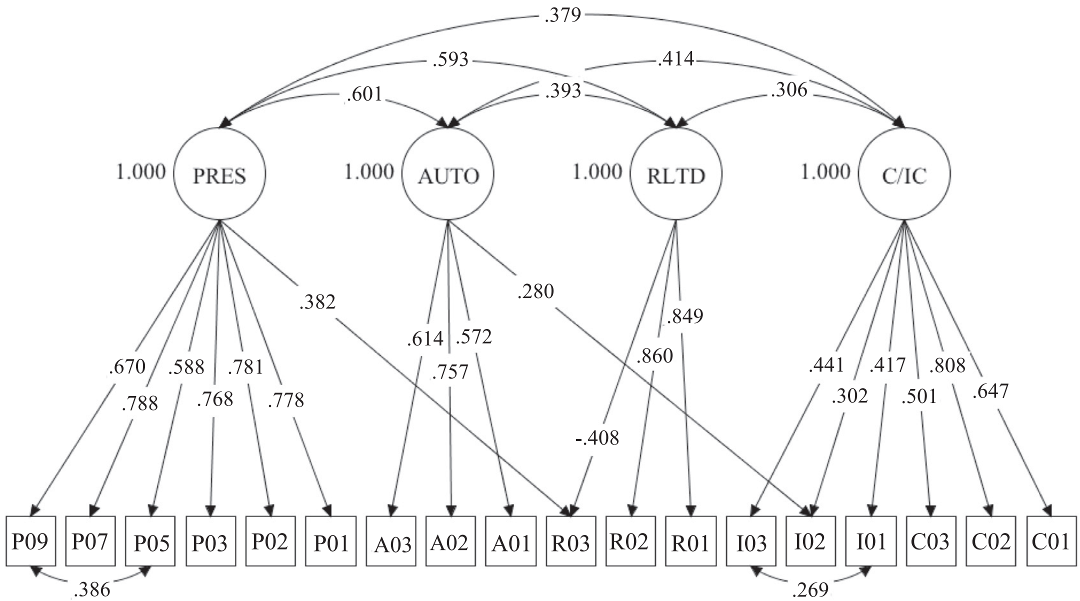

Fig. 2. Standardized parameter estimates for the factor structure of the PENS. Squares indicate the 18 manifest variables (scale items retained from the original PENS). Circles indicate the four latent factors that correspond to each independent subscale of the PENS. PRES = presence, AUTO = autonomy, RLTD = relatedness, and C/IC = competence/intuitive controls. All factor loadings, correlations among the latent factors, and residual variances (not shown in the figure) were significant ($p < 0.01$).

large, or even the general population of university students. Arguably, our sample provides a reasonable estimate of a population of well-educated computer game players within a young-adult age group.

Secondly, this research was conducted in the context of asking people to consider their most recent experience of playing their favourite game, which logically we would expect to reflect games that they largely found enjoyable and played extensively. Thus, our method would tend to exclude people from describing their experience with games that they played very little, and games that they did not enjoy (though this would not rule out what might be thought of as negative experiences e.g., challenge, frustration, tension as part of the gameplay experience). This has several implications for our sample:

- measures of positive affect, enjoyment and the like were probably positively biased;
- measures of negative affect and lack of enjoyment were probably negatively biased;
- measures of control and competence were probably positively biased.

In sum, it cannot be assumed that our proposed revised models would perform as effectively in contexts where people are also interested in measuring less positive videogame experiences.

These implications may explain why we did not find evidence for discriminant validity for the constructs of *Competence* and *Intuitive controls* in the PENS. It may be that these two factors are only really distinguishable when someone is learning to play a new game; in this context we would expect that a sense of controls as being intuitive would have a high degree of variability in early experiences, but that this variability would decrease with repeated practice as controls became more familiar. Conversely, perceived competence would be likely to start at a more consistently lower level, but would be expected to increase with repeated practice. Over time, we would expect that people would no longer be able to distinguish between their sense of competence and their experience of the controls as intuitive.

These implications may also explain why we found such strong overlap between three constructs in the GEQ that dealt with negative game experiences (items in the *negative affect, tension/annoyance*, and *challenge* scales all loaded onto a single *negativity* construct). Arguably, some of these scales (negative affect and tension/annoyance) experienced a floor effect which may have suppressed their variability and made it more difficult for our analyses to distinguish between them. In this case, when studying games that people consider their ‘current favourite’, we could ask whether there is less utility in pursuing such a fine-grained distinction between different types of negative gaming experience. Other than situations where different types of negative gaming experiences are the key focus of the research (e.g., Bopp et al., 2016), a scale that reflects more types of positive experience may be more informative when evaluating activities that individuals enjoy.

It is important to note that while our results may not be applicable for all the different games that people play, they do apply to the most common play context (people playing games they enjoy). Broadly applicable measures of gaming experience should be able to accurately reflect and discriminate between experiences even in situations where the majority of players are experiencing high positive affect, low negative affect and high levels of competence.

Thirdly, our study employed guided recall to prompt participants' recollection of the last time they played their current favourite game, and in most cases there was likely a delay between the last time the game was played and completing our survey. It should be noted that the GEQ manual recommends that the GEQ should be administered immediately after the game session has finished. It may be that both scales would perform differently when administered closer to the time of play and our findings cannot be assumed to apply in this context. However, our findings align with those of Brühlmann and Schmid (2015) who deployed the scales immediately after play. It is also worth noting that other researchers do employ these scales with a post-play delay (e.g., Boletsis and McCallum, 2016b; Souders et al., 2016). Regardless, it is valuable to determine the applicability of measures of the PX in situations where immediate data collection is not possible or preferable (e.g., where play has occurred in a naturalistic setting).

4.2. Conclusions

The GEQ's purported structure is not supported in a context of play where people are playing a game they presumably largely enjoy and a delay occurs between the time of play and response to the scale. It appears that a number of the items link to more than one construct – such items need to be removed or (in an updated version of the scale) need to be replaced with items that link uniquely to a single subscale. In addition, three of the constructs have so much overlap that they do not really deserve to be treated separately – we found a combined construct of *negativity* was more appropriate than the constructs of *negative affect, tension/annoyance* and *challenge*.

The PENS' purported structure is partially supported. The *presence, autonomy* and *relatedness* constructs were supported by the analyses. *Competence* and *intuitive controls* appeared to be a single construct, rather than two separate constructs as hypothesized in the PENS, subject to the caveats described above.

The following constructs appear to have some empirical support: flow (GEQ), immersion (GEQ), competence (GEQ and PENS), positive affect (GEQ), presence (PENS), autonomy (PENS) and relatedness (PENS). Four other constructs, as measured by the current scales, are not empirically supported, with indications that they are conceptually overlapping with other constructs: negative affect, tension/annoyance, challenge (which we suggest should be combined into a single negativity construct), and intuitive controls (which we suggest should be considered part of competence).

References

Al Mahmud, A., Mubin, O., Shahid, S., Martens, J.-B., 2008. Designing and evaluating the tabletop game experience for senior citizens. In: Proceedings of the 5th Nordic Conference on Human-Computer Interaction Building Bridges - NordiCHI '08. ACM Press, New York, New York, USA, pp. 403. https://doi.org/10.1145/1463160.1463205.

Birk, M.V., Mandryk, R.L., Miller, M.K., Gerling, K.M., 2015. How self-esteem shapes our interactions with play technologies. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15, pp. 35–45. https://doi.org/10.1145/2793107.2793111.

Birk, M.V, Atkins, C., Bowey, J.T., Mandryk, R.L., 2016. Fostering intrinsic motivation through avatar identification in digital games. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, Santa Clara, CA, pp. 2982–2995. https://doi.org/10.1145/2858036.2858062.

Boletsis, C., McCallum, S., 2016a. Augmented reality cubes for cognitive gaming: preliminary usability and game experience testing. Int. J. Serious Games 3 (1), 3–18. https://doi.org/10.17083/ijsg.v3i1.106.

Boletsis, C., McCallum, S., 2016b. Smartkuber: a serious game for cognitive health screening of elderly players. Games Health J. 5 (4), 241–251. https://doi.org/10.1089/g4h.2015.0107.

Bopp, J.A., Mekler, E.D., Opwis, K., 2016. Negative emotion, positive experience? Emotionally moving moments in digital games. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, Santa Clara, CA, pp. 2296–3006. https://doi.org/10.1145/2858036.2858227.

Bormann, D., Greitemeyer, T., 2015. Immersed in virtual worlds and minds: effects of in-game storytelling on immersion, need satisfaction, and affective theory of mind. Social Psychol. Personality Sci. 6 (6), 646–652. https://doi.org/10.1177/1948550615578177.

Brühlmann, F., Schmid, G.-M., 2015. How to measure the game experience?: Analysis of the factor structure of two questionnaires. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, Crossings, Seoul, Korea. ACM, pp. 1181–1186.

Cairns, P., Cox, A., Berthouze, N., Jennett, C., Dhoparee, S., 2006. Quantifying the experience of immersion in games. In: Cognitive Science of Games and Gameplay Workshop at Cognitive Science 2006, pp. 7. Retrieved from. http://discovery.ucl.ac.uk/149051/.

Clark, L.A., Watson, D., 1995. Constructing validity: basic issues in objective scale development. Psychol. Assess. 7, 309–319.

Deci, E., Vallerand, R., Pelletier, L., Ryan, R., 1991. Motivation and education: the self-determination perspective. Educ. Psychol. 26 (3), 325–346. https://doi.org/10.1207/s15326985ep2603&4_6.

Denisova, A., Nordin, A.I., Cairns, P., 2016. The convergence of player experience questionnaires. In: CHI PLAY '16, pp. 33–37. https://doi.org/10.1145/2967934.2968095.

Depping, A.E., Mandryk, R.L., 2017. Why is this happening to me? How player attribution can broaden our understanding of player experience. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017), pp. 1040–1052. https://doi.org/10.1145/3025453.3025648.

Drachen, A., Canossa, A., 2009. Towards gameplay analysis via gameplay metrics. In: Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era. ACM, Tampere, Finland, pp. 202–209. https://doi.org/10.1145/1621841.1621878.

Ellick, W., Mirza-Babaei, P., Wood, S., Smith, D., Nacke, L.E., 2013. Assessing user preference of video game controller button settings. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems, pp. 1107–1112.

Engl, S., Nacke, L.E., 2013. Contextual influences on mobile player experience. Entertainment Comput. 4 (1), 83–91. https://doi.org/10.1016/j.entcom.2012.06.001.

Fisher, R.P., Geiselman, R.E., 1992. Memory Enhancing Techniques For Investigative Interviewing: The Cognitive Interview. Charles C Thomas Publisher.

Gajadhar, B., De Kort, Y., Ijsselsteijn, W., 2008a. Shared fun is doubled fun: player enjoyment as a function of social setting. Fun and Games. LNCS, vol. 5294. pp. 106–117. Eindhoven, Netherlands. Retrieved from. http://www.scopus.com/inward/record.url?eid=2-s2.0-57049150761&partnerID=40&md5=ba9fbe9705551a579fd6e93a7e775022.

Gajadhar, B., De Kort, Y., Ijsselsteijn, W., 2009. Rules of engagement: Influence of co-player presence on player involvement in digital games. Int. J. Gaming Comput.-Mediated Simul. 1 (3), 14–27. Retrieved from. http://www.scopus.com/inward/record.url?eid=2-s2.0-78249283925&partnerID=40&md5=418a53da87c73815a9f159b33d092f40.

Gajadhar, B., De Kort, Y., IJsselsteijn, W., 2008b. Influence of social setting on player experience of digital games. In: Proceeding of the Twenty Sixth Annual CHI Conference Extended Abstracts on Human Factors in Computing Systems CHI 08. ACM Press, pp. 3099–3104. https://doi.org/10.1145/1358628.1358814.

Gajadhar, B., Nap, H., De Kort, Y., Ijsselsteijn, W., 2010. Out of sight, out of mind: co-player effects on seniors' player experience. Fun and Games. pp. 74–83. Leuven. Retrieved from. http://www.scopus.com/inward/record.url?eid=2-s2.0-78249253331&partnerID=40&md5=22ecc468df0746a8b7013ab6ce3a958b.

Gerling, K., Birk, M., Mandryk, R., Doucette, A., 2013. The effects of graphical fidelity on player experience. In: Proceedings of International Conference on Making Sense of Converging Media, pp. 229–236.

Gerling, K., Schild, J., Masuch, M., 2011. Exergaming for elderly: analyzing player experience and performance. Mensch Comput. 2011, 11 *Fachübergreifende Konferenz Für Interaktive Und Kooperative Medien. überMEDIEN*-ÜBERmorgen, (Nintendo 2010), 401.

Henseler, J., Ringle, C.M., Sarstedt, M., 2015. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J. Acad. Market. Sci. 43, 115–135.

Hohlfeld, O., Fiedler, H., Pujol, E., Guse, D., 2016. Insensitivity to network delay: minecraft gaming experience of casual gamers. In: 2016 28th International Teletraffic Congress (ITC 28). IEEE, pp. 31–33. https://doi.org/10.1109/ITC-28.2016.313.

Hooper, D., Coughlan, J., Mullen, M.R., 2008. Structural equation modelling: Guidelines for determining model fit. Electronic Journal of Business Research Methods 6, 53–60.

Hu, L.T., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55.

IJsselsteijn, W., Kort, Y., Poels, K., Jurgelionis, A., Bellotti, F., 2007. Characterising and measuring user experiences in digital games. In: ACE 2007 International Conference on Advances in Computer Entertainment Technology, Salzberg, Austria. ACM, pp. 1–4.

IJsselsteijn, W., Poels, K., & de Kort, Y. (2008). Measuring player experiences in digital games. Development of the Game Experience Questionnaire (GEQ). *Manuscript in Preparation*.

IJsselsteijn, W., Van Den Hoogen, W., Klimmt, C., De Kort, Y., Lindley, C., Mathiak, K., ... Vorderer, P., 2008b. Measuring the experience of digital game enjoyment. In: Proceedings of Measuring Behavior. Maastricht, The Netherlands, Retrieved from. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.567.8017&rep=rep1&type=pdf.

Inchamnan, W., Wyeth, P., 2013. Motivation during videogame play: analysing player experience in terms of cognitive action. In: Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death. Melbourne, Australia, . https://doi.org/10.1145/2513002.2513010.

Johnson, D., Gardner, J., 2010. Personality, motivation and video games. In: OZCHI 2010 Proceedings. Brisbane, QLD, pp. 276–279. Retrieved from. http://www.scopus.com/inward/record.url?eid=2-s2.0-79953090423&partnerID=40&md5=716f3854528b9b295619ae24fba04fd2.

Johnson, D., Nacke, L.E., Wyeth, P., 2015a. All about that base: differing player experiences in video game genres and the unique case of MOBA games. In: Proceedings of SIGCHI 2015. Seoul, Korea, pp. 2265–2274. https://doi.org/10.1145/2702123.2702447.

Johnson, D., Watling, C., Gardner, J., Nacke, L.E., 2014. The edge of glory. In: Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '14. Toronto, Ontario, Canada, pp. 141–150. https://doi.org/10.1145/2658537.2658694.

Johnson, D., Wyeth, P., Clark, M., Watling, C., 2015b. Cooperative game play with avatars and agents: differences in brain activity and the experience of play. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, pp. 3721–3730. https://doi.org/10.1145/2702123.2702468.

Johnson, D., Wyeth, P., Sweetser, P., Gardner, J., 2012. Personality, genre and videogame play experience. In: Proceedings of the 4th International Conference on Fun and Games - FnG '12. ACM Press, New York, New York, USA, pp. 117–120. https://doi.org/10.1145/2367616.2367633.

Kao, D., 2016. Exploring the impact of avatar color on game experience in educational games. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 1896–1905. https://doi.org/10.1145/2851581.2892281.

Kao, D., Harrell, D.F., 2015. Exploring the impact of role model avatars on game

experience in educational games. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, pp. 571–576. https://doi.org/10.1145/2793107.2810291.
Kao, D., Harrell, D.F., 2016. Exploring the effects of encouragement in educational games. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16. ACM. https://doi.org/10.1145/2851581.2892335.
Kenny, D., 2015. Measuring Model Fit. http://davidakenny.net/cm/fit.htm.
Lankoski, P., Bjork, S. (Eds.), 2015. Game Research Methods. ETC Press, Pittsburgh, PA, USA.
Liukkonen, T.N., Heinonen, T., Raitoharju, R., Pitkäkangas, P., 2015. Motion tracking exergames for elderly users. IADIS Int. J. Comput. Sci. Inf. Syst. 10 (2), 52–64.
Loevinger, J., 1957. Objective tests as instruments of psychological theory. Psychol. Rep. 3, 635–694.
Madore, K.P., Addis, D.R., Schacter, D.L., 2015. Creativity and memory: Effects of an episodic-specificity induction on divergent thinking. Psychol. Sci. 26 (9), 1461–1468. https://doi.org/10.1177/0956797615591863.
Al Mahmud, A., Mubin, O., Shahid, S., Martens, J.-B., 2010. Designing social games for children and older adults: two related case studies. Entertainment Comput. 1 (3–4), 147–156. https://doi.org/10.1016/j.entcom.2010.09.001.
Mekler, E.D., Bopp, J.A., Tuch, A.N., Opwis, K., 2014. A systematic review of quantitative studies on the enjoyment of digital entertainment games. In: Proceedings of the 32nd Annual ACM Conference on Human factors in Computing Systems - CHI '14, pp. 927–936. https://doi.org/10.1145/2556288.2557078.
Memon, A., Meissner, C.A., Fraser, J., 2010. The cognitive interview: a meta-analytic review and study space analysis of the past 25 years. Psychol. Public Policy Law 16 (4), 340–372. https://doi.org/10.1037/a0020518.
Nacke, L.E., 2010. Wiimote vs. controller: electroencephalographic measurement of affective gameplay interaction. In: Proceedings of the International Academic Conference on the Future of Game Design and Technology. ACM, Vancouver, British Columbia, Canada, pp. 159–166. https://doi.org/10.1145/1920778.1920801.
Nacke, L.E., 2013. An Introduction to Physiological Player Metrics for Evaluating Games. Game Analytics. https://doi.org/10.1007/978-1-4471-4769-5_26.
Nacke, L.E., Drachen, A., 2011. Towards a framework of player experience research. In: Proceedings of the Second International Workshop on Evaluating Player Experience in Games at FDG 2011. Bordeaux, France.
Nacke, L.E., Grimshaw, M.N., Lindley, C.A., 2010a. More than a feeling: measurement of sonic user experience and psychophysiology in a first-person shooter game. Interact. Comput. 22 (5), 336–343. https://doi.org/10.1016/j.intcom.2010.04.005.
Nacke, L.E., Lindley, C.A., 2008. Boredom, Immersion, Flow - a pilot study investigating player experience. In: IADIS Gaming 2008: Design for Engaging Experience and Social Interaction. IADIS Press, Amsterdam, The Netherlands, pp. 103–107.
Nacke, L.E., Stellmach, S., Lindley, C.A., 2010b. Electroencephalographic assessment of player experience: a pilot study in affective ludology. Simul. Gaming 42 (5), 632–655. https://doi.org/10.1177/1046878110378140.
Oksanen, K., 2013. Subjective experience and sociability in a collaborative serious game. Simul. Gaming 44 (6), 767–793. https://doi.org/10.1177/1046878113513079.
Peng, W., Lin, J.-H., Pfeiffer, K.A., Winn, B., Peng, W., 2012. Need satisfaction supportive game features as motivational determinants: an experimental study of a self-determination theory guided exergame. Media Psychol. 15, 175–196. https://doi.org/10.1080/15213269.2012.673850.
Poels, K., de Kort, Y., Ijsselsteijn, W., 2007. It is always a lot of fun!" Exploring dimensions of digital game experience using focus group methodology. In: Proceedings of the 2007 conference on Future Play. ACM, Toronto, ON, pp. 83–89. https://doi.org/10.1145/1328202.1328218.
Przybylski, A.K., Ryan, R.M., Rigby, C.S., 2009. The motivating role of violence in video games. Personality Social Psychol. Bull. 35 (2), 243–259. https://doi.org/10.1177/0146167208327216.
Przybylski, A.K., Weinstein, N., Murayama, K., Lynch, M.F., Ryan, R.M., 2012. The ideal self at play: the appeal of video games that let you be all you can be. Psychol. Sci. 23 (1), 69–76. https://doi.org/10.1177/0956797611418676.
Rigdon, E., 1996. CFI versus RMSEA: a comparison of two fit indexes for structural equation modeling. Struct. Equ. Model. 3 (4), 369–379.
Robb, J., Garner, T., Collins, K., Nacke, L.E., 2017. The impact of health-related user interface sounds on player experience. Simul. Gaming 104687811668823. https://doi.org/10.1177/1046878116688236.
Rotoly De Lima, L.G., De, A., Salgado, L., Freire, A.P., 2015. Evaluation of the user experience and intrinsic motivation with educational and mainstream digital games. In: Proceedings of the Latin American Conference on Human Computer Interaction. Córdoba, Argentina, . https://doi.org/10.1145/2824893.2824904.
Ryan, R.M., E.L., Deci, 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55 (1), 68–78.
Ryan, R.M., Rigby, C.S., Przybylski, A., 2006. The motivational pull of video games: a self-determination theory approach. Motivation Emotion 30 (4), 347–363. https://doi.org/10.1007/s11031-006-9051-8.
Sifa, R., Drachen, A., Bauckhage, C., 2017. Profiling in games: understanding behavior from telemetry. In: Lakkaraju, K., Sukthankar, G. (Eds.), Social Interaction in Virtual Worlds. Cambridge University Press, pp. 1–44.
Simms, L.., Watson, D., 2007. The construct validation approach to personality scale construction. In: Robins, R.W., Fraley, R.C., Krueger, R.F. (Eds.), Handbook of Research Methods in Personality Psychology. Guilford, New York, pp. 240–258.
Sobel, M., Bohrnstedt, G.W., 1985. Use of null models in evaluating the fit of covariance structure models. Sociol. Methodl. 15, 152–178.
Souders, D., Boot, W., Charness, N., Moxley, J., 2016. Older adult video game preferences in practice: investigating the effects of competing or cooperating. Games Culture 11. https://doi.org/10.1177/1555412015603538.
Tamborini, R., Bowman, N.D., Eden, A., Grizzard, M., Organ, A., 2010. Defining media enjoyment as the satisfaction of intrinsic needs. J. Commun. 60 (4), 758–777. https://doi.org/10.1111/j.1460-2466.2010.01513.x.
Van Voorhis, C.W., Morgan, B.L., 2007. Understanding power and rules of thumb for determining sample sizes. Tut. Quant. Methods Psychol. 3 (2), 43–50.
Vella, K., Johnson, D., Hides, L., 2013. Positively playful: when videogames lead to player wellbeing. In: First International Conference on Gameful Design, Research and Applications.
Vella, K., Johnson, D., Hides, L., 2015a. Indicators of Wellbeing in Recreational Video Game Players. ACM, pp. 613–617. https://doi.org/10.1145/2838739.2838818.
Vella, K., Johnson, D., Hides, L., 2015b. Playing alone, playing with others. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15. London, United Kingdom, pp. 3–12. https://doi.org/10.1145/2793107.2793118.
Wylie, J.., Jolly, A., 2013. Understanding recruitment: outcomes associated with alternate methods for seed selection in respondent driven sampling. BMC Med. Res. Methodol. 13, 93.