Design and Preliminary Validation of The Player Experience Inventory

Vero Vanden Abeele
KU Leuven
Leuven, Belgium
vero.vandenabeele@kuleuven.be

Elisa D. Mekler
HCI Games Group,
University of Waterloo,
Waterloo, Canada
elisamekler@gmail.com

Lennart E. Nacke
HCI Games Group,
University of Waterloo,
Waterloo, Canada
lennart.nacke@acm.org

Daniel Johnson
Queensland University of Technology (QUT)
Brisbane, Australia
dm.johnson@qut.edu.au

CHI PLAY'16 Extended Abstracts, October 16-19, 2016, Austin, TX, USA

http://dx.doi.org/10.1145/2968120.2987744

Abstract
We present the design and preliminary results of the validation of the Player Experience Inventory (PXI). Based on the input of 64 experts in the field of player-computer interaction, we designed and refined this new scale. Our scale is based on the MDA framework (and on Means-End theory, underlying MDA). The PXI incorporates two subscales, one with dimensions at the functional level (i.e., dynamics) and one at the psychosocial level (i.e., aesthetics). The initial results, via principal factor analysis, suggest the scale can be used accurately to evaluate player experience. This work is our first step towards presenting a new, validated survey instrument for player experience evaluation.

Author Keywords
Player experience; scale; PXI; evaluation

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

Introduction
In this paper, we present the design and preliminary validation results of the Player Experience Inventory (PXI), a scale designed to measure player experience. Although scales for evaluating player experience concepts exist already, there is a need for a comprehensive scale that serves both designers and user researchers, that is validated within (and originates from)

the player experience and games user research communities, and that is freely available. Much current work in these research communities points to unsatisfactory ways of measuring player experience, e.g. [1, p. 449], [2]–[5].

In their systematic review on game enjoyment, Mekler et al. collected 87 quantitative studies [6] that measured player experiences. They found that the most used standardized questionnaires were the Intrinsic Motivation Inventory (IMI) [7], used in 15 research studies, and the Game Experience Questionnaire (GEQ) by IJsselstein et al. [8], used in 8 studies. However, they also found that most researchers who investigated determinants of game enjoyment utilized self-developed questionnaires rather than standardized questionnaires.

We also reviewed player experience research and can substantiate similar percentages. We used the following Boolean search string in the ACM digital library: "(Title:game*) and (PublicationTitle:SIGCHI Conference on Human Factors in Computing Systems)". This returned 106 results, including 21 papers that contained self-reports to measure player experiences. The PENS [9] and the IMI [7] were most common, each of them used three times. The GEQ [8] was mentioned twice. Self-developed questionnaires constituted the majority of papers (n=9), confirming the pattern found by Mekler et al. These studies suggest that, aside from author-developed scales, the three most commonly used scales are the IMI [7] the GEQ [8] and the PENS [9].

However, these scales have challenges in terms of their usefulness of the measured constructs for game designers and games user researchers. To date, no formal validation of the Game Experience Questionnaire has been published. Moreover, some authors report different factor structures emerging when using this scale [1], [2]. The PENS, though made available to some researchers, is a commercial scale. Therefore, it is not possible to publish and discuss specific items, with associated limitations in a research context. The IMI [7] is perhaps the most established scale as its scientific roots lie in motivational psychology. However, it is not operationalized for player experience, which is conceptualized to be broader than intrinsic motivation.

As for the self-developed scales, there are reasons for tailoring scales to idiosyncratic research objectives. Naturally, different research foci are served by different questionnaires. However, as a consequence, it is difficult to compare results over different studies, with different games or audiences, because of a lack of shared measurements of player experience. This may limit the field's ability to progress.

In sum, there is a lack of freely available, validated, player experience questionnaires. However, the constructs that should be part of such a player experience questionnaire are still being debated. To further this debate, we undertook card sorting studies and interviews with experts, active in the domain of player-computer interaction, asking them to discuss what constructs they envisioned as crucial to understanding player experience. Our aim in this study was to seek consensus regarding key constructs, with the ultimate aim of our larger program of research of developing and validating a freely-available player experience inventory. Relatedly, we seek to inform research being conducted by others in this space by identifying relevant constructs for consideration.

Method

The PXI has been under development for a period of two years and includes two major iterations preceding the current validation of the scale.

Iteration 1

In the early spring 2015, a review was conducted on scales used in current PCI research; 124 scales were discussed, comprising over 700 submissions (for a full list, see https://goo.gl/jxPttB). On the basis of discussions among 3 authors, 9 constructs (Enjoyment, Competence, Autonomy, Ease-of-Control, Cognitive Immersion, Meaning, Effort, Aesthetic appeal) were put forward as possible candidates for such a scale, and 5-7 items per construct were devised.

Next, these 9 constructs and their items were presented to 31 experts in the PCI field as a card sorting exercise. Experts were asked to group items, and to add constructs they felt were missing. For 24 of 31 experts, the card sorting was followed by an interview, discussing how they defined the player experience, what theories they found informative and what player experience dimensions they would include in the scale in addition to discussing the chosen constructs and how to operationalize them into items.

Overall, our participants critiqued the usefulness of some constructs (e.g., Effort) and the labels (e.g., Cognitive absorption, Aesthetic appeal).

They also mentioned many theories and conceptual frameworks during the interview, in line with their varied backgrounds. On one end of the spectrum, theories on human motivation were given, such as Self-Determination Theory [10] or Uses & Gratifications Theory [11]. The researchers who proposed these theories tended to see player experience as defined by motives that are enduring, in some cases so enduring that they might be considered personality or game types [12].

On the other hand, practical game design frameworks were also mentioned such as Octalysis [13], the Book of lenses [14] or the MDA framework [15]. These were typically mentioned by researchers drawn from industry or with a background in design or arts. These frameworks often link game dynamics to player experiences rather than listing psychological constructs. These experts identified missing constructs, situated at the immediate experiential level, for example, constructs that measure challenge, progress feedback, goal perception. They also stressed the importance of audiovisual appreciation, rarely a part of current scales.

It was emphasized by many experts (both academic and industry) that concepts at one level can be causal to the higher psychological constructs. This distinction and causal relationship between the immediate experience during gameplay (i.e., game dynamics) on the one hand, and psychological motives (i.e., aesthetics) on the other hand is reminiscent of MDA, a widely acknowledged game design framework (see Figure 1).

In sum, from this qualitative study, a tension between 'academic' game researchers and 'designer' game researchers was revealed. This duality could be a reason why it is difficult to develop a consensus on how to measure player experience empirically.

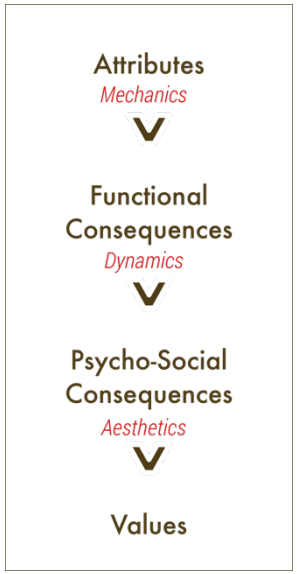


Figure 1. The means-end chain ranging as specified in [14], and mapped unto the MDA framework [13].

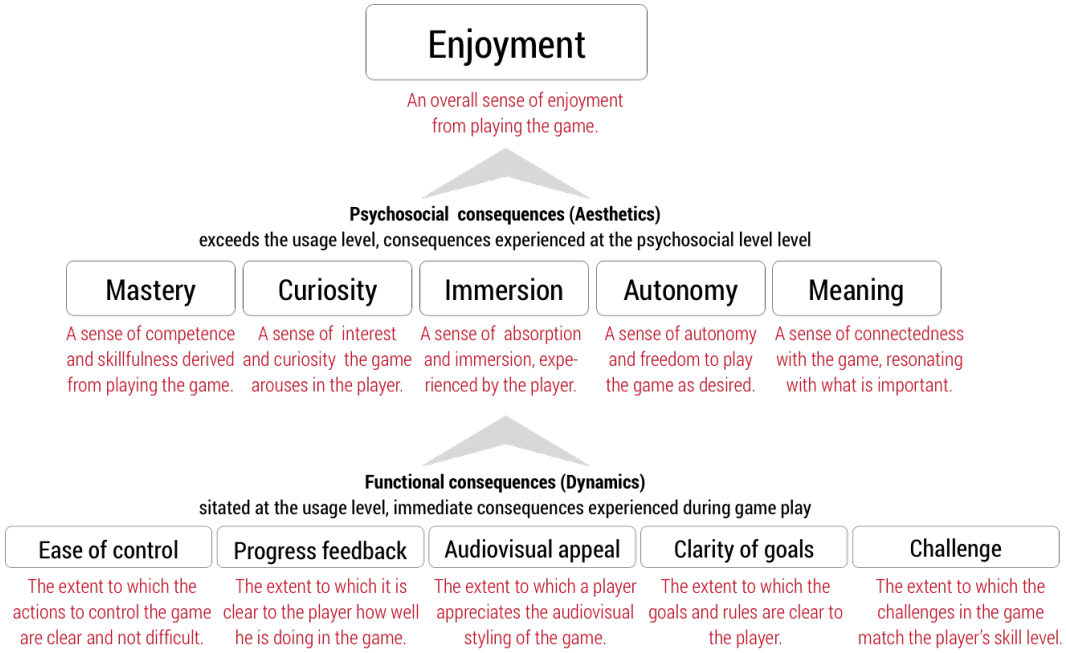


Figure 2. The revised Player Experience Inventory, based on the critiques of 31 experts, and based on the MDA framework[15] and Means-End theory [16] .

Iteration 2.

Armed with the critiques of experts, we revised our player experience inventory: we removed 1 construct (Effort), we added three new constructs (Progress feedback, Clarity of goals, Challenge) and reworded the labels for several constructs. Moreover, we revised our theoretical modal, drawing heavily on the MDA framework and Means-End (ME) theory, which is the scientific model underlying the MDA framework. ME theory posits that attributes are only a means to an end [16]. Applied to games, game attributes are a means to experience a player holds during and after game play. ME theory distinguishes between *functional consequences* and *psychosocial consequences (see figure 1)*. Functional consequences are situated at the usage level, these are the immediate and tangible consequences of game attributes that are experienced directly. Functional consequences align with the concept of dynamics in the MDA framework where "dynamics describes the run-time behavior of the mechanics acting on player inputs and each other's outputs over time." Psycho-social consequences exceed the immediate usage level and list the consequences either at the social or psychological level [17]. Hence, they are situated at the level of aesthetics as specified by the MDA framework. A player experience inventory that is based on the MDA framework (hence ME theory) may be particularly useful in the PCI community. A measurement instrument, incorporating two subscales, with dimensions at both the level of functional (dynamics) and at the level of psychosocial consequences (aesthetics), may be useful for both theoretically-oriented researchers and design-oriented researchers.

In sum, based on the discussion with experts during iteration 1, a set of 11 constructs was devised (see Figure 2). Moreover, these constructs were arranged according to their respective levels, that is, at the functional or psychosocial level. Five to six items were again devised per construct (see Figure 3). The full list can be found online at https://goo.gl/EYuoJf.

These items were again put forward to 33 new game experts who did not take part in the previous study, this time via an *open* card sorting exercise. Hence, experts were not given the labels of constructs, nor the number or constructs we aimed at, but rather the 57 items only. They were allowed to provide labels for their groups if they wanted to, but this was not required. The clustering provided by experts clearly aligned with our 11 proposed constructs. The clustering of items (see Figure 3) by our experts shows the average pair agreements (i.e., how often one item was grouped with another item). For the 11 constructs this ranged between 95.5% and 66.3%, whereas the average pair agreement between items of different constructs was 6.1%.

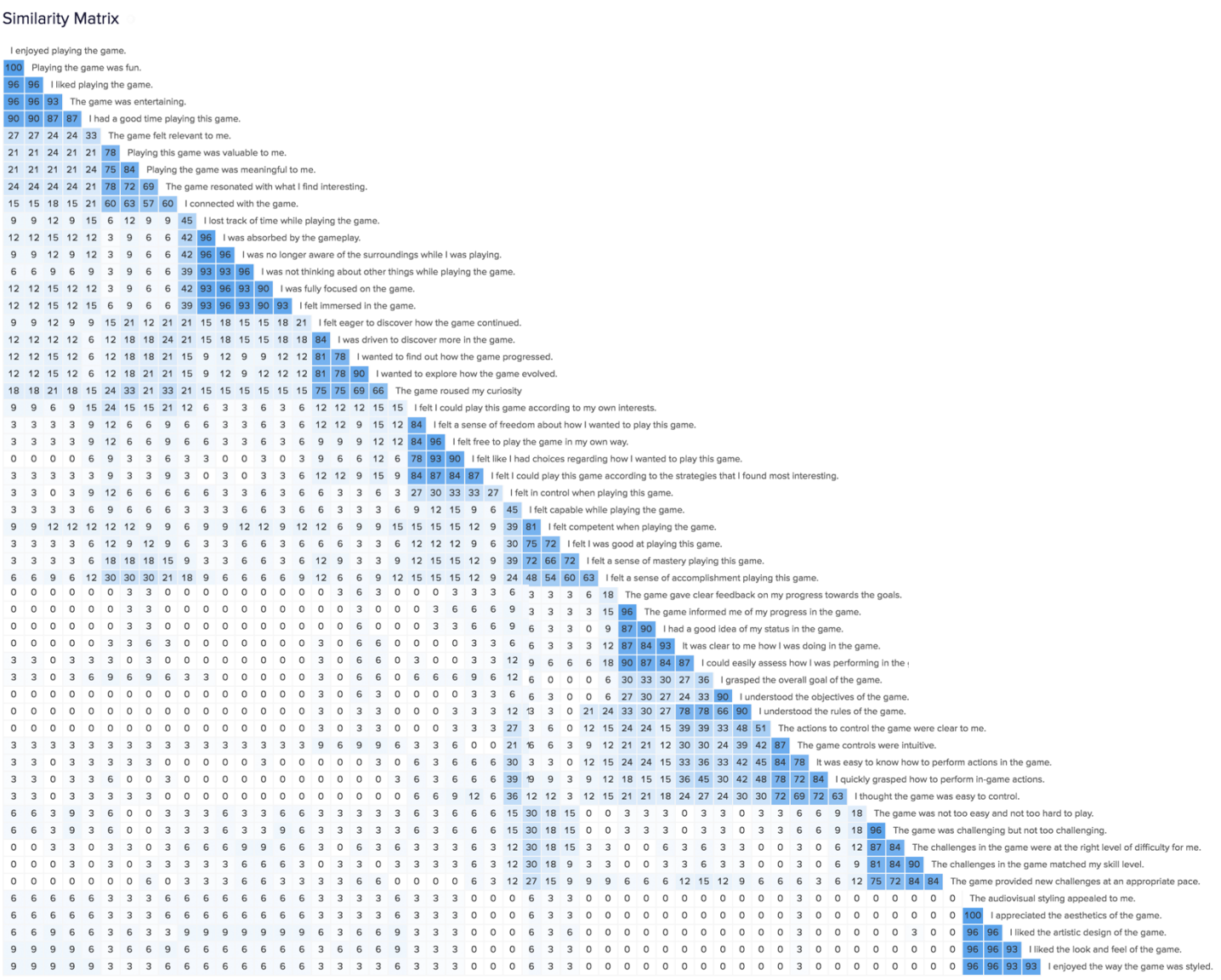
Figure 3. Similarity matrix of the second iteration of the design of constructs for the PXI.

This matrix shows the clustering of 57 items, grouped into to 11 constructs by 33 game experts, via an *open* card sorting exercise; no labels were given, nor the number of constructs desired, nor the number of items per construct required. The clustering results clearly align with our 11 proposed constructs

Average pair agreements (in%)for constructs are:

Enjoyment:93.1
Meaning: 69.6
Immersion:93.8
Curiosity: 77.7
Autonomy: 67.8
Mastery:66.3
Progress feedback: 88.5
Clarity of goals:81.0
Ease of control:75.9
Challenge:83.7
Audiovisual appeal:95.5

Average pair agreement between items of different constructs is 6.1%.



Validation

We are now in the last phase of our scale development, performing exploratory factor analysis for both the subscale of functional consequences (dynamics) and the subscale of psycho-social consequences (aesthetics), to inspect factors and factor loadings. Thus far, we have collected 144 responses[1]. For both subscales, principal axis factoring (PFA) with direct oblimin rotation, on the 25 items at the functional level (i.e., dynamics) and the 27 items at the psychosocial level shows a five factor solution (eigenvalues > 1, total variance explained 71.55% and 71.79% respectively). The pattern matrices (see table 1 and table 2) below show the factor loadings for each factor

Table 1 . Pattern matrix of the funct. consequences subscale: Principal Axis Factoring, Oblimin with Kaiser Normalization.

Factors	1	2	3	4	5
Ease of Control 1					
Ease of Control 2				.406	
Ease of Control 3	.472				
Ease of Control 4				.728	
Ease of Control 5				.654	
Challenge 1			.577		
Challenge 2			.739		
Challenge 3			.712		
Challenge 4			.438		
Challenge 5			.771		
Progress feedback 1					
Progress feedback 2					.444
Progress feedback 3					.711
Progress feedback 4					.626
Progress feedback 5					.918
Audiovisual appeal 1		.836			
Audiovisual appeal 2		.929			
Audiovisual appeal 3		.949			
Audiovisual appeal 4		.874			
Audiovisual appeal 5		.851			
Clarity of goals 1	.757				
Clarity of goals 2	.835				
Clarity of goals 3	.537				
Clarity of goals 4	.871				
Clarity of goals 5	,723				

[1] This is below the heuristic of 300 respondents (or 5-10 cases per item). Yet, the KMO (.905 and .911 respectively) and Barlett test of Sphericity (both $p < .001$) suggest that the dataset is suitable for factor analysis. Yet, more responses will be added in the future, to increase reliability.

Table 2. Pattern matrix of the Psychosocial subscale: Principal Axis Factoring, Oblimin with Kaiser Normalization.

Factors	1	2	3	4	5
Meaning 1	.379				.310
Meaning 2	.661				
Meaning 3	.746				
Meaning 4	.860				
Meaning 5	.744				
Mastery 1					.764
Mastery 2					.780
Mastery 3					.709
Mastery 4					.507
Mastery 5					.372
Mastery 6	.306				.383
Immersion 1		.394	.362		
Immersion 2		.663			
Immersion 3		.556			
Immersion 4		.827			
Immersion 5		.616			
Immersion 6	.347	.535			
Autonomy 1	.371				
Autonomy 2				.645	
Autonomy 3				.790	
Autonomy 4				.835	
Autonomy 5				.601	
Curiosity 1			.830		
Curiosity 2			.886		
Curiosity 3			.696		
Curiosity 4			.875		
Curiosity 5			.500		

Conclusion and future work

Further research is necessary: collecting data from more respondents, and investigating which items can be removed, ensuring both comprehensiveness *and* efficiency of the scale. We are careful regarding multi-collinearity, as we see high loading variables. These items should be removed. As a final step, confirmatory factor analysis will be conducted to verify the goodness-of-fit of the model underlying the player experience inventory. It is our hope if the model fits, such a scale may be useful for both academics and designers.

References

[1] D. Johnson, P. Wyeth, P. Sweetser, and J. Gardner, "Personality, genre and videogame play experience," in *Proceedings of the 4th International Conference on Fun and Games*, 2012, pp. 117–120.

[2] D. Johnson, C. Watling, J. Gardner, and L. E. Nacke, "The edge of glory: the relationship between metacritic scores and player experience," in *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, 2014, pp. 141–150.

[3] K. Verhoeven, V. V. Abeele, B. Gers, and J. Seghers, "Energy Expenditure During Xbox Kinect Play in Early Adolescents: The Relationship with Player Mode and Game Enjoyment," *Games Health J.*, vol. 4, no. 6, pp. 444–451, Jul. 2015.

[4] F. Brühlmann and G.-M. Schmid, "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*, New York, NY, USA, 2015, pp. 1181–1186.

[5] K. L. Norman, "GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers," *Interact. Comput.*, vol. 25, no. 4, pp. 278–283, Jul. 2013.

[6] E. D. Mekler, J. A. Bopp, A. N. Tuch, and K. Opwis, "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*, New York, NY, USA, 2014, pp. 927–936.

[7] E. L. Deci and R. M. Ryan, *Intrinsic motivation and self-determination in human behavior*. Springer, 1985.

[8] W. IJsselsteijn, Y. A. W. de Kort, K. Poels, A. Jurgelionis, and V. Bellotti, "Characterising and Measuring User Experiences in Digital Games," presented at the ACE conference '07, Salzburg, Austria, 2007.

[9] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motiv. Emot.*, vol. 30, no. 4, pp. 344–360, Nov. 2006.

[10] E. L. Deci and R. M. Ryan, *Intrinsic motivation and self-determination in human behavior*. Springer, 1985.

[11] J. L. Sherry, K. Lucas, B. S. Greenberg, and K. Lachlan, "Video game uses and gratifications as predictors of use and game preference," *Play. Video Games Motiv. Responses Consequences*, pp. 213–224, 2006.

[12] L. E. Nacke, C. Bateman, and R. L. Mandryk, "BrainHex: A neurobiological gamer typology survey," *Entertain. Comput.*, vol. 5, no. 1, pp. 55–62, Jan. 2014.

[13] "Octalysis: Complete Gamification Framework - Yukai Chou," *Yu-kai Chou & Gamification*. .

[14] J. Schell, *The Art of Game Design: A book of lenses*, 1st ed. Morgan Kaufmann, 2008.

[15] R. Hunicke, M. LeBlanc, and R. Zubek, "MDA: A Formal Approach to Game Design and Game Research," presented at the Game Developers Conference, 2001.

[16] T. J. Reynolds and J. C. Olson, *Understanding Consumer Decision Making: The Means-end Approach To Marketing and Advertising Strategy*. Psychology Press, 2001.

[17] T. J. Reynolds and J. Gutman, "Laddering theory, method, analysis, and interpretation," vol. Journal of Advertising Research 28 (1), pp. 11–31, 1988.