How to Measure the Game Experience? Analysis of the Factor Structure of Two Questionnaires

Florian Brühlmann
University of Basel,
Department of Psychology
Missionsstrasse 62a
4055 Basel, Switzerland
florian.bruehlmann@gmail.com

Gian-Marco Schmid
University of Basel,
Department of Psychology
Missionsstrasse 62a
4055 Basel, Switzerland
gian-marco.schmid@unibas.ch

CHI'15 Extended Abstracts, April 18–23, 2015, Seoul, Republic of Korea.
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2732831

Abstract
We describe and report the analysis of two widely used questionnaires to measure the player experience in digital games. In order to contribute to the further validation and meaningful application of the PENS and GEQ we examined the underlying factorial structure of both questionnaires. Four hundred and forty-seven participants played two different games and rated them on a set of various variables including the PENS and GEQ. Consistent with previous research we gained additional insight into optimization of both measurements. While the factor structure of the PENS appears to be consistent and invariant across two different games, the GEQ reveals weaknesses in fulfilling these requirements.

Author Keywords
Digital Games; Game Experience; Factor Analysis; Questionnaire

ACM Classification Keywords
K.8.0 [Personal Computing]: Games; J.4 [Social and Behavioral Sciences]: Psychology

Introduction
Within the last decades gaming has become the world's largest entertainment medium [10, 11]. At the same time, a growing interest in assessing the games user experience

Figure 1: Super Hot is a FPS game where the player controls time by moving the avatar. The prototype of the game (`http://superhotgame.com/play-prototype`) was used in this study.

Figure 2: Canabalt (`http://www.adamatomic.com/canabalt/`) is a simple jump and run game where the player has to prevent the avatar from falling down or colliding with objects.

can be observed in both academia and industry. However, since this represents a rather new research domain there is only a limited amount of questionnaire-like measurements available. This is probably also evoked due to commercial interests which appears to be a common issue in related domains of user experience (UX) [1]. Within games user research two questionnaires, the Player Experience of Need Satisfaction questionnaire (PENS) and the Game Experience Questionnaire (GEQ) are considered to be widely used in previous research [5]. In a recent study, Johnson et al. (2014) [5] investigated these measurements in relation to the Metacritic scores (www.metacritic.com). They found that in comparison to the uni-dimensional ratings of the Metacritic scores, multidimensional measurements such as PENS and GEQ deliver a more nuanced understanding of the subjective quality of digital games (e.g. people who base their decision to buy a game on Metacritic scores risk to miss games they would actually like). The flaws of uni-dimensional constructs dealing with rather complex constructs is also discussed in other UX-domains such as aesthetics (e.g. [7]). Johnson et al. noted that the psychometric properties of the GEQ are difficult to evaluate due to the lack of studies dealing with the validation of the factor structure. This is confirmed by Norman [8] who stated that publications by Ijsselsteijn et al. dealing with the GEQ did not involve any empirical results of the questionnaire concerning its reliability and validity. Despite their widespread application in research, measurements of player experience have hardly been subject of replication or further validation [5]. With the intention of analyzing and comparing the underlying structure of the PENS and GEQ, an online study with two different games (See Figure 1 and Figure 2) was conducted. Therefore our research question is whether the proposed factor structure of the two questionnaires can be found in our data. In addition to that, we are interested in first indications of measurement invariance of the questionnaires across different groups. Without the satisfaction of these criteria the application of the measurements and the interpretation of subscale scores is questionable.

Player Experience of Need Satisfaction (PENS)
The proprietary PENS-questionnaire, developed by Ryan and colleagues investigates the "motivational pull" of video games by measuring the intrinsic motivation of players [10]. This approach is based on self-determination theory (SDT) and focuses on the three basic human needs for either intrinsic or extrinsic motivation: *Autonomy* (volitional aspects of an activity), *Competence* (perception of challenge) and *Relatedness* (connection to others) [9]. The applied focus on intrinsic motivation is explained through evidence of its vast impact on sports [3]. Two additional factors - *Presence* (the sense that one is *within* the game world) and *Intuitive controls* - were included in the PENS measurement because they are seen as important aspects of gaming [10]. The PENS-questionnaire measures these needs and the additional factors on 7-point Likert-scales. Because of the nature of the games used in this study, the factor Relatedness was excluded from the questionnaire, which resulted in a final set of 18 items.

Game Experience Questionnaire (GEQ)
The GEQ by Ijsselsteijn and colleagues [4] incorporates seven different dimensions of player experience: *Sensory and Imaginative Immersion, Tension, Competence, Flow, Negative Affect, Positive Affect and Challenge*. The GEQ is a self-report measure for a rather broad investigation of game experience. In comparison to the PENS-questionnaire, which was developed in a top-down theory approach, the GEQ evolved from a bottom-up

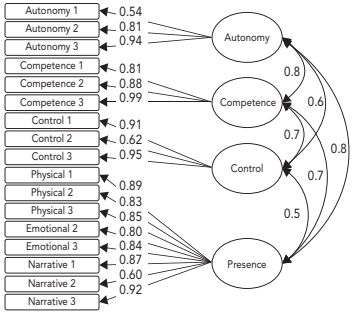

Figure 3: Standardized loadings greater .40 for items of the PENS questionnaire in game 1 and correlations between the extracted factors. *Note:* the item Emotional Presence 1 was excluded from the analysis.

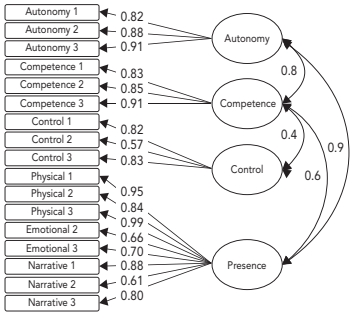

Figure 4: Standardized loadings greater .40 for items of the PENS questionnaire in game 2 and correlations between the extracted factors.

procedure by using focus group research and surveys with frequent players. In addition, the GEQ is measured on 5-point Likert-like scales ranging from "not at all" to "extremely". Each item asks respondents to indicate how they felt while playing the game. The items of the GEQ are freely available in Dutch, English, German and Finnish [4]. However, a detailed report of the development and psychometric properties has never been published [8]. The final set of the GEQ for this study included 36 items (6 spare items were not considered).

Method

Four hundred and forty-seven participants (233 female; 212 male; 3 without gender declaration) took part in an online study on a crowdsourcing platform. The experimental procedure was conducted in a within-subject design. Subjects were randomly assigned to play either game 1 (Super Hot, an experimental first-person shooter) first and then game 2 (Canabalt, a minimalistic jump 'n' run game) or vice versa. After playing each game for 3 minutes, participants were asked to answer the PENS and the GEQ questionnaire. At the end some demographic questions were provided. The participants received a small monetary compensation.

Results

Data Screening

Out of the initial 471 responses, 24 participants reported difficulties with running the games and were excluded from the analysis. A final sample size of 447 remained. Responses to 36 items of the GEQ questionnaire and 18 items of the PENS questionnaire were collected. The items for relatedness of the PENS questionnaire were omitted because neither of the two games included non-player characters or a multiplayer mode. The minimum amount of data for factor analysis was satisfied with over 12 cases per variable.

PENS

Item Analysis

Initially, the factorability of the 18 items was examined. Seventeen items correlated at least .3 with at least one other item for game 1 and game 2, suggesting reasonable factorability. The item Emotional Presence 1 was only marginally correlated with the other two Emotional Presence items and did not correlate with any other item stronger than .3, thus we excluded this item from the further analysis. Secondly, none of the 17 items were correlated higher than .9. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .96 for both games, above the recommended value of .6, and Bartlett's test of sphericity was highly significant (game 1: $\chi^2(136) = 8572$, $p < .001$ and game 2: $\chi^2(136) = 7897$, $p < .001$). Finally, the communalities were all above .3 for both games, which indicates that each item shared enough common variance with other items. Given these overall indicators, the factor analysis was conducted with remaining 17 items because the first item for Emotional Presence did not satisfy all criteria.

Factor Analysis

Principal axis factor analysis was used because the primary purpose was to find the proposed 4-factorial structure, as stated by the authors of the questionnaire [10]), and to confirm its invariance over the use of different game types. It was calculated with oblimin rotation because it is reasonable to expect that some factors (e.g., competence and control) are not independent from each other. The eigenvalues for game 1 showed that the first factor (Presence; 8 items) accounted for 44% of the explained variance, the second factor (Competence) 20%, the third factor (Control) 18% and

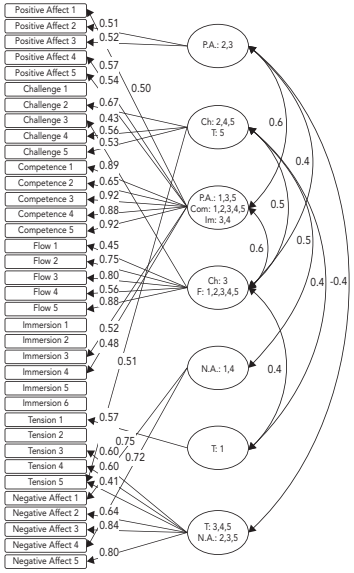

Figure 5: Standardized loadings greater .40 for items of the Game Experience Questionnaire in game 1 and correlations between the extracted factors.

the fourth factor (Autonomy) accounted for 18% of the explained variance. For game 2, results showed that the first factor (Presence) accounted for 45% of the explained variance, the second factor (Autonomy) 21%, the third factor (Competence) 20% and the fourth factor (Control) accounted for 13% of the explained variance. Figure 3 depicts standardized loadings of items on the extracted factors for game 1 and Figure 4 for game 2. The four factor solution explained a total of 80% of variance in game 1 and 77% in game 2. There were no substantial cross-loadings (above .40) observed and every item loaded on one of the extracted factors. The analysis of the factor solutions for both games revealed a congruence coefficient of .98 for Presence, .98 for Competence, .99 for Control, and .95 for Autonomy. Lorenzo-Seva & ten Berge (2006) [6] suggested that a Tucker's factor congruence coefficient of .85–.94 corresponds to fair similarity, while a coefficient equal to or higher than .95 implies that the two factors can be considered as equal.

GEQ

Item Analysis

Initially, the factorability of the 36 items was examined. All items correlated satisfactorily (above .3) with at least one other item and did not show any signs of redundancy (no correlations stronger than .9). The KMO measure was .96 for both games, above the recommended value of .6, and Bartlett's test of sphericity was highly significant (game 1: $\chi^2(630) = 14150, p < .001$ and game 2: $\chi^2(630) = 13634, p < .001$). Finally, each item shared enough common variance with other items for both games (all communalities above .3).

Factor Analysis

Again, principal axis factor analysis with oblimin rotation was used in order to investigate the structure of factors that are not completely independent from each other. The initial eigenvalues for game 1 showed that a first factor accounted for 32% of the explained variance, a second factor 21%, a third factor 15%, a fourth factor 9%, a fifth factor 8%, a sixth factor 10%, and a seventh factor accounted for 5% of the explained variance. For game 1, the items Challenge 1, Immersion 1, 2, 5 and 6, as well as the item Tension 2 did not load substantially (above .40) on any factor. The item Tension 5 loaded on factor 3 (.42; Negative Affect and Tension items) and factor 4 (.51; Challenge and Tension items). Figure 5 depicts loadings of the items in game 1 on the seven factors.

For game 2, results showed that a first factor accounted for 17% of the explained variance, a second factor 20%, a third factor 16%, a fourth factor 16%, a fifth factor 15%, a sixth factor 9%, and a seventh factor accounted for 7% of the explained variance. The seven factor solution explains a total of 68% of variance in game 1 and 67% in game 2. With the factor solution of game 2, the items Positive Affect 5, Challenge 1, Competence 2, Flow 1, Immersion 1, 2, 5 and 6 did not load substantially (above .40) on any factor. The item Challenge 4 loaded on factor 2 (.44; Flow and Challenge items) and factor 6 (.40; Tension and Challenge items). See Figure 6 for loadings of the items in game 2 on the seven factors.

The analysis of the factor solutions revealed congruence values ranging from .56 to .94, with five values greater than .85 (Negative Affect/Tension items with .94, Flow/Challenge with .94, Competence/Positive affect/Immersion with .89, Positive Affect with .91, a second Negative Affect/Tension factor with .86) and two values below .85 (Immersion .56; Challenge .84). A Tucker's congruence value between .85–.94 corresponds to fair similarity, factors with a coefficient equal to or higher

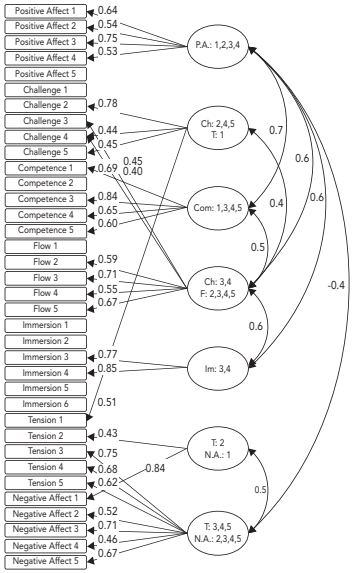

Figure 6: Standardized loadings greater $.40$ for items of the Game Experience Questionnaire in game 2 and correlations between the extracted factors.

than .95 can be considered equal [6].

Consequently, the factorial solutions can not be considered as equal for both games. The two factors Negative Affect/Tension and Flow/Challenge came close to the .95 threshold. The items Challenge 4 and Tension 5 were identified as problematic because they had cross-loading of .4 or above, although none of the items had similar cross-loading in both games. For both games the factor analysis extracted 2 factors with less than three loading items. These factors are probably unstable and indicate that a simpler factor solution might be more appropriate.

Discussion

In general, support for the assumed factor structure of the PENS questionnaire for Presence, Control, Autonomy and Competence related items was invariantly found across the two different games. Invariance is important to ensure that the applied questionnaire does measure the same constructs for different groups (e.g. games) in order to make a reasonable comparison between groups possible. However, further analysis of the data by applying structural equation modeling is needed to be on firm ground about measurement invariance. In the PENS questionnaire, the first item for Emotional Presence ("I am not impacted emotionally by events in the game") did not satisfy all quality criteria. However, this item was the sole negatively-keyed item in the questionnaire and there is a chance that some of the respondents failed to attend to the negative-positive wording. Negatively-keyed items are often included to reduce acquiescence bias but the advantages are controversial [2] and the questionnaire might benefit from removing or rephrasing this item.

The theoretically proposed structure of the GEQ could not be verified. Many items related to different aspects of the player experience loaded on the same factor. Moreover, by applying the proposed factor structure, the observed associations appear to vary across the two games. Only Negative Affect as one factor together with Tension items 3 to 5 and Flow together with items of Challenge can be considered almost invariant across both games. In comparison to the PENS the results suggest that the GEQ struggles to achieve the proposed factor structure and measurement invariance. This might be a consequence of the myriad of ambiguous concepts (e.g., Immersion and Tension) the GEQ tries to cover which results in items that are not clearly associated with one or another factor.

Although this study reports evidence that the GEQ has a different and less stable structure than expected, this does not mean it can not be a valid predictor of player enjoyment and commercial success of a game. It is however questionable whether the proposed subscales (e.g. Tension) measure what they target to measure. This is a possible limitation of the questionnaire which makes it difficult to interpret scores on subscales and guide design decisions in the development of a digital game.

There are at least three limitations of the findings reported in this study: First, participants of this study were not able to play the games for longer than 3 minutes. It is conceivable that some aspects of the player experience can only be adequately measured after a longer session. Second, the games selected were both relatively simple compared to full-fledged commercial games and might not offer a complex experience. Nevertheless, a valid multidimensional game experience questionnaire should ideally also be applicable to comparatively simple web-based games and rather short play sessions. Third, we can not make any statements about the relatedness subscale of the PENS questionnaire. We decided to omit

these items because they would make little sense in a relatively simple single player game. This article provides first indications for the validity of the factor structure and its stability for the PENS questionnaire as well as a comparison with the GEQ which did not reveal the proposed factor structure and appears to be less stable.

In a next step, it is planned 1) to conduct an item-level analysis in order to identify items of the GEQ which should be rephrased or excluded, and 2) run confirmatory factor analysis to examine the fit of this model compared to a simpler factorial structure. In case the found structure can be confirmed, correlations of the PENS and GEQ subscales with outcome variables such as enjoyment, intention to continue playing, and behavioral measures should be examined to study the concurrent and predictive validity of the questionnaires.

Acknowledgments

The authors would like to thank Prof. Dr. Klaus Opwis, Dr. Alexandre N. Tuch, Elisa D. Mekler, M.Sc., Dr. Thomas Ledermann, and Sebastian Hasler for their help and advice during the preparation of this work.

References

[1] Bargas-Avila, J. A., and Hornbæk, K. Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 2689–2698.

[2] DeVellis, R. F. *Scale development: Theory and applications*, vol. 26. Sage Publications, 2011.

[3] Frederick, C. M., and Ryan, R. M. Self-determination in sport: A review using cognitive evaluation theory. *International Journal of Sport Psychology 26*, 1 (1995), 5–23.

[4] IJsselsteijn, W., van den Hoogen, W., Klimmt, C., de Kort, Y., 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), 88–89.

[5] 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.

[6] Lorenzo-Seva, U., and Ten Berge, J. M. Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 2*, 2 (2006), 57–64.

[7] Moshagen, M., and Thielsch, M. T. Facets of visual aesthetics. *International Journal of Human-Computer Studies 68*, 10 (2010), 689–709.

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

[9] Ryan, R., and Deci, E. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology 25*, 1 (2000), 54–67.

[10] Ryan, R., Rigby, C., and Przybylski, A. The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion 30*, 4 (2006), 344–360.

[11] Yi, M. They got game: Stacks of new releases for hungry video enthusiasts mean its boom time for an industry now even bigger than hollywood. *San Francisco Chronicle* (2004), A1.