GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers

1 author:

Kent L. Norman
University of Maryland, College Park
116 PUBLICATIONS   **4,524** CITATIONS

GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers

KENT L. NORMAN

*Department of Psychology and the Human-Computer Interaction Laboratory, University of Maryland,
College Park, MD 20742, USA*
*Corresponding author: klnorman@umd.edu

In recent years, research on the psychological aspects and assessment of video games has become more and more important due to their impact in entertainment and education. The development of psychometric instruments to measure different factors of a player's engagement and skills in playing video games and to measure different factors of the game's playability and attractiveness is essential to this research. This article reviews two papers on measuring the player's subjective experiences playing the game by IJsselsteijn *et al.* (2007, Characterising and Measuring User Experiences in Digital Games. ACE Conference '07, June 13–15, Salzburg, Austria) and the player's level of engagement in playing games by Brockmyer *et al.* (2009, The Development of the Game Engagement Questionnaire: A Measure of Engagement in Video Game-playing. J. Exp. Soc. Psychol., 45, 624–634). While approaching the topic with very different purposes and methodologies, both papers contribute important ideas and useful scales that should be used by researchers in the field.

1. INTRODUCTION

Over the past two decades, video games have become a significant form of entertainment and useful tools for education and training. The two articles that I reviewed emphasize the importance of developing a questionnaire that taps into the construct of engagement in the game and the experience of the gamer. 'Characterising and Measuring User Experiences in Digital Games' by IJsselsteijn *et al.* (2007) acknowledge that it is not an easy task to adequately describe and measure the gaming experience. Yet it is extremely important to do so for the purpose of testing and evaluating games. They briefly talk about a Game Experience Questionnaire (GEQ), but give no information about its content or Development at this point. The second paper, 'The Development of the Game Engagement Questionnaire: A Measure of Engagement in Video Game-Playing' by Brockmyer *et al.* (2009), introduces the Game Engagement Questionnaire (also GEQ) along with a detailed account of its development and psychometric evidence of its reliability and validity. However, in the case of this questionnaire, its purpose is clearly aligned with identifying children who may be at risk of video game violence. Thus, the purposes of these two scales with the same acronyms are entirely different. The first, which will be referred to here as

the GexpQ, was designed for play testing games and assessing different aspects of the subjective experience of the playing the game. The second, which will be referred to here as the GengQ, was designed to assess an individual's tendency to be drawn into a game in an engaging, immersive way and ultimately to identify players that are particularly susceptible to video game violence.

2. THE GAME EXPERIENCE QUESTIONNAIRE

Wijnand IJsselsteijn and his colleagues tend to come from a human factors and usability perspective. However, when it comes to studying video games and the game experience, they rightly note that we must go beyond the now traditional concepts of user-centered design and user testing. These approaches have emphasized productivity in terms of such things as time to perform tasks, reduction of errors, the learning curve, quality of output, ease of use and user satisfaction. As Pagulayan *et al.* (2003) point out, what makes a game good is quite different from what makes a good word processor or spreadsheet. Work productivity is replaced by an interesting challenge in a game. For games, the reward is intrinsic rather than based on some material outcome. In work applications, consistency

and predictability are essential whereas in games variety and surprise are important.

On the other hand, as IJsselsteijn *et al.* note, some aspects of game software and hardware harken back to traditional usability. There are a number of out-of-game functions that are required to play games, which are often ignored in play testing. These would include the player logging in, setting user preferences, the configuration and calibration of interface devices, storing and retrieving game data, and managing resources. These functions provide necessary support for the game but are not in and of themselves fun.

Nevertheless, deriving an assessment of player experience involves a number psychological subjective feelings and emotions on the part of the game player. One is trying to assess a broad spectrum of experience that leads an individual to want to play a game, continue to play the game, to repeatedly come back to again play the game and to recommend the game to others.

A common approach to the problem of assessing a number of ill-defined or implicit factors is to start with a relatively large number of survey items and use exploratory factor analysis to reduce the number of items to a reasonable number of factors that reflect meaningful components of player experience.

At the end of their article, IJsselsteijn *et al.* report that they are in the process of developing the Game Experience Questionnaire to cover a range of digital game experiences that have been identified through reviewing theoretical accounts of player experiences as well as focus group explorations with both hardcore and casual gamers.

A copy of the questionnaire made available by the authors reveals that there are 33 items that are scored to obtain measures on seven different components labeled: competence (e.g. 'I felt skillful', 'I felt competent'), sensory and imaginative immersion (e.g. 'I was interested in the game's story', 'I felt imaginative'), flow (e.g. 'I was fully occupied with the game', 'I lost track of time'), tension/annoyance (e.g. 'I felt annoyed', 'I felt irritable'), challenge (e.g. 'I thought it was hard', 'I felt pressured'), negative affect (e.g. 'It gave me a bad mood', 'I found it tiresome') and positive affect (e.g. 'I felt content', 'I felt happy').

The GexpQ seems reasonable and applicable in studying player experiences with video games. However, it will more than likely need to be tailored for different genres of video games that do not involve a narrative or for which the story is intended to put the player in a bad mood (e.g. survival horror) and noncompetitive games where it is just fun to explore the game environment or play with the simulation.

However, the problem with the GexpQ is that IJsselsteijn and his colleagues have not yet published any of the empirical results using the questionnaire or psychometric evaluations on its reliability and validity and the properties of its subscales in a peer-reviewed journal. While the GexpQ appears to have face validity based on the semantics of the items, there is currently no published empirical support for its use. In materials from the Game Experience Lab, they report that they have published a version of the GexpQ in several languages and that it is currently being tested in a large-scale study. As it turns out, the Game Experience Lab did produce a deliverable as a part of a contract titled 'FUGA—The fun of gaming: Measuring the human experience of media enjoyment' funded by the European Community under the FP6 New and Emerging Science and Technology (NEST) programme (Poels *et al.*, 2009). While the report is very complete and does support the psychometric qualities of the GexpQ, it was neither generally publicized nor made available online.

Consequently, the use of the GexpQ questionnaire has preceded published evidence of its reliability. IJsselsteijn reports, 'The questionnaire has been successfully applied in many studies across different labs worldwide' (personal communication, 13 March 2012). The problem with this approach is that the GexpQ may very well gain an acceptance as the 'standard test' for game experience prior to any refereed publication of its worth. In a way, by publicly fielding the GexpQ prior to journal publication, IJsselsteijn and his colleagues in the Game Experience Lab at the Eindhoven University of Technology call into question the practical utility of the peer review process as the gatekeeper of standardized questionnaires.

3. THE GAME ENGAGEMENT QUESTIONNAIRE

Turning to the second article, Jean Brockmyer (formerly Funk) and her colleagues come from a social psychology background with a history of research on the negative effects of video games on children. Brockmyer *et al.* (2009) report on the development of another questionnaire pertaining to video games. In stark contrast to the first paper, they take great pains to develop the questionnaire over several testing iterations and validate the psychometric properties of the questionnaire prior to its use in the field. But as noted, the motivation for this questionnaire is quite different from the GexpQ of IJsselsteijn *et al.* Instead of exploring the gaming experience as play, Brockmyer *et al.* focus on the negative issue of video game violence. They note that violent video games are popular among many children (Funk, 2008) and that exposure to violent video games could result in increased aggression (Anderson *et al.*, 2007) and desensitization to violence (Carnagey *et al.*, 2007). The motivation for the GengQ is to develop a 'measure of engagement in playing video games that should be useful in assessing the potential impact of playing video games, particularly violent games' (p. 625).

Brockmyer *et al.* present a literature review similar to but more extensive than IJsselsteijn *et al.* to define and distinguish between the experiences of immersion, presence, flow and psychological absorption. In addition, they refer to a publication by Witmer and Singer (1998) that developed two measures for engagement in simulators and virtual environments. One has to do with a person's tendency to become immersed and the

second having to with aspects of the simulation that lead to a feeling of presence.

In the case of the GengQ, Brockmyer *et al.* lean toward an assessment of the individual child and his or her tendency to become involved in a violent video game rather than an assessment of the game itself. So, while IJsselsteijn *et al.* focus on the aggregate player's experience of a particular game, Brockmyer *et al.* focus on a particular individual's tendency to become engaged across a number of games. This seems appropriate since their concern is to identify individuals that are particularly susceptible to violent video games. They assert, 'When one is psychologically absorbed in a violent video game, the normal connections between the performance of violent actions and moral evaluation are broken' (p. 626).

In Study 1, Brockmyer *et al.*, similar to IJsselsteijn *et al.*, use theoretical construction from the literature to obtain a set of items. The primary difference, however, is that while the GexpQ sought to capture a broad range of player experience, the intent of the GengQ was more concerned with developing a one-dimensional scale. Consequently, rather than generating a relatively large number of items and using empirical results to eliminate redundant, unreliable or items meaningless to the participants, Brockmyer *et al.* started with only 10 items with a five-point rating scale. These were administered to 17 children in fourth through sixth grade. Additional items were created to expand the coverage of engagement. These were administered using a three-point rating scale to two different samples: 213 middle school students and 51 college students. Analyses indicated that additional items were needed and four more were added to complete the 19-item questionnaire. While dropping items from a questionnaire can be empirically justified, the addition of items can remain a mystery. Where did they come from? The authors do not explain why particular items were added. What happened during focus groups and in the thinking of the researchers coming up with new items? The final version was administered to a sample of 153 junior high school students as part of a larger study.

The authors collected important demographics and report on them. However, what was not clear was how participants' preference for violent games was computed. The authors report that it was 'computed by adding the number of games coded by participants into one of the three previously developed violence categories' (see Funk and Buchman, 1995 for a list of categories with descriptions). The categories and the instructions to the participants however were not given in the paper itself.

While Brockmyer *et al.* do a good job of reporting the details of their studies, they do not give much information about how the QengQ was formatted or what the instructions were to the students filling out the questionnaires. In addition, it is not clear whether the ratings refer to the games they listed or to games in general?

Nevertheless, the main strength of the paper is the psychometric development of the GengQ. Interestingly, Brockmyer *et al.* use the Rasch rating scale model (Andrich, 1978; Rasch, 1961) and reliability indices from Classical Test Theory (Spearman, 1907, 1913). These methods are more often applied in educational assessments where successively higher integer scores are intended to indicate increasing levels of competence. The Rasch is elegant in that it is a mathematical embodiment of the principle of invariant comparison. The comparison between two items in the test should be independent of which particular individuals were instrumental for the comparison; and it should be independent of which other items within the instrument were also compared. Moreover, a comparison between two individuals should be independent of which particular items within the instrument were used for the comparison; and it should also be independent of which other individuals were also compared. Rasch models embody this principle because their formal structure permits the algebraic separation of the person and item parameters. This forms the basis and justification for obtaining person locations on a continuum from total scores on assessments. Inherent in this is the assumption that the items and the persons are located on a single continuum unlike multidimensional models that position items and individuals in an *n*-dimensional space. So, unlike IJsselsteijn *et al.* who emphasize the importance of a broad spectrum for assessing player experience on many different dimensions, Brockmyer *et al.*, focus on a single unidimensional continuum which they call 'engagement'. They provide additional evidence of the unidimensionality of the scale using a Rasch principal components analysis, reporting that the total variance explained by the measure was 69%.

Brockmyer *et al.* present a table in which the items are ranked by difficulty. But of course, since there are no correct answers on a rating scale, the term 'difficulty' seems a bit inappropriate harkening back to educational testing. Instead since the ranking is merely based on the mean rating for the items, it would make more sense to use another construct relevant to the current application such as 'level of engagement'.

On the other hand, since the purpose of the GengQ was to create a single-dimensional engagement scale, it initially seemed odd that the authors categorized the items in the table and associated them with the labels absorption, flow, presence and immersion. But apparently they intended to imply that increasing levels of engagement are associated with the constructs in the order that immersion is the lowest, presence is mediate low, flow is mediate high and absorption is the highest. It is clear that these constructs came from the review of the prior literature; however, it is not entirely clear how the particular items came to acquire these labels. Were they items on previous questionnaires that purported to measure these constructs? If so, their association is very dubious. Moreover, since the number of items in each category varies from only one in the immersion set up to nine in absorption set, the items may not be a representative set of items for these constructs. The authors provide no empirical evidence in this study to support these constructs.

However, it does seem appropriate to have items in the questionnaire that range from mild to extreme in order to capture the variability in the sample of individuals. Similarly, it is important that the sample be diverse, which seems to be the case, although participants were all in the same age range.

The authors report good reliability statistics for the GengQ with a Cronbach's alpha of 0.85. They also report a Rasch estimate of person reliability (the Rasch analog to Cronbach's alpha) of 0.83 and the item reliability of 0.96. In Study 1, they base the validity of the instrument on the match between the empirical ordering of the item difficulties and the theoretical order and on the fit of the items and rating scale categories to that unidimensional measure.

Study 2 was designed to validate the GengQ by demonstrating that individuals with higher scores would demonstrate more engagement while playing a video game. In Study 2, participants completed the GengQ before playing a game in terms of their usual reactions to video games. Then they played the game, 'S.T.A.L.K.E.R.: Shadow of Chernobyl' in the laboratory for 25 min. At this point, an audiotape was played with three statements at slightly increasing conversational levels.

The first two statements were 'Excuse me, did you drop your keys?' The last was 'Did you drop your keys?' Observers coded the participant's reaction or lack thereof. They were allowed to play the game for an additional 5 min and then completed the GengQ in terms of their reaction the game they played.

The GengQ before and after were highly correlated and the score on the first GengQ tended to predict participant reaction to the audiotape. The GengQ was also significantly correlated with the dissociative experiences scale (DES) (Carlson Bernstein and Putnam, 1993) and the aggression questionnaire (AQ) (Buss and Warren, 2000).

Consequently, the GengQ seems to be a valid measure of engagement with some limitations. The authors report that they used frequent male gamers to increase the chance that some participants would experience deep engagement in playing video games in a laboratory situation. However, it is clear today that over 92% of females play video games. In addition, the age range of 18–26 is highly restricted, since the average age of gamers is 40 and the initial motivation for the study pertained to middle school children (Lenhart *et al.*, 2008).

4. PSYCHOMETRIC ISSUES

Considering the two papers reviewed here in contrast, several issues become apparent which will be discussed below.

4.1. Uni- or Multi-dimensional Assessments

Both IJsselsteijn *et al.* (2007) and Brockmyer *et al.* (2009) refer to previous literature trying to define a host of concepts of involved with game play experience and engagement; namely immersion, flow, presence, absorption and dissociation.

The question is whether these are separate constructs and consequently should be assessed as different factors in a multidimensional space or whether they are different names for the same construct and only one scale is needed.

IJsselsteijn *et al.* lean toward a multidimensional space but only partially related to these constructs. The GexpQ has subscales labeled: competence, sensory and imaginative immersion, flow, tension/annoyance, challenge, negative affect and positive affect. But, unfortunately, there is yet no empirical evidence as to the validity of these scales.

Brockmyer *et al.* strongly assert that engagement is a unidimensional scale. Moreover, they provide empirical evidence using the Rasch principal components analysis to support this claim. To them, the constructs of immersion, presence, flow and absorption are actually points along one scale indicating increasing levels of engagement, but this may be an artifact of the particular 19 items they selected to be a part of the questionnaire. Whether it is unidimensional or not, it is likely that multiple components of engagement may be introduced by different aspects of the digital game. Thus, the GexpQ attempts to tape into factors relating to engaging aspects of skillful performance and challenge, the game narrative and environment, and positive and negative emotions.

4.2. Are We Evaluating the Individual or the Game?

In studies of usability, we emphasize the importance of user characteristics. But generally, we do so for the purpose of obtaining representative participants or for designing for different levels of experience. For power users, we can design more complex systems with more features; and for novice or first time users, we will need simpler, more intuitive interfaces. But generally, we are in the business of assessing the usability of the interface rather than the abilities or characteristics of the user. Even when we evaluate user performance on different systems, we are interested in assessing the system, not the user. In fact, in many usability tests, the task administrator instructs the user, 'We are not evaluating you or how well you perform, but rather we are evaluating the software and the system'.

However, with the shift to user experience, this distinction can start to get a little muddied. Since the experience we want to study is in the mind of the user, an interaction can occur between how the system creates that experience and how the user's prior experiences, susceptibility and engagement affect the experience. The user is not going to have a good experience on an ill-designed system that is hard to use, frustrating or boring. But a user who is a good problem solver, a persistent worker and handles difficulty well may actually have a better experience with a challenging interface than one who has a low tolerance, low self-esteem and feelings of low self-efficacy. This is particularly true in the area of game usability and playability where we focus on player experience. So, user characteristics can affect usability, as does the design of the interface. Only recently have studies attempted to tease the

two apart. For example, Campbell and Norman (2010) looked at users with high versus low spatial visualization abilities searching for information in Wikipedia. Spatial visualization ability and system experience predicted system effectiveness measures, while spatial orientation ability, spatial visualization ability and general computer experience predicted system efficiency measures.

In the case of the two articles reviewed here, Brockmyer *et al.* focus on a scale of engagement as a player characteristic, while IJsselsteijn *et al.* focus on player experience as a way to evaluate games, but neither article attempts to reconcile the interaction between the two. It does seem that there is an inherent interaction in individual differences and games played.

My own lab is working on the separation of characteristics of the player and features of the game (Norman, 2010). Following the lead of Witmer and Singer (1998), we have developed two measures, one for the susceptibility of the player to become immersed (immers ability of individuals, IAI) and the other for aspects of the game that make it immersive (immersive ness of games, ING). The IAI is composed of four susceptibility factors: daydreaming, video games, media and sports. The video game factor is most similar to the GengQ. The ING is composed of four immersive factors: visual presence, audio presence, sensory engagement and sense of control. While related to the GexpQ, the ING focuses more on the attributes of the game.

5. CONCLUSION

In my opinion, the articles by IJsselsteijn *et al.* and Brockmyer *et al.* were worthy of publication. IJsselsteijn *et al.* makes an important contribution emphasizing the multifaceted nature of the game experience and the need for an instrument to assess what it is about the game that contributes to the engaging experience.

Brockmyer *et al.* also emphasize the need for a measure of game engagement as factor of one's personality, and they actually develop and test the questionnaire. My complaint is that while somewhat justifiable from a policy and funding prospective, their focus on violent video games and aggressive behavior is problematic. It associates the GengQ with the anti-video game group represented by Anderson *et al.* (2007) which implies that engagement in video games has a negative valence. It thereby limits the application of the GengQ to use by the opponents of violent video games rather than providing a standardized assessment tool for the wider community who value the positive side of engagement. This group, of course, includes the video game industry; but also includes the advocates of video games who point to the positive neurological effects of video games (e.g. Green and Bavelier, 2006), their educational value (e.g., Gee, 2003), and their use in educational settings (e.g., Squire, 2003). It is precisely in these contexts that engagement is a positive trait of the individual and desired effect of a video game.

Consequently, it is unfortunate that the one instrument, the GengQ, that has proven reliability and validity is the one that is most politically distanced from the field and the other, the GexpQ, most used in the study of video game experience has yet to be published results on its psychometric qualities.

REFERENCES

Anderson, C.A., Gentile, D.A. and Buckley, K.E. (2007) Violent Video Game Effects on Children and Adolescents: Theory, Research, and Public Policy. Oxford University Press, Oxford.

Andrich, D. (1978) A rating formulation for ordered response categories. Psychometrica, 43, 357–374.

Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M. and Pidruzny, J.N. (2009) The development of the Game Engagement Questionnaire: a measure of engagement in video game-playing. J. Exp. Soc. Psychol., 45, 624–634.

Buss, A.H. and Warren, W.L. (2000) Aggression Questionnaire: Manual. Western Psychological Services, Los Angeles, CA.

Campbell, S.G. and Norman, K.L. (2010) Cognitive abilities and the measurement of World Wide Web usability. In Human Factors and Ergonomics Society Annual Meeting Proceedings, vol. 54, pp. 1092–1096.

Carlson Bernstein, E.B. and Putnam, F.W. (1993) An update on the dissociative experiences scale. Dissociation, 6, 16–27.

Carnagey, N.L., Anderson, C.A. and Bushman, B.J. (2007) The effect of video game violence on physiological desensitization to real-life violence. J. Exp. Soc. Psychol. 43, 489–496.

Funk, J.B. (2008) Video Games. In Strasburger, V., Wilson, B. and Jordan, A. (eds), Children, Adolescents, and the Media (2nd edn), pp. 435–470. Sage, Thousand Oaks, CA.

Gee, P.J. (2003) What Video Games Have to Teach us about Learning and Literacy. Palgrave/Macmillan, New York.

Green, C.S. and Bavelier, D. (2006) The Cognitive Neuroscience of Video Games. In Humphreys L. and Messaris, P. (eds), Digital Media: Transformations in Human Communication, pp. 211–224. Peter Lang Publishing, New York, NY.

IJsselsteijn, W., de Kort, Y. and Poels, K. (2007) Characterising and Measuring User Experiences in Digital Games. In ACE Conference'07, June 13–15, Salzburg, Austria.

Lenhart, A., Kahne, J., Middaugh, E., Mcgill, A.R., Evans, C. and Vitak, J. (2008) Teens, Video Games, and Civics. Pew Internet and American Life Project.

Norman, K.L. (2010) Development of Instruments to Measure Immerseability of Individuals and Immersiveness of Video Games. Technical Report LAPDP-2010-03, HCIL Technical Report 12-5-10, University of Maryland, College Park, MD 20742.

Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R.L. and Fuller, T. (2003) User-centred Design in Games. In Jacko, J.A. and Sears, A. (eds), The Human–Computer Interaction Handbook, pp. 883–906. Erlbaum, Mahwah, NJ.

Poels, K., de Kort, Y.A.W. and IJsselsteijn, W.A. (2009) Game Experience Questionnaire. STREP/NEST-PATH Deliverable D3.3,

Game Experience Lab, University of Technology of Eindhoven, NL.

Rasch, G. (1961) On General Laws and the Meaning of Measurement in Psychology. In Proc. 4th Berkeley Sympos. Math. Statist. Probab., IV, pp. 321–334. University of Chicago Press, Berkeley, 1980.

Spearman, C. (1907) Demonstration of formulae for true measurement of correlation. Am. J. Psychol., 15, 72–101.

Spearman, C. (1913) Correlations of sums and differences. Br. J. Psycho., 5, 417–426.

Squire, K. (2003) Video games in education. Int. J. Intel. Simul. Gaming 2, 49–62.

Witmer, B.G. and Singer, M.J. (1998) Measuring presence in virtual environments: a presence questionnaire. Presence, 7, 225–240.