
Usability, Playability, and Long-Term Engagement in Computer Games

Alessandro Febretti

Milestone Games
Via Gustavo Fara 35 - 20124 Milano, Italy
and
University of Illinois at Chicago
Electronic Visualization Laboratory
842 W. Taylor Street - 60607 Chicago (IL), US
alessandro.febretti@gmail.com

Franca Garzotto

Politecnico di Milano
School of Industrial Design and
Department of Electronics and Information
Piazza Leonardo da Vinci 32- 20133 Milano, Italy
franca.garzotto@polimi.it

Copyright is held by the author/owner(s).

Abstract

Does usability affect long term user engagement in computer games, or are other factors more influential? This paper explores this issue, discussing an evaluation study that measured the relevance of usability versus playability factors for long-term user engagement in eight commercial games.

Keywords

Interactive game, engagement, usability, playability, heuristic evaluation, user testing, correlation

ACM Classification Keywords

H5.2. User Interfaces – *Evaluation/Methodology*

Introduction

Long term engagement can be defined as the degree of voluntary use of a system along a wide period of time (i.e., weeks, months, or years), involving dozens, if not thousands, of interactions, each one spanning for significantly longer than few seconds or minutes. Achieving long term engagement is an important concern in a number of domains, one such domain is represented by the so-called “long games”, which represent a large portion of interactive entertaining products developed at industrial level. In long games, a single play sessions normally lasts one hour or more, but a common industry practice is to design for a total

play time in the order of tenths of hours; several games (especially the ones based on multiuser interaction, like MMOGs - Massive Multiplayer Online Games [3]) can even offer a theoretically unlimited gameplay experience. The fact that (many) people are using or have used a game repeatedly, for significantly longer than few seconds or minutes, and will keep using it after weeks or months or years, is an indicator of users' satisfaction, makes the product more reliable for players, increases their pleasure of "being part of the community", and ultimately contributes to the "quality of the experience".

The general scope of our work is the investigation of design factors that may affect long term user engagement in the above class of systems. Long games, like any interactive application, expose an interface between the user and the application itself (in this case, the game world). Thus we may suppose that *usability*, which is considered a key factor for the "quality of the experience" in any interactive system, should in principle affect user engagement in long games too. In particular, we may expect that usability problems could reduce user engagement in the long term. On the other hand, a games' primary objective is not being efficient and effective in completing a specific task: they instead need to succeed in providing *overall fun* to the user, which is the fundamental factor that motivate users to continue use of the game over time.

We therefore want to address the following questions: does usability truly influence long term engagement? In particular, do usability problems reduce long-term engagement? Or is long term engagement more strongly related to other design features, those

addressing "playability" per se, which may overwhelm the influence of usability defects?

The Evaluation Study

As an initial step towards answering these questions, we have carried on an evaluation study on *eight commercial long games*. The overall process followed in our work is sketched in figure 1.

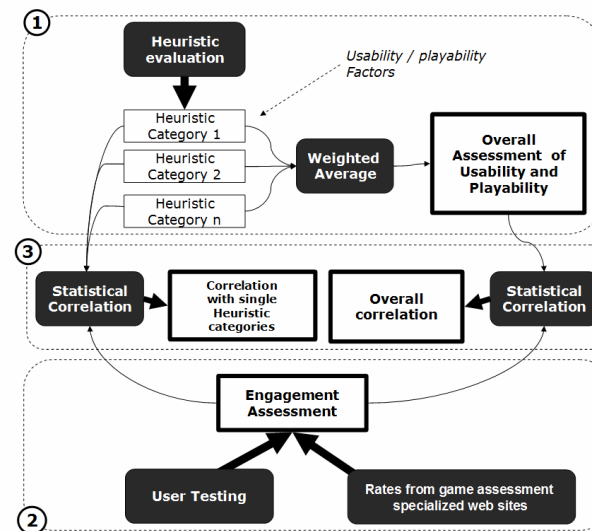
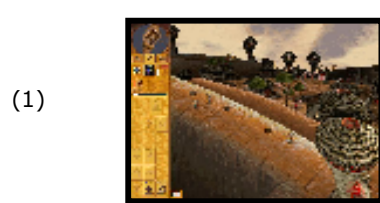


Figure 1. The evaluation study at a glance-

In each of the evaluated games we measured user engagement, usability and playability, and the correlation between engagement and usability/playability. We decomposed usability and playability into a number of more measurable factors, defined on the basis of the current literature on usability and enjoyment in games; we assessed them using heuristic evaluation [10] carried on by *twenty experts* in game design and usability.



Evaluated
Games

User testing was adopted to empirically measure engagement, involving *forty-seven* users. To increase the reliability of these empirical measures, we integrated them with the engagement “scores” for the inspected games provided by specialized web sites for game quality assessment [7, 9]. Table 1 reports the set of eight games that were considered in this analysis. We selected them from a larger pool of pre-analysed games, according to the following criteria: i) to be well known, professionally-developed, successful titles published in the last ten years ii) to be referred by specialized web sites for game quality assessment [7, 9] (to enable us to derive additional engagement measures); iii) to have at least one significant usability problem that clearly emerges at some point of the gameplay.

	Title	Year	Genre
1	Popolous	1999	God-game
2	Age of Empires 2	1999	Real-
3	Diablo II	2000	Action /
4	Stronghold	2001	Real-
5	Neverwinter Nights	2002	Single
6	World of Warcraft	2005	Online
7	Dragonball Z	2006	Fighting
8	Shayia http://shaiya.aeriagames.com/	2007	Online RPG

Table 1. Evaluated games.

Playability and Usability Heuristics


From the various proposals of playability heuristics available in the current state of the art [1,2,5,12], we selected *twenty-two* heuristics we deemed more


significant for our games, and we categorized them in seven different classes. What follows is the list of the identified heuristics classes, with an example heuristic taken from each class:


- **Concentration and Immersion (7 heuristics)** – Example: “*The game is able to capture and keep the player’s attention*”.
- **Challenge (2 heuristics)** - Example: “*The game should provide different challenge levels for different players*”.
- **Player Ability (4 heuristics)** - Example: “*Actions available to the player should increase as the player progresses through the game*”.
- **Control (2 heuristic)** - Example: “*The player should feel in control of his/her own actions in the game*”.
- **Objectives and Feedback (3 heuristics)** - Example: “*The player should always be aware of his/her objectives inside the game*”.
- **Social Interaction (for multiplayer games only, 2 heuristics)** - Example: “*The game should support and ease communication between players*”.
- **Artificial Intelligence (2 heuristic):** “*The artificial intelligence should provide unexpected behaviors*”.


Similarly, the general usability heuristics found in literature [10] were filtered to discard those that were not significant in a computer game scenario, we slightly customized some of them, and introduced some new ones basing on game industry practice [4, 5, 11]. Five heuristics classes were identified for a total of fifteen usability heuristics:


- **Customization (3 heuristics)** - Example: “*The game should support user customization of audio and video settings*”.
- **Controls (4 heuristics)** - Example: “*Default commands should be mapped intuitively and*

- (5)  consistently with respect to standards defined by the game genre".
- **Game Views (2 heuristics)** - Example: "Game views should be designed appropriately depending on the specific activity inside the game".
 - **Interface/Layout (4 heuristics)** - Example: "The game interface should be intuitive and immediate to let the player keep focus on gameplay".

- (6)  Three or four experts were assigned to each game to carry on the inspection. Before any inspection activity, the sets of categorized heuristics, their metrics, some general procedural guidelines, and the analysis and reporting format were discussed and agreed among all inspectors, to perform a common and consistent process and to achieve a coherent and comparable set of judgments.

- (7)  **Evaluating Engagement**
A qualitative measure of user engagement was first estimated through supervised and unsupervised user testing. Each game was submitted to a different user group, recruited by the experts assigned to the inspection of that game.

- (8)  A group always consisted of five to eight users, that had different levels of experience in the specific game under analysis, from no previous exposure to previous play once or more times.

- (8)  In a test session, lasting approximately sixty or seventy minutes, each user had to complete a number of scenarios. Scenarios were custom for each game and were designed to expose users to all the most significant gameplay situations of the system, so that players could have a realistic experience of the game as it would be played in a spontaneous, "normal" condition.

Evaluated Games

Evaluators' observations during gameplay, interviews and questionnaires at the end of the play experience provided the qualitative and quantitative data that allowed us to elaborate an initial assessment of the engagement level offered by the game. These measures were then compared and integrated with data coming from game rating aggregators available on specialized web sites for game quality assessment [7, 9], to gain a more precise and reliable measure of game engagement.

Investigating Correlations

Heuristics-based inspection of each game resulted in a set of measures for each usability and playability attribute. Through a weighted average of these values we obtained a global estimation of the game compliance to the chosen usability and playability criteria. Statistical correlation analysis was carried out using Spearman's rank correlation test [8] between categorized and global indicators of usability and playability, and the game engagement rating. The resulting correlation coefficients between global playability/usability estimates and game engagement are shown in Table 2.

Correlation type	Correlation value
Playability ↔ Engagement	0.760 (> 0.738, p=0.05)
Usability ↔ Engagement	0.515

Table 2. Correlation between global heuristics / usability estimates and game engagement

Comparing the estimated correlation coefficients with thresholds for 0.05 significance level allows us to identify a *significant relationship* between the analyzed *gameplay heuristics* and the actual level of user *engagement* in the game. The *relationship* between the considered *usability* factors and *engagement* instead, does *not* appear to be *statistically relevant*. This difference in correlation levels is evident also in estimations between single heuristics classes

and game engagement (figure 2): *Correlation levels for playability heuristics classes are significantly higher on average than correlations for usability classes.* It is worth noting how the playability heuristics in the *challenge* category appear to be strongly linked with engagement. These factors cover the aspect of *balance* in the game challenge level, in relation to both player initial skills and progress throughout the game.

The fundamental nature of these aspects for enjoyment in gameplay is well known [1, 2, 12], but it has been seldom analysed in relationship to long term

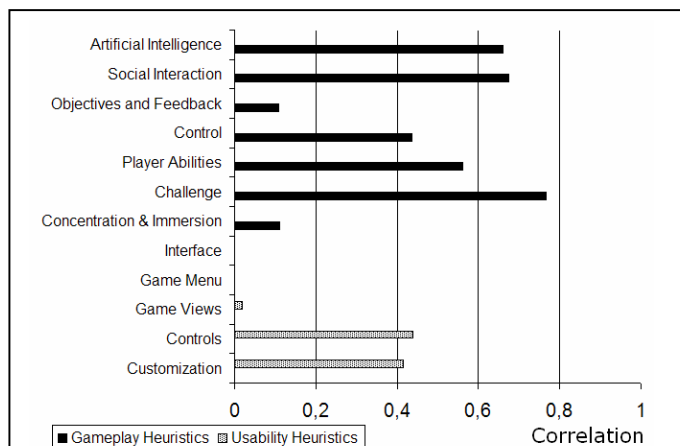


Figure 2: Correlation between heuristic classes and long term engagement

engagement. It is also interesting to observe how the two usability heuristics classes that show a stronger correlation with engagement are the ones that deal with *customization* and *control*. User control on the actions required by the game and the capability of tuning game features to one's needs may be a fundamental aspect of gameplay per se, and in general contributes to enjoyment and engagement [12]. Thus an interface that makes the performance of such actions difficult slows down the game flow and is more annoying than, for example, a menu that is hard to navigate or a lay-out that is not fully intuitive.

Discussion and Conclusions

Although a number of studies have investigated separately usability and playability in gaming experiences [1, 2, 4, 5, 6, 11, 12], to our knowledge this is the first attempt that explicitly investigates the relationships among these factors and long term engagement, in the context of long professional games. Our study shows that engagement appears to be significantly related to playability factors in long professional computer game

experiences, but statistical analysis highlights a *low correlation between usability and long term engagement* in the same context. The latter result is apparently surprising, since during user testing we indeed observed a *local, momentary* decrease in the user feeling of "being engaged" with the game experience in each situation where a usability problem manifested itself during a play session. Still, unless the usability defect was so severe to prevent any further action, the same users tended to find original strategies to overcome the problem and continued playing. Figure 3 offers a qualitative, intuitive illustration of this behavior. When patterns like this manifested, most players declared at the end of the session to be satisfied of the experience anyways, and would play the game again.

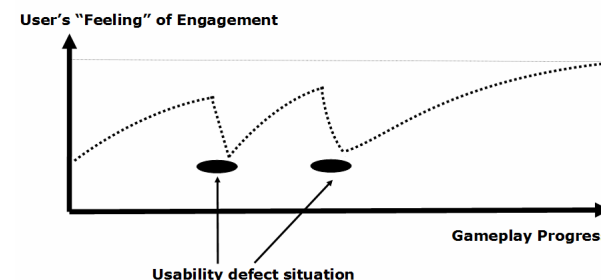


Figure 3. Typical Flow of Engagement Perception

The results of our study suggest a general methodological hypothesis on the relationship between usability and long term engagement: we will formulate this hypothesis using the concept of *Density of Usability Defects (DUD)*. DUD indicates the amount of usability defects a user will face for a specific gameplay length, defined as $DUD = UP/T$, where UP is the *number of usability problems* for the *game experience duration T*.

Our hypothesis is that *long term engagement is affected by the value of DUD, rather than the absolute value(s) of usability (heuristics)*, because factors that have a *constant, pervasive* influence on the quality of

the gaming experience (like the ones captured by playability heuristics) dominate on usability.

Playability factors are more relevant for users' *overall feelings of engagement* during the game and represent the main drivers that foster or dissuade them to play again. If DUD is low, (i.e., a specific usability problem manifests itself with limited frequency during a play session), the decrease in user engagement in the game experience tends to be *temporary* and *local*, remaining confined to the moment of defective interaction, without significantly affecting the *overall* engagement.

The above hypothesis is supported by the results of our study, but further empirical investigation is needed to confirm it.

Also, our work considered a specific category of games - long industrial games - which are normally subject to strong quality assessment (QA), so that frequent user interface problems are noticed, and removed, during the QA phase of industrial production. In this kind of systems, even when usability is non excellent, DUD is usually low, i.e., usability problems have an *episodic, sporadic* nature, and a specific interface weakness manifests itself in a limited number of hot spots inside the gameplay experience.

Our future work will verify our methodological hypothesis investigating other categories of games, e.g., those involving short play sessions like web-based minigames or casual games. We expect that a game of this kind having the *same* type of usability problems as a long game would tend to manifest these defects more frequently, due to the simpler and shorter gameplay structure, so that usability weakness can have a deep influence on overall engagement. Similar considerations may apply to games developed by non-professional teams, or games that do not go through a consistent Quality Assessment process. In these development scenarios, even complex and long gameplay structures

can be damaged by unidentified usability problems, affecting long term engagement.

References

- [1] J. Chen, "Flow in Games", <http://jenovachen.com/flowingames/thesis.htm>
- [2] Desurvire, H., Caplan, M., Toth J.A., "Using Heuristics to Evaluate the Playability of Games". Proc. CHI 2004, April 2004, pp.457-466
- [3] Ducheneaut, N., Yee, N., Nickel, E., Moore, R., "Alone Together" Exploring the social dynamics of massively multiplayer online games". Proc. CHI2006, pp. 407-416
- [4] Federoff, M. "User Testing for Games: Getting Better Data Earlier". Game Developer Magazine, June 2003
- [5] Federoff, M. "Heuristics and Usability Guidelines for the Evaluation of Fun in Video Games". Thesis at the University Graduate School of Indiana Univ., Dec. 2002
- [6] Garzotto F. "Investigating the educational effectiveness of multiplayer online games for children" Proc. ACM IDC 2009, pp. 29 - 36
- [7] "IGN", IGN Entertainment, <http://www.ign.com>
- [8] Graziano A.M., Raulin M.L. "Research Methods - A process of Inquiry". Pearson, 2007
- [9] "Metacritic", <http://www.metacritic.com>
- [10] Nielsen, J., Molich, R.L., "Usability Inspection Methods", John Wiley & Sons, 1994
- [11] Pinelle D., Wong N., Stach T., "Heuristic Evaluation for Games: Usability Principles for Video Game Design", Proc. CHI 2008, April 2008, pp. 1753-1762
- [12] Sweetster P., Wyeth P., "GameFlow: A Model for Evaluating Player Enjoyment in Games ", ACM Computer and Entertainment, 3, 3. 2005, pp. 1-23.