Top Versus Bottom: Game Evaluation from an Expert or Player Perspective

Björn Strååt
Stockholm University, Department of Computer and Systems Sciences/DSV
Forum 100
SE-164 40 Kista, Sweden
+46-709-723 222
bjor-str@dsv.su.se

Harko Verhagen
Stockholm University, Department of Computer and Systems Sciences/DSV
Forum 100
SE-164 40 Kista, Sweden
+46-8-16 16 94
verhagen@dsv.su.se

ABSTRACT

In recent years, many video game researchers have developed domain specific heuristics for video games. Heuristic evaluation is a common way to measure software usability, both during development and to assess existing systems. Video games are very different from productivity software in terms of purpose, design and execution, and thus require specially developed heuristics. Heuristics developed for video games can however only guide the game developers, and are in themselves no guarantee of a positive game experience.

The questions examined in this paper concern the extent to which heuristics developed by experts can address the issues that the end users – the players – experience. Furthermore, we examine what kinds of design issues are most frequent, and whether these issues are reflected in online user comments. The results show that users focus their criticism on narrative, game graphics and sound rather than the issues that the heuristics involved in this study were designed to counter.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: *valuation/methodology*; H.5.2 [**User Interfaces**] *Evaluation/methodology, User-centered design, Style guides*

General Terms

Design, Documentation, Human Factors, Standardization

Keywords

Games, Experience, Heuristics, Evaluation, User comments

1. INTRODUCTION

Commercial off-the-shelf Video games (video games) are designed to be entertaining. They should give the user an experience of joy, thrill, challenge, and so on. The video game medium should deliver the best possible experiences as envisioned by the game designers. Game designers of today have access to a plethora of tools that help prevent or remedy poor design choices. Unfortunately, even with these tools and processes at hand, many video games are released with design flaws; flaws that might ruin the gaming experience for the user and in the end lead to poor sales for the developers.

Many of the available software evaluation methods were developed with productivity software in mind and not designed to meet the specific needs of video game software [1]. While productivity software should be completely free of challenges, video games are designed to be challenging, thus making evaluation tools that are meant to remove challenges inadequate to evaluate video games. In order to address this issue, many game researchers have developed, and are still developing, lists of heuristics for evaluation of video games. Game usability is an active research field and several methods for evaluation of game design have been developed by researchers and put into use by the game design industry. Federoff [2] was an early pioneer and others have followed. Desurvire, Caplan and Toth [3] and Desurvire and Wiberg [4] are examples of heuristic evaluation sets, which game designers have recognized and are applying on their work. Many researchers in the field of game usability have developed their tools following the method for producing heuristics for productivity software, established by Nielsen and Molich [5]; by evaluating usability issues identified by experts. This implies a top-down approach, where the researchers' knowledge and the expertise of video game designers has been the conceptual basis for development.

The purpose of this paper is to examine to what extent a selection of video game heuristics can help prevent poor design choices, what type of issues heuristics fail to measure, and how well heuristics designed by a top-down approach matches the players' view on video game issues. To this end, we performed two studies. In the first study we evaluated a set of contemporary video games from an expert view, using domain specific heuristics. In the second study, we analyzed user comments from metacritic.com on the same games. The results from the studies showed that the findings from the expert evaluation did not coincide with the user comments.

We have only worked with commercial off- the-shelf video games in our studies. This does not suggest that we do not acknowledge the immense world of independent games, art project games, and serious games, and so on and so forth. The limitation is due to the original objective of the heuristic lists involved – they were all created for commercial video games.

Proceedings of the 10th International Conference on the Foundations of Digital Games (FDG 2015), June 22-25, 2015, Pacific Grove, CA, USA.
ISBN 978-0-9913982-4-9.

2. BACKGROUND

2.1 Origins of Heuristics

Usability inspection methods, such as heuristic evaluations and cognitive walkthroughs, are cheap and quick to perform, as no users need to be recruited. The evaluations can be performed on early prototypes and throughout all stages of a project. The heuristic evaluation method is a "quick-and-dirty" evaluation and rating method to identify usability issues in digital interfaces. A small number of evaluators - two to five - individually perform tasks in a system, while taking notes of any issues they encounter. The issues are rated according to severity, and matched against the list of heuristics. The method is low cost and does not require many resources in time or setup. Often applied at early stages of design, it allows the designers to evaluate their work and to avoid issues before release. If used on a finalized interface, it can evaluate and suggest improvements.

Nielsen and Molich [5] proposed a set of ten guidelines to be used as design principles for digital interfaces. The guidelines are intended both for production of new systems and evaluation of existing ones. The name "Heuristic", as in "A rule of thumb" is used since the design principles are no more than principles and should not be regarded as a hard ruleset. Nielsen [6] collected 249 usability issues and based on factor analysis developed ten heuristics.

2.1.1 Top-down Heuristics or Bottom-up Heuristics

Jaferian et al introduce the idea that there are two main approaches to creating heuristics: top-down or bottom-up [7]. The "top-down" approach consists of data gathered from theories, domain high-level expert knowledge, and other heuristics within the domain. If the data is derived from field studies, interviews with end users and domain observations, it should be considered to be "bottom-up" (ibid). It would be presumptuous to claim that one approach is better than the other, but it is important to acknowledge that there is a difference. There is a risk that a predominantly top-down approach will fail to capture certain aspects of the end users needs, and a predominantly bottom-up approach may fail to take expert advice into consideration. To overcome the shortcomings of the approaches, a combination of both data gathering methods is recommended.

2.2 Video Games and Heuristics

Productivity software, such as a word processors, automatic bank teller machines, or spread sheet applications are tools designed to assist or automate as much of the users' tasks as possible. The software should be as free of obstacles and challenges as possible to minimize the stress of the user. Productivity software applications are tools first and foremost, and designers of productivity applications should always strive to optimize the ease of use for the end user. Video games are designed for entertainment and should be optimized for the users' entertainment, while still being easy to use. The interaction of video games should be user friendly, and the tasks of the video game should be challenging. While productivity software often is designed to automatically solve complex tasks for the user and strive to minimize the required input, most video games require constant interaction from the user during play. These differences mean that games need domain specific heuristics, separate from productivity software. Laitinen [8] showed that Nielsen's 10 heuristics [6] worked to a certain degree, but they do not fully grasp the nuances of video games

2.3 Video Game Heuristics

Most of the heuristics in table 1 are based on expert opinion. Very few take actual end- user considerations into account; Pinelle et al come close with their reviewer evaluations, Desurvire and Wiberg's survey may count as a measure of user experience; however they do not directly address the actual users when creating their lists of heuristics.

Author	Approach	Details
Federoff [2]	Top Down	Literature study combined with on site participant observation
Desurvire et al [3]	Top Down	The HEP (Heuristic Evaluation for Playability) list is based on Federoff [2] articles, and interviews with domain experts
Pinelle et al [9]	Top Down /Bottom Up	10 heuristics based on video game reviews from the Gamespot.com website. The authors of the reviews are domain experts (professional game reviewers) but acting as end users, which makes this a combined approach.
Desurvire and Wiberg [4]	Top Down	The PLAY (Principles for Game Playability) is based on the HEP list [3]

Table 1: Lists of video game heuristics used in this paper, and the predominant design approach methods.

3. VIDEO GAME ISSUES

3.1 Our Definition of Issues in Video Games

Good usability enhances the user experience, sometimes even turning the users focus completely on the task. Csikszentmihalyi [10] call this phenomenon Flow and mentions numerous situations when Flow can be reached: playing an instrument, athletic performances, playing games etc. Csikszentmihalyi [ibid] describes the sense of flow: *"the state in which individuals are so involved in an activity that nothing else seems to matter*".

Jennet et al [11] suggest that Flow and immersion are closely related; immersion in video games requires concentration, a sense of challenge, control and emotional involvement. Immersion is the way a player feels connected to the activity of playing the game or *"the sensation of being surrounded by a completely other environment […] that takes over all of our attentio*n…" [12]. The concept is a state of mind when a player experiences a high degree of focus and enjoyment, loses track of time and sometimes the awareness of the real world (ibid.). If there are design elements that do not support immersion or actively break it, the player's experience of focus and enjoyment is interrupted.

Rollings and Adams [13] describe video games as a set of rules that define obstacles, or challenges, that players need to overcome if they want to win the game. Suits [14] defines a game as "the voluntary attempt to overcome unnecessary obstacles", where the obstacles are challenges set by the rules of the game. Both Suits and Rollings and Adams describe a specific end state or result that the player needs to achieve to win the game or to conquer a challenge. The rules of the game tell the player what actions are allowed. In a video game, the interface of the game can be considered to be a part of the rules, as it dictates the possible actions.

If we, as players, accept the rules of the game, and willingly partake in playing, we voluntary attempt to overcome obstacles set by the game rules. These obstacles are often represented by various challenges. The goal of the game should be clear, and the challenges should be possible to identify and to interact with. Challenges in a video game should appear in the game-play, not in the interaction with the video game software. Regardless of whether a challenge lies in finding out what to do next or in performing a specific task, the design of the video game must afford all possible options to the player. Juul [15] states that even if a game is easy to use in terms of manipulating the game world, it still can be difficult in terms of challenging game events.

Linderoth [16] describes how challenges in a game can be expressed as either a search for things players can do in the game, called exploratory challenges, or in the execution of actions, called performatory challenges. For example, deciding what move to make in a chess game would challenge the player's exploratory skills while hitting bull's eye with a bow and arrow would challenge the player's performatory skills. Linderoth uses the concept of affordance, as defined by Gibson [17], to describe the ability to find and execute actions - the world around us affords using, based on our abilities. If a performatory or exploratory challenge is not properly afforded, we will not be able to act on it and instead have a performatory or exploratory restriction, which in essence would indicate a game challenge issue.

To conclude: the game experience lies in the game's ability to engage the player's emotions, through immersive elements such as audiovisual, challenges, or narrative. Our definition of video game issues is when design elements or game challenges have poor usability and/or does not support and/or actively breaks the player's sense of immersion.

4. THE STUDIES

4.1 Study 1: Evaluation of Game Heuristics

The purpose of Study 1 [18] was to see if there was a connection between violation of heuristics and game rating. A total of ten games were evaluated according to the heuristic evaluation method as described by Nielsen [6]. . Four evaluators evaluated each of the ten games by playing them for an average of two hours. The evaluators are double experts, in the sense that they all have a vast experience of video games as well as usability evaluations. Each evaluator reported all the issues they found during the play sessions. The issues were then analyzed using a set of heuristics. The set of heuristics was compiled from the lists presented in table 1, and each heuristic value was picked based on the definition mentioned in section 3.1. The resulting list of heuristics is called The Net Heuristic List and consists of 14 heuristics.

The video games were selected based on their popularity rating on metacritic.com[1]; low rated games with a score of 40/100 or less and high rated with a score of 80/100 or more.

4.1.1 Results Study 1

We found that the low rated games violate more heuristics than the high rated games do, and that some issues were not represented in the heuristics at all. The study also showed that the low rated games violated game usability heuristics much more than high rated games do.

The majority of the issues found were usability issues such as lack or loss of player control, unclear interaction interface and poor visibility of possible actions.

4.2 Study 2: Analyzing Player Reviews

2 [19] focused on the players' ideas on game design issues. We gathered all user comments at metacritic.com from three of the games that were used in Study 1 [18]. Based on the metacritic ranking we selected one high ranked, one medium ranked and one low ranked game. By using AntConc[2], which is a word concordance software, we derived three word categories of the most frequent words of each game. The categories were "Adjective or Adverb", "Genre Description" and "Game specific". The resulting word lists were then viewed in the context sentences, which in turn were analyzed with the same heuristic set that was used in Study 1.

4.2.1 Results Study 2

The result showed that the players focused on issues with storyline, audiovisual elements and sense of immersion. This strongly indicates that the view of the experts (the top-down approach) does not coincide well with issues that the users find relevant. Moreover, users did rarely mention usability aspects. Comments about controls were predominantly negative, which indicates that players mention game controls when they are poorly implemented. If the controls are good, the players don't seem to mention them much at all.

5. CONCLUSION

In our studies, we can see that none of the games involved are completely without issues. There seems to be a strong connection between the types of issues and the overall estimated quality of the game. Severe game usability issues were prevalent in the low quality games, but hardly occurred at all in the high quality games, and when they did occur, the severity rating was low.

The most frequently violated issue concerned the player's sense of control. Both low and high quality games contained several examples of this issue.

In study 2, it was clear that the users felt strongly about the graphic and sound, as well as the story line. No such issues were found in study 1, which implies that the expert evaluations and the user opinion differs.

Furthermore comments on three tiers of games (high, medium and low) showed that the users also are concerned about game performance, such as video graphic routines, processor speed, and game optimization vis-à-vis computer hardware. Performance and code optimization is not reflected in any of the heuristic lists we used for our studies, which implies that there is room for other sets of more technically oriented heuristics.

6. DISCUSSION AND FUTURE WORK

Analysis of the user comments from metacritic.com shows that users do not care to comment on the same type of issues the heuristic evaluation from Study 1 [18] show. That is not to say that the users are uninterested in good game usability, they just focus on the aspects of the game that they find more important – the game experience.

[1] All ratings were made on metacritic.com, a website that collects professional game reviews from over the world. The ratings are on a 1 to 100 scale, where a score of 100 is best.

[2] AntConc is a freeware corpus analysis toolkit for concordancing and text analysis. It was created by Professor Laurence Anthony and is available at http://www.laurenceanthony.net/

Even though we worked from our own definition of video game issues (when design elements or game challenges have poor usability and/or does not support and/or actively breaks the player's sense of immersion) few issues regarding immersion were found in Study 1 [18]. Notwithstanding this fact, our studies show that users have many opinions regarding immersion, however expressed in their own words. In Study 2 [19] we saw that players are looking for a challenging, immersive and intriguing game experience and engaging narrative. This discrepancy is s strong indication that heuristics are limited in finding and evaluating immersion enhancing design elements.

Many of the heuristic lists that we have been working with in this study do, to some extent, mention attention to storyline and narrative. However, heuristic evaluations are supposed to be automated and quick, executable within a couple of hours per evaluator. A heuristic evaluator is looking high and low for issues, with limited time to evaluate whether a game story is intriguing or not. The general nature of heuristics and their focus on usability issues entail that a heuristic evaluation is not an efficient method to evaluate the narrative aspects of a video game. Alternative methods should be developed for this purpose.

We suggest a closer examination of the applicability of narrative analysis methods applied in other experience based media, e.g. literature, theater, and film. The interactivity of the video game narrative suggests that the bottom up, end-user perspective is the essential starting point. This, we believe, is a clear indicator that heuristic evaluations are limited in their use.

6.1 The Value of Heuristics

While clearly immersion breaking issues can be found using heuristic evaluations, immersion-enhancing elements cannot. The heuristic lists involved in this paper exist in a grey-zone of what can and what cannot be evaluated. To a large extent, the lists work well as checklists for inexperienced developers, but regarding the aspects of evaluating immersion, they are strongly limited. If the heuristics we have examined are not enough, other heuristics are needed, or alternative methods must be developed.

In the process of product development, the aim of heuristic evaluations is to, within a reasonable time frame, provide expert supervision and support to critical design decisions. In video game design, heuristics have a clear usability focus and serve a strong purpose, in particular when applied in the early iterations when the groundwork for the interface is set. Great interface design, however, will not make players buy, play, and enjoy a video game. While it may be true that players have a low tolerance for ill-attended usability design, they are not likely to reflect too deeply on an interface that does not disrupt immersion or playability.

The heuristic lists created from a top-down approach are likely to assist the designers in preventing usability issues. In Study 2 [19] we found that games that receive poor reviews and low ratings from end users are games that Study 1 [18] showed that the user experience is frequently compromised by usability issues. In all likelihood most high budget productions involve some kind of usability references in the design process. This assumption is corroborated by our results in [18]: High scoring games are often high end productions, and do not manifest usability issues to the same severe degree as low budget productions do.

At any budget, standard methods and solutions make more sense than reinventing the wheel. In any game production, a further developed version of our Net Heuristic List could provide a quick and cost effective "shortcut" to pinpointing usability issues that might interfere with the player experience and ultimately the market performance of the product.

6.2 Next Step

So where do we go from here? A natural next step is to create a new heuristic list, working both bottom up to get the users' perspective, and top-town, to get the developers' point of view. Further work could include ways to evaluate the different types of immersion.

Users have high demands on the immersive qualities of video games. Other disciplines e.g. such as literature and film analysis may have evaluation methods that are more suitable than heuristic evaluations; movie companies have pre-screenings, showing alternative endings to test audiences; drama plays on theaters uses the same scrip but the various interpretations can alter the theater experience altogether. It is not a farfetched thought to believe that a closer look at these methods can be useful to the video game market.

7. REFERENCES

[1] K. Isbister and N. Shaffer, Game Usability: Advancing the Player Experience, CRC Press, 2008.

[2] M. A. Federoff, "Heuristics and usability guidelines for the creation and evaluation of fun in video games," Doctoral dissertation, Indiana University, 2002.

[3] H. Desurvire, M. Caplan and J. A. Toth, "Using heuristics to evaluate the playability of games," in *CHI'04 extended abstracts on Human factors in computing system*, ACM, 2004, pp. 1509-1512.

[4] H. Desurvire and C. Wiberg, "Game usability heuristics (play) for evaluating and designing better games: The next iteration," in *Online Communities and Social Computing*, Springer Berlin Heidelberg, 2009, pp. 557-566.

[5] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 1990, pp. 249-256.

[6] J. Nielsen, "Nielsen Norman Group," 1 January 1995. [Online]. Available: http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/. [Accessed 28 02 2014].

[7] P. Jaferian, K. Hawkey, A. Sotirakopoulos, M. Velez-Rojas and B. Konstantin, "Heuristics for Evaluating IT Security Management Tools," *Human–Computer Interaction*, pp. 311-350, 2014.

[8] S. Laitinen, "Do usability expert evaluation and test provide novel and useful data for game development," *Journal of usability studies*, vol. 2, no. 1, pp. 64-75, 2006.

[9] D. Pinelle, N. Wong and T. Stach, "Heuristic Evaluation for Games: Usability Principles for Video Game Design," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2008)*, 2008.

[10] M. Csikszentmihalyi, Beyond boredom and anxiety: The experience of, San Francisco: Jossey Bass, 1975.

[11] C. Jennet, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs and A. Walton, "Measuring and defining the experience of immersion in games," *International journal of human-*

computer studies, vol. 66, no. 9, pp. 641-661, 2008.

[12] J. H. Murray, Hamlet on the Holodeck: The Future of Narrative in Cyberspace, Simon and Schuster, 1997.

[13] A. Rollings and E. Adams, Andrew Rollings and Ernest Adams on game design, New Riders Publications, 2003.

[14] B. Suits, The grasshopper: Games, life and utopia, Broadview Press, 2014.

[15] J. Juul, A Casual Revolution: reinventing video games and their players, MIT Press, 2010.

[16] J. Linderoth, "Beyond the digital divide: An ecological approach on gameplay," in *Proceedings of DiGRA*, 2011.

[17] J. J. Gibson, "The theory of affordances," *Hilldale, USA*, 1977.

[18] B. Stråat, M. Johansson and F. Rutz, "Does Game Quality Reflect Heuristic Evaluation? – Heuristic Evaluation of Games in Different Quality Strata," *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 2015.

[19] B. Stråat and H. Verhagen, "VOX POPULI - A Case Study of User Comments on Contemporary Video Games in Relation to Video Game Heuristics?," United Kingdoms, 2014.