DESIGNING ENJOYABLE VIDEOGAMES:
DO HEURISTICS DIFFERENTIATE BAD FROM GOOD?

Timothy V. Ballew and Keith S. Jones
Texas Tech University
Lubbock, Texas

How does one design a fun video game? One way to guide design is through the use of heuristics. Federoff (2002) compiled a list of design heuristics based on the gaming literature and interviews with game designers. However, these heuristics were not validated. The present study attempted to validate these heuristics by having participants play either a lowly or highly rated PC game. Participants were asked to provide a rating of overall enjoyment and then rate the applicability of each heuristic. Analysis of 133 participants' ratings of overall enjoyment indicated that non-gamers felt both games were equally enjoyable while moderate and heavy gamers found the highly rated game most enjoyable. Further analyses indicated that participants considered approximately 75% of the heuristics to be more applicable to the highly rated game than the lowly rated one. The implications for the design of new video game heuristics are discussed.

INTRODUCTION

Video game industry sales for 2005 were $10.5 billion, surpassing those of the film industry (NPD Group, 2005). A recent top seller, Microsoft's Halo 2, has sold over 6.4 million copies (Reuters, 2005). So what makes this game so popular?

To answer this question, it is important to look at the gaming media. The Web site Metacritic.com combines reviews from multiple media outlets and presents them as an averaged rating on a scale from 1 to 100. The Metacritic rating for Halo 2 is 95 (Metacritic.com). This rating suggests that the game, as reported by multiple reviewers, is enjoyable.

We can see that both the volume of sales and the media ratings for Halo 2 suggest that, overall, this is a fun game. However, neither of these pieces of information tells us which design elements contribute to making the game fun.

When faced with design decisions, designers often rely on heuristics (i.e. guidelines or rules-of-thumb), which can "guide a design decision or be used to critique a decision that has already been made" (Dix et al., 2004). Existing heuristics, such as those for productivity applications, are not useful because games and productivity applications have different goals (e.g. fun vs. efficiency). Thus, a set of heuristics specifically for game design is needed.

To date, Federoff (2002) has completed the most comprehensive effort, compiling a list of forty two video game design heuristics based on the gaming literature and interviews with game designers (Examples can be found in Table 1). These heuristics relate to "game interface" (e.g. players should be able to save the game in different states), "game play" (e.g. one should design games with high replayability), and "game mechanics" (e.g. feedback should be given immediately to display user control; Federoff, 2002).

While Federoff's study is beneficial, it does have limitations. Federoff's data came from video game designers rather than video game consumers. Developers could be considered experts, and thus their view of the video games and their design might not truly reflect that of the video game consumer (Schneiderman, 1998). In addition, Federoff did not validate the heuristics that she compiled. The current study addresses these two issues.

In order to begin to validate these heuristics, participants were asked to play either a lowly or highly rated game. If the heuristics are valid, then participants should rate them as more applicable to the highly rated game and less applicable to the lowly rated game.

The genre chosen to validate Federoff's (2002) heuristics was First-Person Shooter (FPS) due to its reputation as being one of the easiest to learn and play. Two games, both military themed, were selected based solely on their overall ratings on Metacritic.com. After participants played the game, they completed a set of questionnaires assessing their overall enjoyment and the applicability of Federoff's heuristics.

METHOD

Design

A fully between-subjects design was employed. The independent variables were Game Quality and Gamer Type. Game Quality referred to the games used, Navy Seals – Weapons of Mass Destruction (18/100 on Metacritic.com) and Call of Duty (91/100 on

Metacritic.com). Gamer Type referred to how often participants actively played video games. These three groups were non-gamers (0 hours per week), moderate gamers (1 – 6 hours per week) and heavy gamers (7 or more hours per week).

The dependent variables were ratings of overall enjoyment and Federoff's gaming heuristics. Overall enjoyment was assessed on a Likert scale ranging from one (horrible) to five (excellent). The gaming heuristics were taken from Federoff (2002) and assessed on a five point Likert scale ranging from one (strongly disagree) to five (strongly agree).

Participants

One hundred and fifty four students (48 women, 106 men) participated to fulfill course credit. They were between 17 and 27 years old ($M = 19.69$). Fifty nine participants (38.3%) were non-gamers, sixty-five participants (42.2%) were moderate gamers, and thirty participants (19.5%) were heavy gamers. The unequal group sizes are reflective of the sample population.

Information was collected regarding what games participants had recently enjoyed. It should be noted, however, that no specific questions were asked regarding whether or not they had previously played either of the games used in this study. Therefore, no speculation may be made as to what effect previous exposure might have on the collected ratings.

Apparatus

The games were presented on IBM-compatible computers with Pentium 4 processors. All ran Windows XP and had 17-inch monitors. Each computer had identical hardware and software (e.g. ATI Radeon 9600 video cards).

Procedure

Participants completed a demographic survey that inquired as to their current gaming habits. Based on the information from this survey, participants were classified into one of the three levels of Gamer Type.

Participants in each group were then randomly assigned to play one of the two video games. Seventy four participants (28 non-gamers, 39 moderate gamers and 13 heavy gamers) played Navy Seals while eighty participants (31 non-gamers, 26 moderate gamers and 17 heavy gamers) played Call of Duty.

To begin the gaming portion of the experiment, participants browsed through the menu system of their assigned game in order to obtain a better understanding of that game's controls. After becoming familiar with the menus and adjusting the control scheme to their liking, participants played through as many game levels as they could in 40 minutes. To finish levels in either game, participants had to complete all of the objectives for that level (e.g. eliminating all opposing forces or destroying enemy equipment).

After the gaming section of the experiment, participants answered a series of questions. It should be noted that, although participants were informed before playing that game that they would be completing this questionnaire, they were unaware of its contents or purpose.

The first question asked participants to rate their overall enjoyment of the game. The remaining questions asked about each of the forty two gaming heuristics included in Federoff (2002).

RESULTS

Analysis Approach

Data were analyzed separately for each Gamer Type. For each group, two separate analyses were conducted. First, independent t-tests were conducted on overall enjoyment ratings to determine if the ratings for the two games differed.

In addition, the heuristics were analyzed using a set of forty two independent t-tests. Results are presented in two ways. One set of results is presented using a modified Bonferroni correction (Keppel, 1991) to control for family-wise error. The second set of results is presented without the Bonferroni correction in order to lower the likelihood of making a Type II error. The choice of presenting results in both formats seems most prudent given the preliminary nature of this work.

Did Enjoyment Ratings Differ for the Two Games?

Non-Gamers. It was important to determine whether participants were able to distinguish between the two games in a manner consistent with the Metacritic.com ratings. Analysis revealed that the non gamer group enjoyed the two games equally, $t(56.256) = 1.185$, $p>.05$, $M=3.00$ (Navy Seals) and $M=3.35$ (Call of Duty). This finding was expected, as the reviewers on Metacritic.com were most likely avid game players.

Moderate gamers. An independent t-test showed a significant Game Quality effect on overall enjoyment for moderate gamers, $t(61.389) = 4.772$, $p<.001$). This finding suggests that, unlike non-gamers, moderate gamers enjoyed the highly rated game, Call of Duty

(*M*=4.12), significantly more than the lowly rated game, Navy Seals (*M*=3.08).

Heavy gamers. Like moderate gamers, heavy gamers enjoyed the highly rated game more than the lowly rated one, $t(22.067) = 6.522$, $p<.001$, Call of Duty (*M*=4.12), Navy Seals (*M*=2.15). It is interesting to note that the disparity between heavy gamers overall enjoyment ratings seems to be a bit larger than that of the moderate gamers. This difference is driven by the fact that, although ratings for highly rated game were the same for these two groups, heavy gamers enjoyed the lowly rated game less than moderate gamers.

Did the Heuristics Apply to the Two Games?

Non-Gamers. Despite the fact that non-gamer ratings for overall enjoyment were not different for the two games, it is still possible that their heuristic ratings would be different for the highly and lowly rated games. Thus it remained important to analyze the heuristic ratings for this gamer type.

Results of the analysis show that, when using the Bonferroni correction, non-gamers considered only two heuristics (5%) as more applicable to the highly rated game as opposed to the lowly rated one. Without the correction, fifteen (36%) heuristics were considered more applicable to the highly rated game. A list of

heuristics with significant ratings differences, categorized by type, can be found in Tables 1, 2 and 3. Values indicate the probability associated with each t-test ($α$ = .05) comparing the highly and lowly rated games.

Moderate gamers. Moderate gamers provided the largest set of applicable heuristics. Overall, eighteen (43%) heuristics were considered more applicable to the highly rated game as opposed to the lowly rated one when using the Bonferroni correction. When significance values were viewed without the correction, thirty (71%) heuristics were considered more applicable to the highly rated game. Tables 1, 2 and 3 contain a list of heuristics with significant ratings differences.

Heavy gamers. While heavy gamers had the largest difference in their mean overall enjoyment ratings, they only considered eleven (26%) heuristics as more applicable to the highly rated game when the Bonferroni correction was implemented. Twenty six (62%) heuristic ratings were considered more applicable without the correction. Heuristics with significant ratings differences can be found in Tables 1, 2 and 3.

It should be noted that heavy gamers, in contrast to moderate gamers, considered fewer heuristics to be more applicable to the highly as opposed to lowly rated game. It is suspected that the sample size for each group is responsible for the difference.

Game Interface Heuristics	Non	Moderate	Heavy
Warning messages aided in the prevention of and recovery from errors	.019**	.022**	.029**
Objects in the game reflected their real world counterparts	.033**	.000*	.002**
Control options were plentiful but not overwhelming	.004**	.013**	___
The sound in the game provided meaningful feedback	___	.004**	.002*
The interface of the game felt consistent (i.e. colors, text)	.020**	___	.025**
Control customization and default setup were consistent with my expectations	___	.000*	___
Controls were mapped in a way that felt natural	___	.000*	___
The in-game interface distracted me from the gameplay	___	___	.003**
The interface of the game felt non-intrusive	___	___	___
Identifying my status in the game was not difficult	___	___	___
The learning curve for this game was short	___	___	___
The menu system in the game contained too many levels	___	___	___
Reading the game manual would have made gameplay easier	___	___	___
I was able to save the game when I felt it necessary	___	___	___

Table 1: Game Play Heuristics Considered More Applicable For The Highly Rated Game. (* indicates significance with Bonferroni correction, ** without)

Game Mechanics Heuristics	Non	Moderate	Heavy
The physics in the game felt correct	___	.001*	.000*
Feedback about user control was given in a timely manner	___	.000*	.018**
The game got me involved quickly and easily	___	.000*	.001*

Table 2: Game Mechanics Heuristics Considered More Applicable For The Highly Rated Game. (* indicates significance with Bonferroni correction, ** without)

Game Play Heuristics	Non	Moderate	Heavy
The storyline was interesting and enjoyable	.044**	.000*	.000*
The game gave hints, but not too many	.004**	.000*	.000*
The goal of the game was clear from the beginning	___	.000*	.006**
I felt there were multiple goals in the game level(s) I played	___	.005**	.016**
The game was easy to learn but hard to master	___	.005**	.009**
I felt rewarded by the game	___	.000*	.000*
The pace of the game applied pressure without being frustrating	___	.000*	.004**
I felt that the game would be enjoyable if it were played again	___	.000*	.000*
Many different strategies could be used to win this game	___	.040**	.000*
The audio and visual representations in the game aroused my interest	___	.000*	.000*
I was able to interact with many objects in the game	___	.000*	.012**
Early in the game I learned skills that were useful later in the game	___	.005**	.047**
There were multiple paths that could be taken to complete the game	___	.010**	.008**
I felt that after playing this game my skills had increased	___	.015**	.007**
I felt as though the game world would continue to exist even if I were not currently playing	___	.039**	.003**
The puzzles in the game related to the story	___	.000*	.003**
The enemies and other characters in the game acted reasonably, yet in an unpredictable manner	___	.000*	___
The difficulty level of the game could be varied	___	.049**	___
I felt that I could win the game, given enough time	___	.002*	___
The tutorial was interesting and absorbing	___	.002**	___
The gameplay felt fair	___	___	.000*
The gameplay was balanced in a way that there was no definite way to win	___	___	___
I had the ability to create and implement my own content in the game	___	___	___
The game felt modeless	___	___	___

Table 3: Game Play Heuristics Considered More Applicable For The Highly Rated Game. (* indicates significance with Bonferroni correction, ** without)

DISCUSSION

Federoff's (2002) study was limited in that it used game designers rather than consumers to construct game design heuristics. The current study provides a more suitable sample with which to validate the heuristics by using participants who were both gamers (i.e. game consumers) and non-gamers and who varied in the amount of time they spent playing video games.

A second limitation of Federoff's (2002) study was that no attempt was made to validate the heuristics. The current study addressed this limitation by assessing the validity of each of Federoff's forty two heuristics in one genre of video games.

Non-gamers enjoyed a lowly and highly rated video game equally well. Yet non-gamers still considered fifteen heuristics more applicable to the highly rated game without the Bonferroni correction. Non-gamers might not rate the games differently in terms of overall enjoyment, but when the design elements were broken down, it seems that non-gamers knew, to an extent, what made a better or worse game. Further research should be conducted to determine in more detail what design aspects non-gamers consider when making decisions about overall enjoyment. Knowing this might allow game designers to better accommodate both gamers and non-gamers.

Separately, moderate and heavy gamers had a large number of heuristic ratings that were considered more applicable to the highly rated game. Considering moderate and heavy gamers together, thirty four (81%) of the heuristic ratings (without the Bonferroni correction) were considered to be more applicable to the highly rated game. It should be noted that, with the Bonferroni correction, only twenty four (57%) of the heuristics were considered applicable.

The findings of this study suggest that not all of Federoff's (2002) heuristics are valid for designing a video game in the first-person shooter genre. Further validation of these heuristics is needed in other genres to determine which heuristics function cross genre and which are useful only in specific genres. A follow-up study is being conducted to validate these heuristics in a second genre of video games. It is hoped that with further research, a fully validated set of video game design heuristics will be created.

REFERENCES

Dix, A., Finlay, J., Abowd, G, D., & Beale, R. (2004). *Human-Computer Interaction* (3rd ed.). Europe: Prentice Hall.

Federoff, M. (2002). *Heuristics and Usability Guidelines for the Creation and Evaluation of Fun in Video Games.* Unpublished master's thesis, Indiana University, Bloomington, Indiana.

Keppel, G. (1991). *Design and Analysis: A Researcher's Handbook*. New Jersey: Prentice Hall.

Metacritic.com Video Game Ratings (n.d.). Retrieved October 16, 2005 from http://www.metacritic.com/games/

NPD Group. (2006, January 16). The NPD Group Reports Annual 2005 U.S. Video Game Industry Retail Sale. Retrieved February 15, 2006, from http://www.npd.com/dynamic/releases/press_060117.html

Reuters (2005, January). Halo 2 Video Game Sales Hit 6.4m units. *Reuters*. Retrieved October 25, 2005, from http://www.msnbc.msn.com/id/6848748/

Schneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human Computer Interaction* (3rd ed.). Massachusetts: Addison-Wesley.