Games User Research Methods to Book-End the Game Development Lifecycle

Marina Kobayashi
Electronic Arts
209 Redwood Shores Parkway
Redwood City, CA 94065 USA
mkobayashi@ea.com

Rich Ridlen
Electronic Arts
209 Redwood Shores Parkway
Redwood City, CA 94065 USA
rridlen@ea.com

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract

The beginning and ending of a videogame's lifecycle are crucial times in its development. Game teams generally have questions early in production about how consumers will interact with their new mechanics as they build them and then as they approach Alpha, how they will experience the totality of the final game. This talk will introduce two methodologies that are utilized at EA Game Lab to help game teams answer these questions: white box testing and full play through testing, respectively.

Author Keywords

Games user research, methods, game design, playtest, playthrough, usability, video games, evaluation, metrics, low-fidelity prototype testing, white box

ACM Classification Keywords

H.5.1. Information interfaces and presentation (e.g., HCI): Evaluation/methodology.

General Terms

Design, Measurement

Introduction

This talk will be broken into two parts. The first will focus on white box testing that occurs early in production and the second will focus on full play through testing that occurs late in production. The two different methodologies associated with each will be explained and examples of them in practice will be shared.

White Box Testing Overview

Many key game mechanics do not see the light of user testing until well into the development of a game. This can present significant complications as fixing some issues later in the cycle can have a weighty cost. Early in production the first efforts are to create mechanics that work within the game's concept. The game designers use this time to experiment with design solutions, interactions, and/or approaches to accomplishing objectives within a set of constraints. Testing them within this "white box" space prior to the completion of its in-game context is a good method to gather early feedback. Moreover, issues can be fixed relatively cheaply as the mechanic is generally so early that few other resources (aesthetic or otherwise) have yet been spent on it.

Methodologies Used for White Box Testing

Think-aloud protocol is commonly thought of as a method testing software content, navigation, and interaction designs (such as a "call to action" or "feedback loop") which can be a series of inputs and outputs. The user explains his/her thought process and the researcher/designer is able to identify issues with the tested implementation. For example: unclear content, navigation that fails to make the correct path clear, or lack of feedback from interaction designs.

However, aside from usability testing, think-aloud protocol can be invaluable to getting feedback on game mechanics. When testing mechanics, there may be little UI/content or other elements that may be traditionally considered testable for usability. However, mechanic testing could be considered more parallel to testing interaction designs, but with additional visceral considerations. With conventional software, the feedback on an interaction may cover the look and feel and the system response, but may often be 2-dimensional. With game mechanics, the feedback systems can include 3-dimensional aspects of physics, such as weight, force, speed, velocity, and power. A few examples are: effects such as the power of an explosion and the ability to rapidly turn 180°-270° to counter an attack from behind—both of which require conveying to the player the physics.

Beyond mechanic testing, levels or multiplayer maps are also good candidates for white box testing (**figure 1**). Player navigation paths, enemy spawning and pathing, and placement of cover or obstacles are all examples of what can be found out via testing in a white box environment.

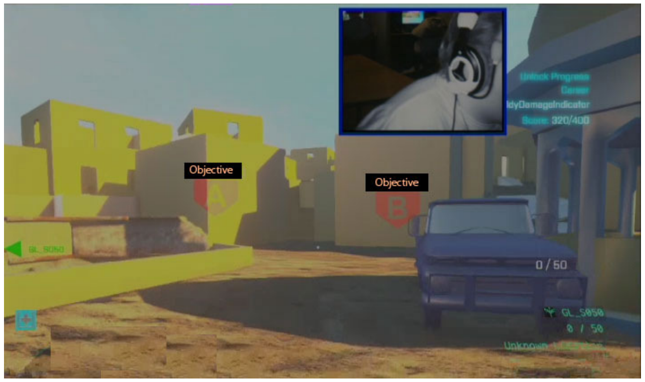


Figure 1. A player scans structures in the background to identify a path to his objectives.

Think-aloud Protocol: White box testing uses a conventional one-on-one think-aloud protocol to understand the players' mental model, successes/failures, and determine breakdowns with mechanics while they are being developed. In contrast with conventional think-aloud usability, the objective of games user research is not to make it easier for a user to accomplish feats, such as defeating a boss enemy or solving a puzzle, but to understand what causes frustration with these activities and how the experience can be improved through better design. The development team may address these issues as they continue to iterate on the designs, but time, effort, and resources have been saved.

Specific Benefits of White Box Testing:

- Catching usability errors early
- Minimum impact on resources
- Gut checking of key mechanics early on
- Avoidance of costly impact on resources if not caught until Alpha
- Collecting feedback that can provide design direction, help identify optimal solutions

Full Play Through Testing Overview

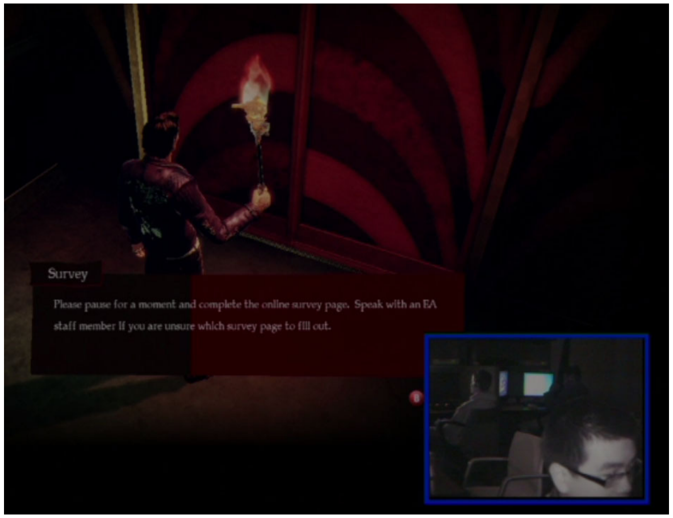
Traditionally many game teams never have consumer data on the entirety of their game experience before its release; they rely mostly on themselves, QA, and/or other internal resources. While some mechanics and levels will be tested over the lifecycle of the game, not understanding its totality can lead to disharmony between levels, challenge, advanced mechanic tutorialization, and story comprehension, among others.

Methodologies User for Full Play Through Testing

10 – 20 participants are hired through an external agency[1] to come to EA's lab facility for 2 to 5 days, for 4 to 8 hours per day. A screener is provided to the agency to use, but they are also phone screened again by internal staff. Upon arrival the expectations and general method over the test's duration is explained. Data is gathered during the test in a variety of ways: survey stops, telemetry, spontaneous core sampling, cognitive probing, and/or hand-written metrics.

Survey Stops: These can be integrated into the game itself as pop-ups that appear similar to tutorial dialogues or other on-screen messaging (**figure 2**). The pop-ups inform players to stop gameplay and to turn to the PC survey. Their technical implementation is done by the development team, with the GUR informing the frequency required to capture the desired data granularity. Due to time and resource constraints, some development teams cannot implement them. In this case, players will complete surveys at the end of levels, quests, and/or at regular timed intervals.

[1] Per California law, EA is not allowed to have participants "work" for more than 3.5 hours without compensating them a wage; thus the use of an external agency.



Survey

Please pause for a moment and complete the online survey page. Speak with an EA staff member if you are unsure which survey page to fill out.

Figure 2. Automated in-game survey prompts appear on-screen and pause the game.

Telemetry: Contingent on the development team, telemetry hooks are implemented into the game code to collect specific metrics. These are identified by the game team with the assistance of the GUR. Generally, they are metrics such as player deaths, weapon usage, time to complete levels or puzzles, etc.

Spontaneous Core Sampling: During the course of a multi-day test, issues can surface that weren't expected or new questions about weapon usage or interest in a particular feature can suddenly gain importance to the game team/GUR. Either quickly programmed in Survey Monkey or answered with paper and pen, these can be beneficial in understanding how players are processing the experience.

Cognitive Probing: While various game levels might previously have been usability tested before coming together into the whole of the game, the fixes from that test might not yet have been seen in practice. Additionally, some levels, mechanics, or boss fights might not have been seen by the GUR at all. In this case, following one player's experience and then probing him/her to talk through his/her thinking can be very valuable.

Hand-Written Metrics: Telemetry might not have been implemented into the game, but metrics are still needed to ascertain aspects of the player experience. These can be timings of a level or timing the game as whole; it can be deaths during boss fights or how long a puzzle takes. They can serve to help the team tune gameplay as well as help the GUR convince the team of necessary changes to gameplay with actual numbers.

Analysis of Full Play Through Data

The collected data from the various sources can serve to reinforce, refine, or repeal one another. This creates a confidence with the team and the GUR that the decisions being made are as validated as they possibly can be.

Specific Benefits of Full Play Through Testing to Better Understand Players and Uncover Game Design Issues:

- Challenge tuning over the course of the entire game
- Overall story and narrative comprehension
- Advanced mechanic tutorialization effectiveness
- Ending satisfaction
- Campaign and multiplayer synchronization

- Uncovering of puzzle issues
- Uncovering of navigation issues
- Understanding the average length of levels and the campaign via the consumer experience (not QA/development teams)
- Vetting of menu and UI screens as players progress
- Engagement of players over the course of the entire game

Acknowledgements

We thank all the members of the EA Game Lab for their support: Joseph Iloreta, Tim Toy, Jami Novotarski, Amber Nalu, Damian Rainey, and Cory Garza. Authors Marina Kobayashi and Rich Ridlen gratefully acknowledge Electronic Arts for supporting and growing the Game Lab and its practice of Games User Research. They would also like to acknowledge their management: Scott Taylor and Jim Lewis.

References

[1] Wikipedia: Design. http://en.wikipedia.org/wiki/Design

[2] Davis, J; Steury, K; and Pagulayan, R. A survey method for assessing perceptions of a game: The consumer playtest in game design. *The International Journal of Computer Game Research,* (2005). http://www.gamestudies.org/0501/davis_steury_pagulayan/

[3] Fulton, B. Beyond Psychological Theory: Getting Data that Improves Games, *Gamasutra* (2002). http://www.gamasutra.com/gdc2002/features/fulton/fulton_01.htm