Lecture 24

Player Testing
Why Player Testing?

A good game no one can learn to play...

...is a bad game
Material for Today

- Not an area of expertise
  - Material from Dan Cosley (danco@cs.cornell.edu)
  - With additional material from Hrönn Brynjarsdóttir (hb47@cs.cornell.edu)
- Dan Cosley is great resource
  - Has guest lectured in past
  - INFO 3450 is a big part of the game design minor
Today’s Outline

• Questions to start with
• Different metrics of usability
• Conducting a user study
• Data collection/analysis
Questions to Start With

• Why are you conducting the test?

• What are you going to learn?
  ● Sources of player difficulty?
  ● Typical player strategies?

• How will you use the results?
  ● Sometimes, to persuade or justify
  ● Often, to iterate the design
Brainstorming Exercise

• What matters in a game?

• And how do you measure it?
Quantitative Metrics

• Time to learn to use a game verb
• Time (ability) to complete a specific task/quest
• Usage (or lack of usage) of gameplay features
• Errors (how many, where)
• Player satisfaction (Likert scale)

Problem: need many users for good stats
Qualitative Metrics

- What does the user say?
- Where/how do they run into trouble?
- What’s the first reaction/impression?
- How would they describe the gameplay?
- Would they play it again? Recommend it?

**Advantage**: More amenable to small groups
The Key Idea

- Put the challenge **where you want it**
  - Some things are meant to be difficult
  - If not explicitly a challenge, should be easy

- **Example**: Fast Travel
  - Allow the user to explore a vast world
  - But keeps constant travel from being boring

- Even bigger problem in educational games
  - Are they encountering the “right” challenges?
The User Study

- **Participants**
  - Who are you studying?

- **Artifact(s)**
  - What are you studying?

- **Tasks and scripts**
  - How do you plan to study?

- **Experimenter roles & best practices**
  - How are *you* involved in the study?
Participants: How Many?

- People’s time is valuable
  - Theirs: how often do they want to play?
  - Yours: you have to administer the test

- Tests yield **diminishing returns**
  - Especially at prototyping stage
  - Rule of 5 (or 3) for qualitative results
Participants: Who?

- Not you (usually)
  - You have privileged knowledge
  - But okay in early stages
- Representative of target audience
  - *Actimates Barney* vs. *Grand Theft Auto*
  - Back up your concept document claims
- Someone(s) old, someone(s) new
Ethics and Benevolence

- Remember...
  - Your participants are real people
  - They are doing you a favor
Hi, XXX.

My name is XXXX, and I’m going to be walking you through this session.

You probably already know, but let me explain why we’ve asked you to come here today: We’re testing a web site that we’re working on to see what it’s like for actual people to use it.

I want to make it clear right away that we’re testing the site, not you. You can’t do anything wrong here. In fact, this is probably the one place today where you don’t have to worry about making mistakes.

As we go along, I’m going to ask you to think out loud, to tell me what’s going through your mind. I might stop you from time to time and ask you specific questions regarding your activities. This will help us get a fuller picture of what this interface is like to work with.
Artifacts: What is Tested?

- Working with incomplete product/prototype
  - Some stuff is clearly not finished
  - Do not want comments on unfinished bits
  - Be very clear of the scope of your test

- The Wizard of Oz
  - Can hide unfinished details with hacks or tricks
  - Great for AI in games (replace AI with human)

- Test should focus on finished details
Tasks & Scripts: Direction

- “Climb up to the blue square”
- “The game has put information in your codex. You may want to read it before continuing”
- “What do you think should happen if you go here, touch this, hit that?”
  - Example of pre-interactive direction
- “Just try things out, explore”
  - Sometimes no direction is direction
Tasks & Scripts: Design

- Directedness of tasks depends on **goal**
  - Do **you** know what you are looking for?
  - Are some goals more important than others?

- How long should tasks be?
  - When should you mercy-rule them?

- How long is the test/how many tasks?
  - User attention wanes over time
  - Do you want to order by priority?
Some Great Resources

- http://www.usability.gov
  - Standard government usability guidelines

- http://www.irb.cornell.edu
  - Ethical guidelines for usability testing
  - Covers all “human experiments” at Cornell
  - Professors need approval before research
Roles & Best Practices

- At least two testers
  - **Experimenter**: run the show
  - **Observer(s)**: record what happens

- Be unobtrusive as possible
  - Will you be there when they play?
  - Your input will bias participants
  - But do not frustrate the user
Think Aloud Method

- While you shut up, they should talk
  - About **everything** going on in their head
  - Gives you “inside the head” data

- Complements and explains observations
  - Separate player **failures** from **frustration**
  - Think of games like *Super Meat Boy*

- A little unnatural, may need reminders
Do Not Take it Personally

• People will criticize the interface
• It can be sad watching people fail
• But do not coach them
  • This will bias your results
  • If you mercy rule them, move on
The Study: Data Collection

- **Notes**
  - What did they say, what did they do?

- **Videotape & Audio**
  - To capture what you might have missed
  - Audio okay as long as the player thinks aloud

- **Game state logs**
  - Log the state of the game to a file
  - Can replay back as a cinematic
Advanced Data: Heatmaps

Zone of Death!
Filter on:
How do I…
SWTOR Example: Chat Logs

Filter on:
Bug, Broken
SWTOR Example: Player Deaths

Legend:
Orange = group
Green = solo
SWTOR Example: Player Deaths

Legend:
Orange = group
Green = solo

Enemy level - player level
SWOTOR Example: Patrol Paths

Encounter “pull” radius
Low Tech Data Collection

• **Questionnaires**
  - When? Is it a pretest or posttest?
  - Multiple choice vs. open-ended questions

• **Interviews**
  - Again, before or after?
  - Don’t have to be super-formal
Post-Test Team Debrief

• Fresh observations taste better
  • Do not wait too long to debrief
  • But do not debrief with player in room

• Talk about each session post-session
  • Comparing results to previous sessions
  • But ignore sessions that are too old

• Talk about general issues every day
Problem Chart Spreadsheets

- **General Format:**
  - Statement of the problem
  - Observation(s) that prompted it
  - Estimated importance
  - Ease of fixing

- This allows you to prioritize
  - And also define “problem”
  - In games, some things should be hard.
Problem Chart Spreadsheets

- **General Format:**
  - Statement of the problem
  - Observation(s) that prompted it
  - Estimated importance
  - Ease of fixing

- This allows you to prioritize

- And also define “problem”

- In games, some things should be hard.

Not unlike bug tracking!
Summary

- Find representative users
- Have a plan for your test
- Let the player play
- Observe and notice
- Summarize and act

- Rinse and repeat: frequently
A Worthy Goal

• Make a test plan (5-10 minutes)
  • Your artifacts
  • Welcome script
  • Task or two
  • A couple of questions to ask

• Meet with another group
  • Swap members for testing once as a twice
  • Debrief as a group, and with both groups
  • About specific game, about testing overall