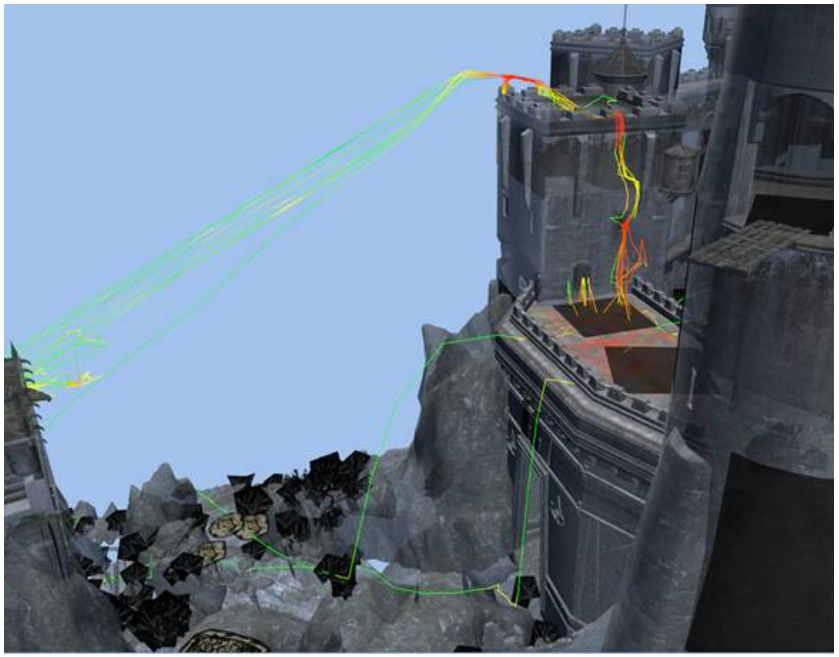


GUR Methods and Game Analytics

INF2300H
iSchool, University of Toronto

Velian Pandeliev
Dec. 2, 2019

HCI/UX Reading Group

Wednesdays, 3:30 pm - 4:30 pm in BL 626

10-minute intro to a paper,
followed by discussion.

hcitang.github.io/hci-reading-group/#!index.md

Ubisoft Visit Post-Mortem

Research methods revisited

Data Axes

		Qualitative	Quantitative
Behavioural	**Objective**	*"Yeah, I **played this game**…*	*… and got to **level 19**"*
	Subjective	*"I was **playing stealthy**…*	*…like, **40%** of the time"*
Attitudinal	**Objective**	*"I **understood** base upgrades…*	*… by hour **2** or so"*
	Subjective	*"I **liked** the game a lot…*	*… and I give it an **8.5 / 10**"*

GUR methods

Player profiles

- Ethnographic field study
- Focus group
- Market segmentation
- **Questionnaire**
- Diary study
- Personas

Usability/Clarity

- A/B testing
- Cardsort
- **Heuristic evaluation**
- Initial player experience
- **Interview**
- Narrative testing
- **Think-aloud test**

Balance/Appreciation

- **Benchmark**
- Critical facet playtest
- Extended playtest
- RITE (Rapid Iterative Testing and Evaluation)
- **Telemetry**
- **Review analysis**

Mixing GUR methods

In practice, GUR studies employ a variety of methods in sequence for any given study.

This allows researchers to gather and use different kinds of data to answering their research questions.

A **usability** study might involve a think-aloud protocol and an interview, coupled with in-game telemetry.

An **extended playtest** might include direct observation, questionnaires, telemetry, and follow-up interviews.

A **benchmark** might include the above plus a review analysis component.

Benchmarking

Benchmarking in GUR allows researchers to **compare games** directly based on player input on the same well-established research instruments.

Benchmarking can provide some clarity to game developers about how their game measures against similar games in its genre.

Benchmarks are often performed on **other studios' games**.

The **same instruments** must be used on all games, which means they have to be quite **generic** in nature.

Even with the same instruments, the context of a game's release or anticipation varies, so benchmark conclusions must be **made carefully**.

Terminator: Resistance

47

December 10, 2019

This story takes place in a post-apocalyptic Los Angeles, nearly 30 years after Judgement Day, and stars a brand new hero in the Terminator universe, Jacob Rivers, a soldier in the…

Expand

Asterix & Obelix XXL 3: The Crystal Menhir

59

November 28, 2019

Expand

Sniper Ghost Warrior Contracts

63

November 22, 2019

Fulfill contracts that offer clear objectives with a fixed monetary reward and options to complete bonus challenges for payouts. With hundreds of ways to take down a wide range of targets,…

Expand

Sid Meier's Civilization VI

88

November 22, 2019

Civilization VI offers new ways to interact with your world, expand your empire across the map, advance your culture, and compete against historys greatest leaders to build a civilizatio…

Expand

Review Analysis

A common and recurring benchmark in the video game industry is a game's aggregate **review score**.

Rigorous procedures exist for using game reviews as a source of post-launch GUR improvement data (e.g., **RVA** by Ian Livingston). Why?

- Reviewers are articulate, **experienced**, context-aware players whose opinions influence other players'
- Reviewers are plugged into the expectations for the game, the genre, and the **social and cultural climate** around it, offering insights inaccessible to regular players
- **Teams already read reviews** and may form inaccurate, non-rigorous opinions based on them

Under no circumstances
should designers and developers
be allowed to form
their own opinions of their games.

/s?

Questionnaire Design

Questionnaire vs Survey

A **survey** is a systematic approach to obtaining data from a sample so insights from it can be extrapolated to an entire population.

A **questionnaire** is a research method that poses (typically written) questions to participants.

Questionnaires are often administered to large samples and used to answer survey questions such as demographics, player profiles, etc.

Questionnaires are also used to obtain attitudinal data about player experiences and preferences.

Pros: comparability, relatively inexpensive to administer to large samples

Cons: inflexible, tedious, can be hard to verify respondent attributes

Types of Questions

Closed-ended questions provide a fixed set of possible responses.

They are suitable for situations when the range of answers is **known and limited** (e.g., *"Please indicate what gaming devices you own"*), or when a short answer has **no established scale** (*"How fun was the game overall?*)

Open-ended questions use a freeform text field to record responses.

They are suitable for **long-form** attitudinal answers, for answers where the options are **unknown or too numerous** (e.g., age), or for **existing natural scales** of response (e.g., duration of play)

Single and Multiple choice

On which platform did you play this game?

- ❍ XBox One
- ❍ PlayStation 4
- ❍ PC or Mac
- ❍ Android or iOS

Which of the following classes have you played before?

- ☐ Barbarian
- ☐ Crusader
- ☐ Demon Hunter
- ☐ Monk
- ☐ Necromancer
- ☐ Witch Doctor
- ☐ Wizard

Ranking

Please rank the following game genres from your favourite (1) to your least favourite (5):

- ☐ Action game
- ☐ Adventure game
- ☐ Strategy game
- ☐ Simulation game
- ☐ Role-playing game

Rating

Rating scales are used to elicit responses about an attribute along a continuum:

1	2	3	4	5

Excellent	Good	Average	Poor	Terrible

Strong accept	Weak accept	Weak reject	Strong reject

These work best with 4-7 rating levels.

Consider whether an even or odd number of levels is more appropriate.

Open-Ended Questions

I identify my gender as:

Number of hours spent playing video games this past week:

Age:

(Consider ranges for age)

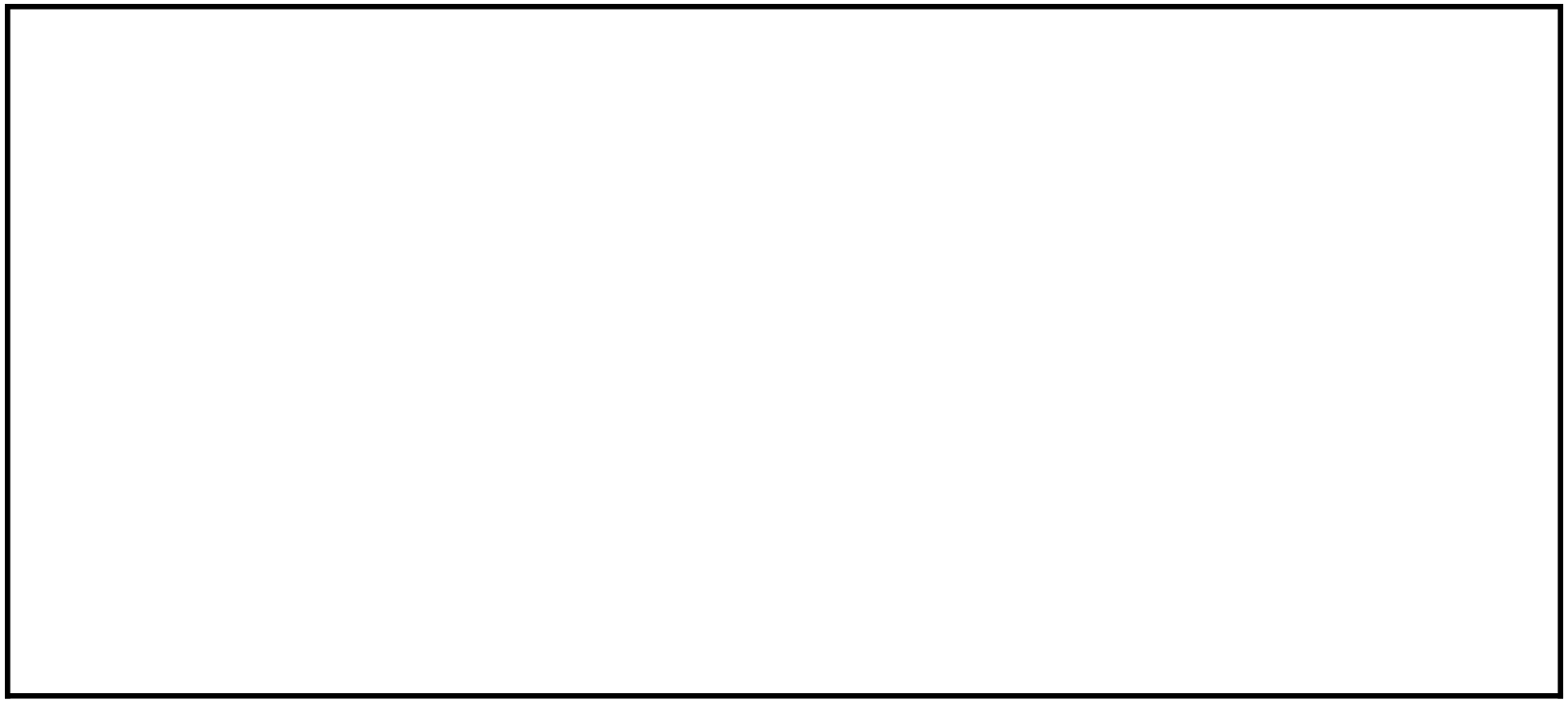
0-17	18-30	31-50	51-70	71+

Writing Questions

- Ask as few questions as you can, as precisely as you can
- Avoid jargon, complicated questions, cultural references
- Mix closed-ended and open-ended questions
- Do not offer neutral or low-effort answer options,
  e.g., "no opinion", "don't know", etc.
- Let participants clarify, comment, and explain in open-ended fields
- Use a progress tracker

Open-Ended Questions

What do you like best about Dark Souls?



Open-Ended Questions

Please recall an outstanding **positive** experience you had in your most recent game session in Dark Souls.

Please try to **describe** this experience as accurately, concretely, and in as much detail as possible. Don't worry about grammar or spelling.

This was my experience:

Established GUR Questionnaires

- **Game Experience Questionnaire** (GEQ) (Ijsselsteijn et al. 2008)
  Immersion, competence, flow, affect, and challenge.
- **Player Experience of Need Satisfaction** (PENS) (Ryan et al. 2006)
  Self-Determination Theory (autonomy, competence, relatedness)
- **Immersive Experience Questionnaire** (IEQ) Jennett et al. (2008)
  Cognitive and emotional involvement, real-world disassociation, challenge & control.
- **Positive and Negative Affect Schedule** (PANAS) Watson et al (1988)
- **Ubisoft Perceived Experience Questionnaire** (UPEQ)
  Azadvar et al. 2018

Reminder: Self-determination theory

Self-determination theory states that certain internal motivations are universal mental **needs** rather than just wants or desires:

- **Competence:** I need to feel good at something
- **Autonomy:** I need freedom to do things my own way
- **Relatedness:** I need to connect to other people

This is highly relevant to video games since we can already think of game aspects and properties that connect to each of these needs.

DECI & RYAN

Example: UPEQ

Initial development study on 563 participants (10% female), 69% US/CAN, 28% France.

Validation study conducted with 7997 participants (3% female) and matched to in-game telemetry

Uses 7-point **Likert scales** (*'LICK-art'*) on a series of declarative statements.

Rate your agreement with the following statements:

I was free to decide how I wanted to play.

1	2	3	4	5	6	7
Not at all			Somewhat			Very true

Evaluate this question:

How many iPhones do you own?

0-1	**1-2**	**2-4**	**4-8**	**8+**

Evaluate this question:

The iPad 360 is a revolutionary device with an amazing set of features.

Indicate how much you like it on a scale of 1 to 5:

1	2	3	4	5

Evaluate this question:

Which do you prefer:
being able to call people while riding the subway, or having to wait?

Evaluate this question:

Do you feel tired while driving because your children are noisy?

Always	Sometimes	Sometimes, but it's not important	No, I don't	No, they are very quiet

Evaluate this question:

The graduate courses I have taken in this department failed to be helpful in my academic career.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Evaluate this question:

What do you think?

☐ **I agree with my MP Lawrence Toet!** We must take additional action to protect Canada from terrorism.

☐ **I disagree!** Terrorists are victims too.

HOW TO FOLD & MAIL:

For your privacy, fold the newsletter with the

Evaluate this question:

After 1 hour of play, how would you rate your experience with the game overall?

1	2	3	4	5
Terrible		**Okay**		**Great**

Evaluate this question:

Overall, what did you LIKE and DISLIKE about the game?

LIKE:

DISLIKE:

Evaluate this question:

How was the DIFFICULTY of the game?

1	2	3	4	5
Easy				Hard

Evaluate this question:

How was the DIFFICULTY of the game?

1	2	3	4	5
Much easier than I expected		**Just right**		**Much harder than I expected**

Interview Design

Interviews

Interviews provide detailed attitudinal data about individual participants, their context, and the challenges they encountered in their gameplay immediately after a gameplay session.

Pros: Flexible and adaptable, rich qualitative data

Cons: Expensive, impossible to directly compare respondents, hard to generalize

Types: structured (**script**), semi-structured (**outline**), unstructured

Data: verbatim responses need to be transcribed, coded, and analyzed

Interview questions

Ask questions that elicit rich but accurate answers.

Avoid **yes/no** questions:

Did you like the game?

Ask **open-ended** questions:

Tell me about a positive experience with the game.

Ask questions about specific **recent** occurrences:

Tell me about a positive experience with the game during your latest session.

"Say more about that."

Interview sequence

Introduction: introduce yourself, explain goals, outline ethics, ask permission to record / take notes.

Warm-up: first questions should be easy and non-threatening

Main body: present questions in a logical order, follow-up on tangents but keep interviewee on topic

Cool-off period: include a few easy questions to defuse tension at the end

Wrap-up/Closure: thank interviewee, signal the end of the interview, stop recording, offer compensation (if applicable)

Warm-up questions

Overall, what did you think about the game?

What did you like about the game?

What did you not like about the game?

Was there anything more difficult than you expected in the game?

Was there anything more confusing than you expected in the game?

Challenges

- Build trust
- Follow the script but allow digression
- Refocus or prompt participants without interrupting
- Formulate questions on the fly
- Avoid repetition
- Keep track of time
- Record the interview and record it another way
- Multitask: talk, listen, filter, take notes
- Avoid bias

Types of prompts

Silence

When your respondent stops talking, remain silent until they say more.

Echo

Repeat the last thing they said and ask “and then what happens”

Agreement sounds

Say “Oh”, or “I see” to prompt the participant, but be careful not to lead or interrupt

“Say more about that”

Yields confirmation / elaboration and helps focus discussion

Interviewing tips

- Repeat rephrased answers back to the participant to check your understanding.
- Deflect participant questions: *"How would you expect it to work?"*
- Avoid leading/loaded questions: *"What made you think that?"* implies the participant was wrong to think it, vs *"How do you know that?"*
- *"I figured it out, but kids might have a hard time."*
  Don't let participants project: *"How was it for you?"*
- Verify responses: *"Did you understand the cover mechanic?"* vs *"Please explain how cover works in the game"*

Cautions

Users remember:

- The **lowest** points, when the game frustrated or failed them
- The **highest** points, when things felt exceptionally good
- The **most recent** points, which are still fresh in their memory

BAUMEISTER 2001

Research Bias

Bias

As researchers, we are always trying to avoid **bias** (skewing data based on how we think it should be).

We need real data that is not affected by our expectations, opinions, or beliefs.

How we phrase and construct our questions is crucial to data integrity.

Confirmation Bias

Avoid asking questions that signal to the participant what you would like the response to be:

"How much did you like the main boss fight?"

Prior judgment or consensus:

"This game has a Metacritic score of 93.
What did you think about it?"

Your own evaluation:

"That was mean!"

Unanswerable question bias

Avoid asking about information that the participant wouldn't know or be able to give you reliably.

Highly specific:

"What time did you have dinner 2 weeks ago?"

Foresight:

"How do you think you'd feel after beating this game?"

Other people's experiences:

"What did your daughter think was the best item in the game?"

Social acceptance bias

Be careful asking about socially desirable or undesirable topics. Participants are likely to skew their responses and you should take them with a grain of salt.

"When was the last time you flossed?"

"Have you ever committed a crime?"

In **central tendency bias**, participants are reluctant to give extreme answers. (e.g., 1/7 or 7/7)

In **acquiescence bias**, participants tend to agree with statements as presented.

Analyzing Qualitative Data

Qualitative studies involve collecting and analyzing texts, observations, video, and artifacts to understand complex situations.

Qualitative goals

Fieldwork produces large amounts of data

Difficult to interpret ambiguous comments & understand complex situations

There may not even be a discoverable “truth” — multiple researchers might (and do) have different perspectives on the same situation

Your task is to turn unstructured, qualitative data into **actionable insights** about the game and convince the design team that you have enough evidence to support them.

Steps of analysis

Data reduction

Select, focus, simplify, and abstract the raw data

Data organization

Organize reduced data so explanations can begin to be generated

Data explanation & insight

Draw conclusions from explanations and present them

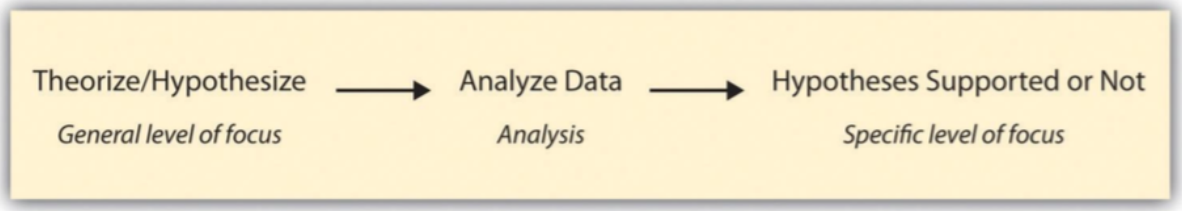
Approaches

Inductive analysis

Let the themes emerge from studying data (bottom-up)

Deductive analysis

Determine if data supports preexisting hypothesis (top-down)



Coding qualitative data

The most common measures of quantitative data: averages, differences, numbers of participants affected are inaccessible or difficult to obtain with qualitative data.

Data coding involves reducing freeform comments to countable and easier to work with codes. E.g.,

"I couldn't figure out where the damage was coming from; suddenly it was like BLAM you're at 65%. BLAM 30% and your mech suit alarm is going off. I had no idea what to do."

COMBAT: Damage source unclear. ***AGENCY: Felt helpless.***

Generating codes

Deciding **how to code** a given piece of data is a significant challenge.

Different researchers may choose different codes for the same data, making comparisons between them difficult.

Even so, codes are very **rarely prescribed** ahead of time.

Instead, GUR largely uses **open coding**: inductively generating codes based on the data.

As more data is analyzed, codes are compared, merged, and grouped into higher-order emergent axes or **themes**.

Researchers **debrief** throughout the process to generate code books, align their coding practices and discuss how codes relate to each other.

Grounded Theoryish

Open coding (Data reduction)

Describe each data point with a code that summarizes its core concept. As you encounter more codes, compare and merge into new concepts.

Axial coding (Data organization)

Combine concepts into categories and group them by theme or issue.

Themes or issues often emerge before all the data is analyzed.

Selective coding (Data explanation)

Pick the core themes/issues and scan through the remaining data specifically looking for instances that support or refute your conclusions so far.

Supporting qualitative insights

Counts and consensus

Your coding indicates that over 70% of players mentioned feeling frustrated with the first mini-boss.

Exemplars

Video clips or verbatim descriptions of an issue that illustrate it clearly and show a specific player's frustration as a direct result.

Triangulation

Find the issue in other sources of data, e.g., questionnaire and interview, questionnaire and observation, observation and telemetry

Game Analytics
Telemetry

Some of our methods

	Quantitative	Qualitative
Objective	**???????**	heuristic
	observation	think-aloud
Subjective	questionnaire	interview

Game Analytics

Game analytics is a professional discipline focused on collecting and analyzing vast amounts of in-gam **telemetry (event data logged from a distance)** to support decisions about designing games, improving player experience, and prioritizing features.

Game Analytics:

- Deals with massive data streams from all players
- Requires technical infrastructure and skills
- Is only ever objective, behavioural, and quantitative.

GUR and GA

Game Analytics is an emerging discipline made possible by:

- Advances in data collection and management technologies
- Games extending to data-rich environments (e.g., Facebook)
- Freemium games optimizing revenue through behaviour analysis

It supplements GUR methods with:

- Precise behavioural recording
- Triangulation and new data streams
- Reach and scale
- Live A/B hypothesis testing

Questions

Game Analytics

What is happening?

What has happened?

What is likely to happen?

What will happen if we change X?

How do we make X happen?

Why did X happen?

How do we optimize X?

Games User Research

Game Analytics process

Game Analytics is relevant **throughout development** of a game, and preparation for its appropriate use can begin in the concept phases of a game, peaking during **beta** and **live testing** post-launch.

Events are simple, trackable units of data that the game is instrumented to collect:

> Enter a game area, die, start a game session, upgrade an item

Measures of these events:

> Frequency, number/proportion of players performing it, frequency by player type, by stage in game, by session, player progress, player experience

Event structure

Common event logging fields:

- Event name
- User ID
- Timestamp
- Session ID
- Build version
- Platform
- Additional parameters

Event structure

Event_Name	StartMission	EndMission
User_ID	P0344	P0344
Timestamp	2019/07/31 19:44:21	2019/07/31 20:15:01
Session_ID	5593096	5593096
Platform	PS4	PS4
Build_Version	1.12.7.4	1.12.7.4
Player_Level	6	6
Parameters	Name:'A1M3_Hilltop'	Name:'A1M3_Hilltop' stars:3

Key Performance Indicators (KPI)

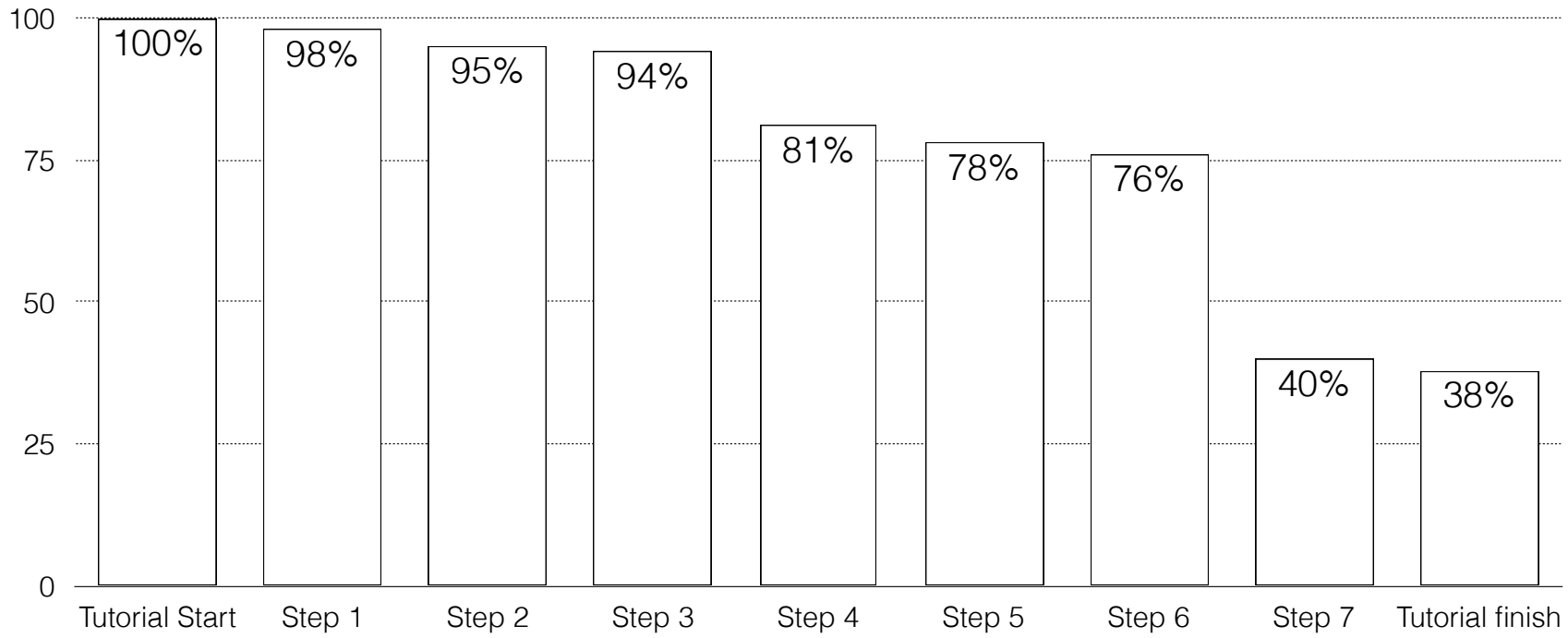
In addition to user experience an game design, GA is responsible for monitoring key business and benchmark metrics after launch.

- Downloads / installs / new users
- Active users (daily, weekly, monthly)
- Ad / download revenue
- Retention Day (1, 3, 7, 14, 30)
- Average Revenue Per User / Per Paying User
- Percentage of Paying Users

Games do not automatically log all player actions.

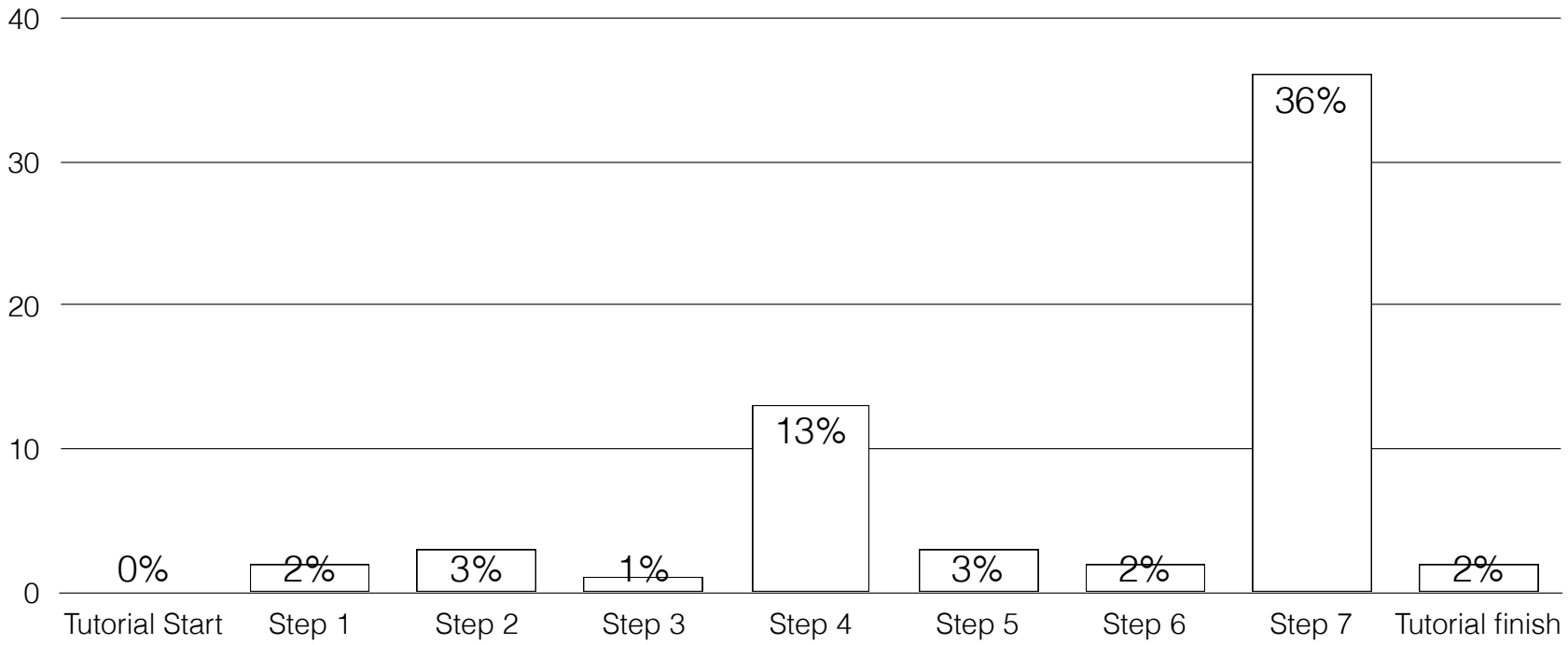
GA Example 1: Tutorial funnel

Percentage of players continuing tutorial after each step:

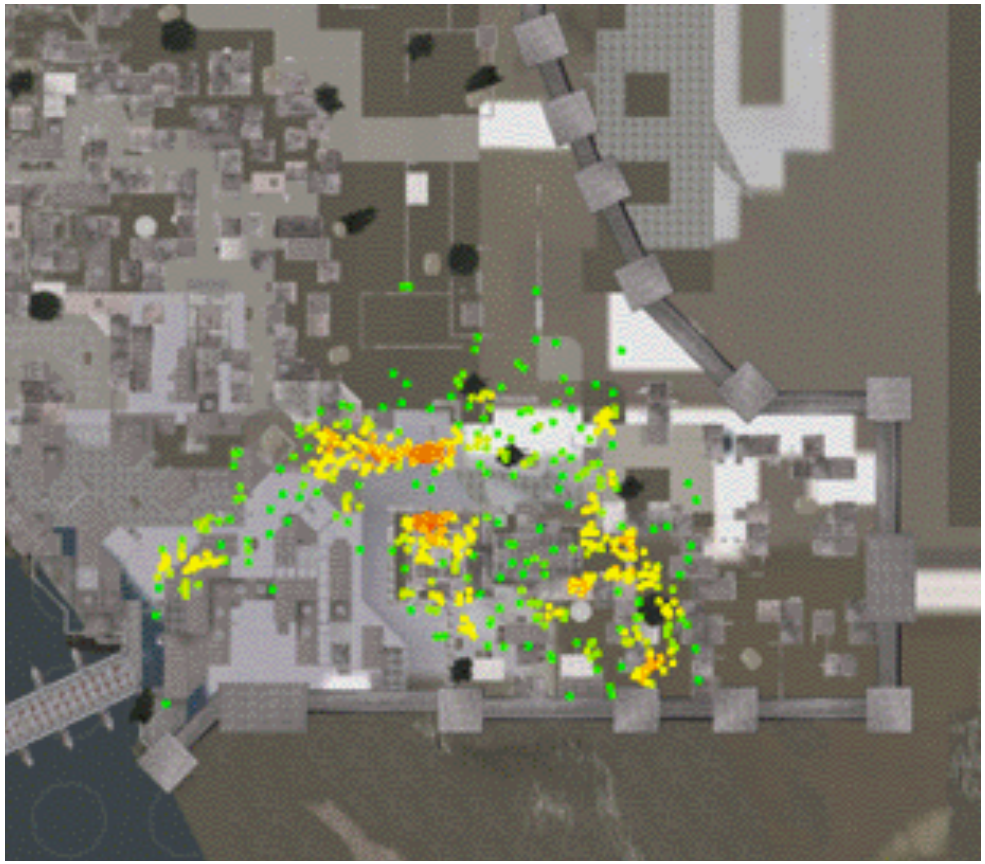


GA Example 1: Tutorial funnel

Percentage of players dropping out after each step:



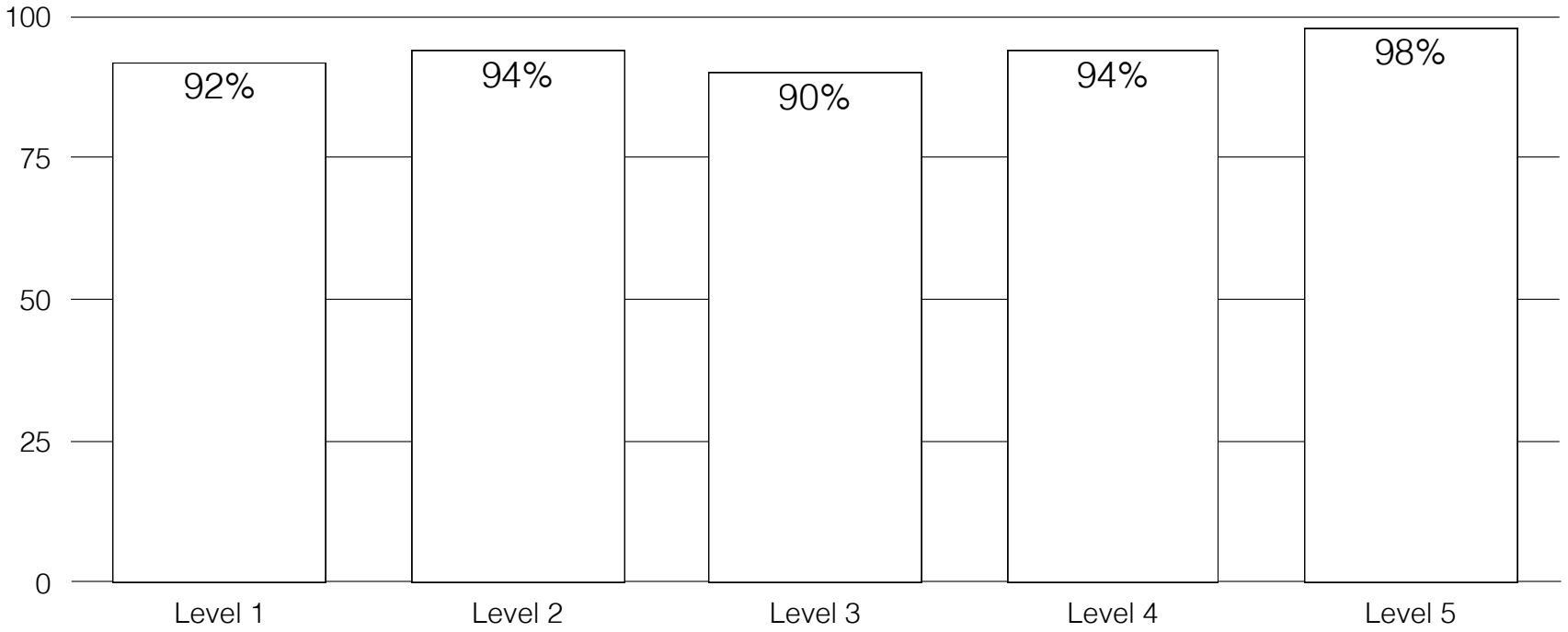
GA Example 2: Deaths Heatmap



J. DANKOFF (AC: BROTHERHOOD)

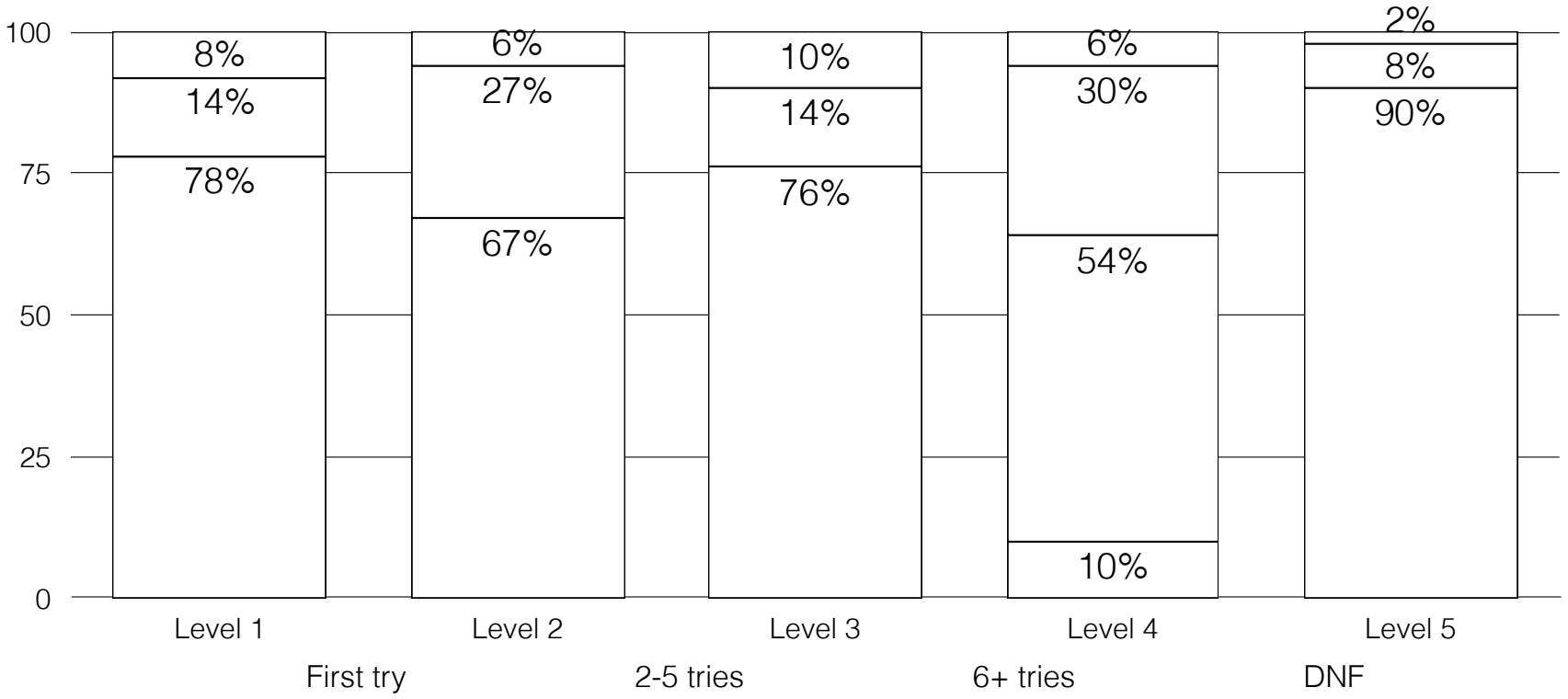
GA Example 3: Game Balancing

Percentage of players who finished each level

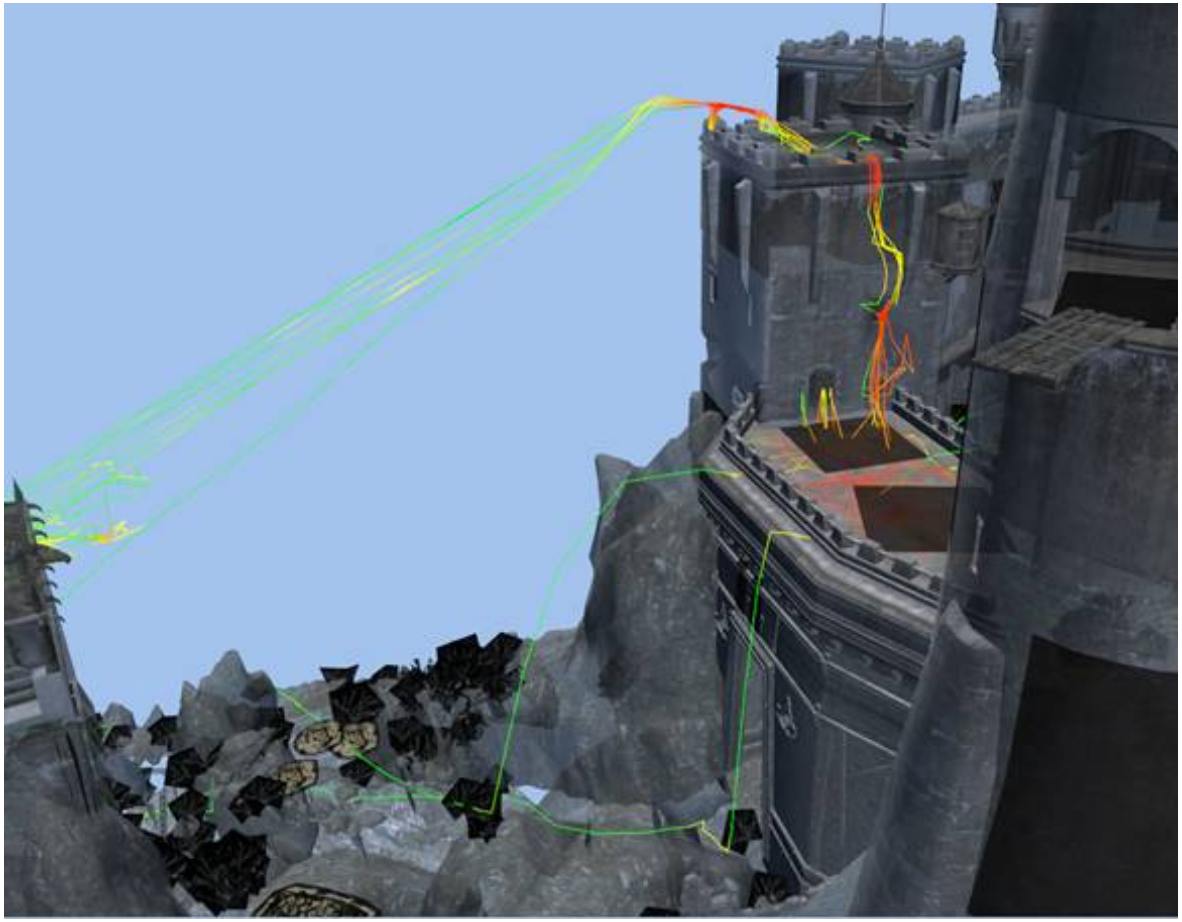


GA Example 3: Game Balancing

Percentage of players who finished each level



GA Example 4: Supporting Observation



J. DANKOFF (AC: BROTHERHOOD)

Week 12 Workshop v1

Write a generic **Initial experience (Hour 1)** questionnaire for AAA AA game.

- Overall rating (1-5)
- Likes / Dislikes
- Difficulty
- Difficulty compared to expectation
- Frustration
- Frustration comments
- Controls
- Visuals
- Likelihood to abandon (1-5)
- Reasons to abandon

Week 12 Workshop v2

Write a generic **Combat** questionnaire for use on AAA AA titles.

- Combat rating (1-5)
- Likes & Dislikes
- Amount of combat
- Variety of combat
- Difficulty of combat
- Difficulty of combat compared to expectations
- Frustration
- Hit clarity
- Player health
- Additional comments