



Good afternoon, Everyone! Welcome to my talk called “Moving the Needle: User Research Towards the End of Production”. Today we are going to delve into how user experience evaluation is executed to make the best possible game once you’re advanced into the development of your game. We are going to use PvZ Garden Warfare as a case study. Let’s get started!

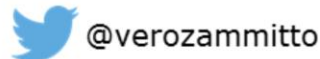


# Veronica Zammitto

Electronic Arts

Lead Games User Researcher

Background in Psychology,  
Human-Computer Interaction (HCI)  
Game Studies



GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



My name is Veronica Zammitto, and I'm the Lead Games User Researcher at Electronic Arts.

One of the amazing aspects of working at EA is the huge portfolio of games. I have the privilege of working on a variety of EA games ranging from sports to shooters, such as FIFA, Madden, NHL, UFC, Dragon Age, Battlefield, and Plants vs Zombies Garden Warfare, among others titles.

# Test early, test often

Good mantra to follow

But every project reaches...

The last stretch of production

## **MOVING THE NEEDLE**

We are all well aware of the mantra 'test early, test often'. However, regardless of how smooth or bumpy the road has been during development, we all arrive to the last stretch of production where core game mechanics are locked and the game is close to its final shape. But, it's not quite there yet. It is that time when we need to keep designing and testing to 'move the needle'.

The background of the slide is a photograph of a snowy, mountainous landscape. The snow is white and appears to be blowing or falling, creating a sense of movement. The sky is a pale blue-grey. The overall tone is cold and serene.

**How can we go from good to great?**

Bottom line, moving the needle is: How can we go from good to great?

## UR Challenges near the End of Production

- No time for radical changes
  - Fine-tune the gameplay
- Deconstruct problem areas
  - Design a methodology and metrics
- Focus on the weak spots
  - Pivot focus based on players' feedback



User research in this stage of development presents particular challenges:

- There is no room for radical changes. the game portrays the design intention but there are still gaps in user experience. This constrain should be seen as a welcome rule in late production. It creates a great scenario for doing user research with a 'fine-toothed comb' where the mechanics are going to be tested closely and only modified based on precise findings. This is fine tuning. This is the core of 'moving the needle'.

- In order to do this fine-tune, it is critical to deconstruct the 'problem areas'. This is related to the part of the game and its game elements to evaluate. At this step, we must be critical of the methodology (session strategy) and the key metrics for success (such as fun, frustration, engagement, navigation, and 'metacritic-like' scales)

- Lastly, prioritization by focusing on the weak spots. Don't be afraid of pivoting the focus on what really needs to be fixed based on players' feedback for which you need to have multiple points of entry (multiple questions) to triangulate data.



# UR near the End of Production

	Early	Later
Goals	Core game mechanics User flow Usability	Progression Fine tuning Balance
Techniques	1-on-1 think aloud Heuristic evaluation	Questionnaires, telemetry
Data	Observation of players' actions Players' utterances Guidelines	Affective <ul style="list-style-type: none"><li>• Feelings, emotions, and attitudes</li></ul> Behavioral <ul style="list-style-type: none"><li>• In-game actions</li></ul>

UR is an integral part of the development process. How it is executed will vary as the creation process of a game varies.

Early in development, such as pre-pro, the game development is more concerned on strengthening the foundations, such as core mechanics and character concept. You'll likely just have wireframes for flows and very early prototypes, generally just bare bones with only very few actions to do. Which are perfect for the first evaluations where changes can be done quickly and cheap.

Whereas later in production you will have a full(er) working build, your characters are fleshed-out and you want to be absolutely positive you're providing the best experience. UR will accompany that switch of focus, and for that there are some techniques that work better than others.

In early stages typical goals are understanding core game mechanics: do they work? Do people get it? 1-on-1 and think alouds are great for that.

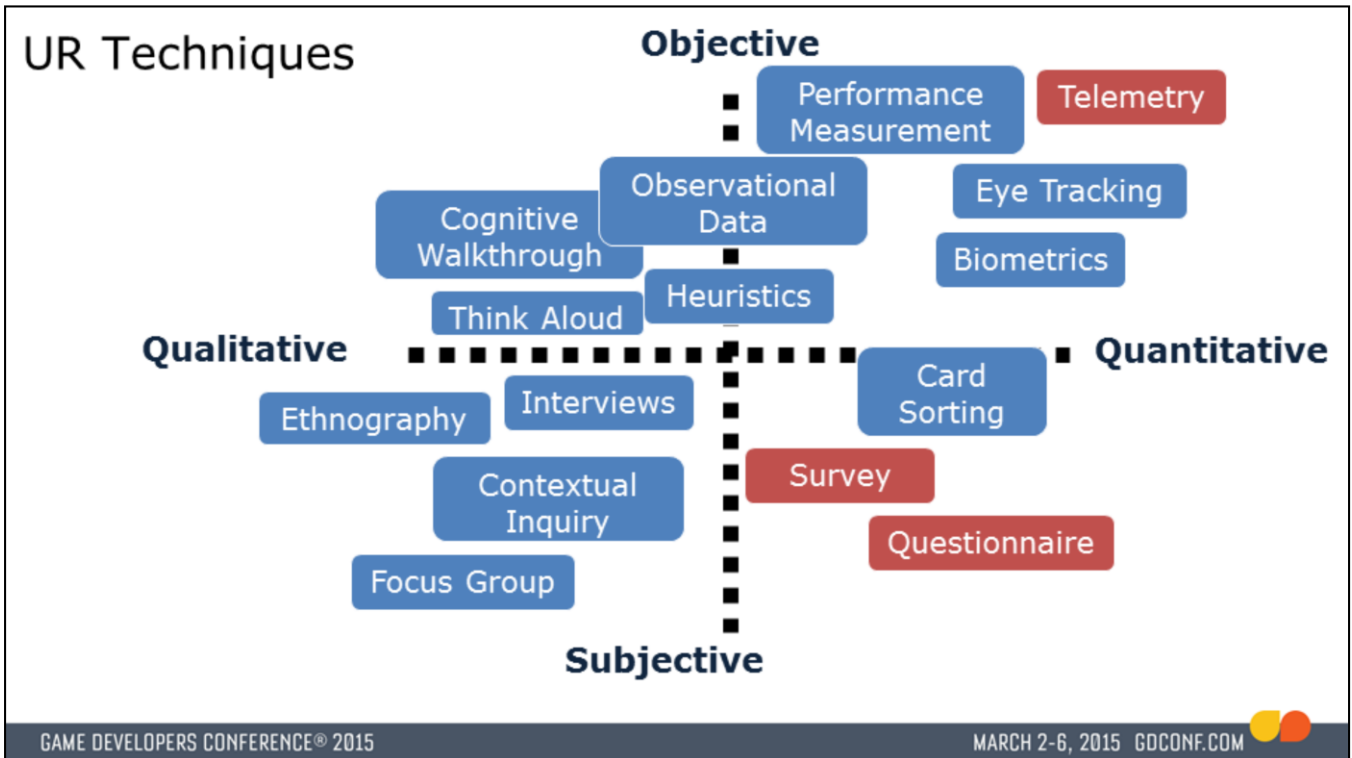
Later goals are more related to players' emotions (affective responses –feelings, emotions, attitudes- e.g. Are they having fun?) and about what they do in-game (how much are they using

the shotgun?)

For instance in FIFA: do players feel empowered as a professional football player? Do advanced players perform skill moves? If so, how that adds to the experience?

Technique wise, questionnaires and telemetry are excellent for that.

Let me give you the full picture of techniques.



There are multiple techniques for UR. Each of them will measure a specific aspect of the UR which makes them more suitable for different moments of game development. I will not get into details of each one, but I want to point out those that are most suitable for the last stretch of development: surveys, questionnaire, and telemetry.

Notice that those are more in the quantitative side, because at this stage of production you need to have a strong reliable numbers. This means bigger sample size of participants in your sessions.

Survey and Questionnaire are key to aggregate users' feedback on how they feel about the game. These techniques will have scales, such as in questions with a sentence indicating from strongly agree to strongly disagree, or from 1 to 5, indicating which options represent better how you feel about the game (e.g. fun). This is the more quant approach to questionnaire, however there will also be open-ended questions to bring up the users' reasons about why those ratings.

Telemetry is key to have direct, automated input about what people do in-game. Bear in mind that telemetry will not appear by itself on the last stages of development, it needs to be integrated and design from early on. Plus, you want to be strategic about what to collect. Designers and Researchers should work



collaboratively to define which data points to collect. Not because you 'can' collected means you 'should' collected everything, you will drown on data. And you want to make sure you collect those that will inform you best.

# Case Study



GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 [GDCONF.COM](http://GDCONF.COM)





Plants vs Zombies Garden Warfare was bringing a known brand into the shooter genre.

Plants vs Zombies Garden Warfare took a traditional, loved tower defense game and turned it into a 3<sup>rd</sup> person shooter. The goal for Garden Warfare is to immerse players in the world of Plants vs Zombies where they get the opportunity to play as their favorite plant or zombie character and join the battle for Suburbia. Garden Warfare was designed for fans of action/shooter games and made aspirational for kids. A game where parents and kids can enjoy the game together. Accessibility was a pillar for PvZ GW.

PvZ GW design vision was to emphasize the plants and zombies qualities (plants being charming and endearing while zombies were the primary source of humour), in a highly cooperative, fun environment, and a healthy dose of PvZ humor.

Here is a video clip playing as a plant.



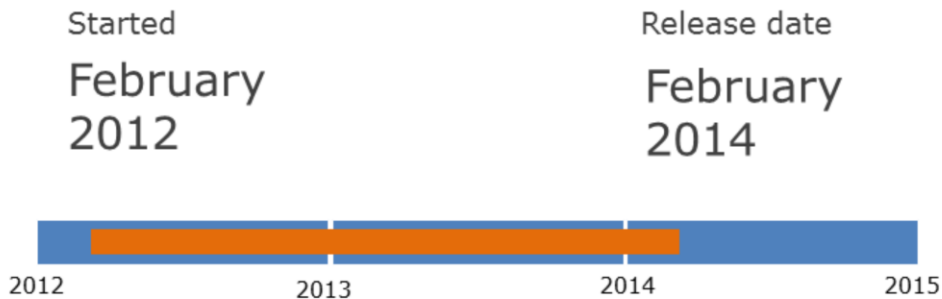
GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



And now a video clip from the zombie perspective. This video is from the mode Gardens and Graveyards.

# Timeline



UR questions: character robustness, maps and factions balance, progression (short and long term engagement)

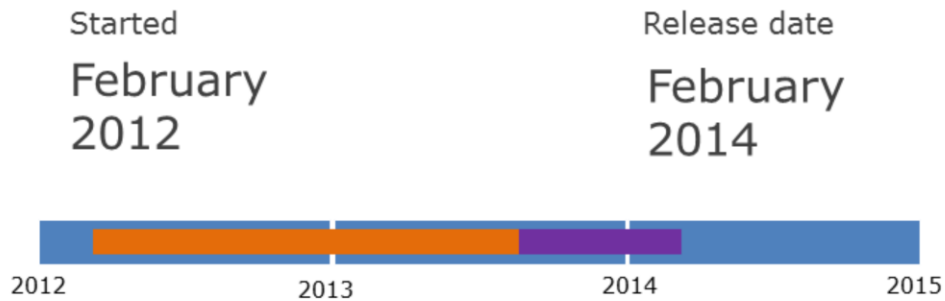


This multiplayer only game had key user research questions at the latest stages of production: plants and zombie factions balance, character leveling having a timely progression ranging from teaching you the abilities to unlocking new characters (early and long term engagement).

This was a multiyear development process, here we'll cover the last couple of months.

And notice that we were testing until the month it was released, which even more underlines the last possible improvement that could be done.

# Timeline



The last stretch of UR testing:  
September 2013 to February 2014



The last stretch of UR testing, the last 6 months of the 2 years project: September 2013 to February 2014



# User Research goals

- Assessing map perception and usage
- Ensuring Balancing of factions
- Validating Characters' playability and experience



We were at a point where the game modes was already defined and the initial reception of the game was on the positive side; however how the map, factions, and characters needed to be scrutinized and improve. We were aiming for more than just a reception, we wanted sustained fun throughout the mode with whichever character you play.

# UR set-up towards end of production

## Study design: Mixed method

- Quantitative and qualitative

## Instrument: Questionnaire

- Affective and Behavioral data

## Playtest session

- 24-48 players
- 4 hours to 2-days



The UR set-up towards the end of production had certain characteristics.

They typically have a study design that involved mixed method, which means including both quantitative and qualitative data. Triangulating data is critical.

Fundamental techniques to use, as it was for PvZ GW, were questionnaire and survey to gather affective and behavioral data.

As mentioned earlier, to strengthen the quantitative side you want a 'big' user session. This is not big data but a experimental session. In each session for PvZ GW, we had between 24 and 48 people, in sessions that lasted from 4 hours to 2 full days.

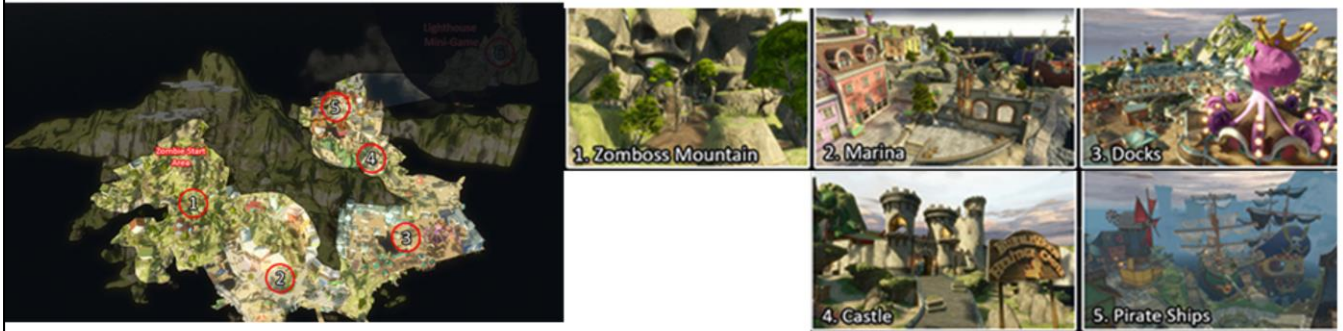


Here is our lab with 24 players at once playing PvZ GW.

# Gardens & Graveyards

## Mode with Expanding map:

- Attack and defend target
- Capturing targets unlocks new sections of the map



GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 [GDCONF.COM](http://GDCONF.COM)



Gardens & Graveyards is a mode where Plants and Zombies face off in an expanding map. This is a similar concept to Battlefield's Rush mode. The zombies are on the attack and need to capture the plant's garden locations – turning them into graveyards. The plants need to stop the zombie invasion. If the timer runs out before the zombies capture the garden the plants win – the timer is extended after each garden is captured. If last garden is captured, zombies win.

A map in Gardens & Graveyards is Coastal map which is actually composed by 5 sub-maps: Zomboss Mountain, Marina, Docks, Castle, and Pirate Ships.

# G&G - Coastal Map

## Deconstructing Maps

- Metrics
  - Map size
  - Cover
  - Population
  - Visual appeal
  - Respawn
  - Navigation



I mentioned how important is to deconstruct the game areas and define metrics.

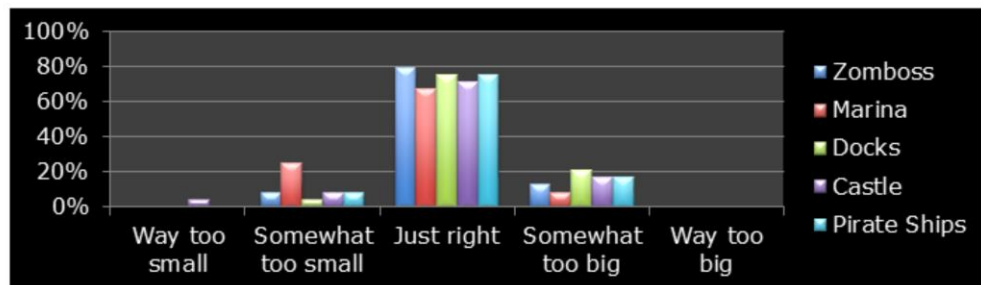
This is how we deconstructed Coastal map for user research. There were 6 metrics that each single map was measured:

Map size, amount of available cover against incoming fire, population (how empty or crowded the map felt), the visual appeal, respawnig locations, and navigation for getting around the map.

# G&G - Coastal Map

## Deconstructing Maps

- Map size
  - Scale question type



GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



All those metrics were implemented into a questionnaire. There were scale type questions for each metrics.

For example, for map size we asked how the map was perceived ranging from way to small to way to big. We wanted to hit that sweet spot of 'just right'.

As you see, overall all sub-maps were performing fairly well, 70-80% of players' experience were matching the design intentions.

The same work for the rest of the metrics was done.

Data supported doing minor tweaking for more amount of cover in Zomboss Mountain and respawning in Castle.

However...



# G&G - Coastal Map

**Ranking** of the 5 maps within Rush Coastal accordingly to users' PREFERENCE and FUN

1.



- (+) Open air fun
- (+) Several interesting ways of approaching
- (+) Good cover
- (+) Good scenery
- (+) Empowers teamwork

2.



- (+) Interesting choke points
- (-) Felt small
- (+/-) Difficult to breach

3.



- (+) Good cover
- (+) Good scenery
- (+) Easy to navigate

4.



- (+) Good scenery
- (+) Felt balanced
- (-) Felt generic

5.



- (+) Vertical combat
- (+) Good cover
- (-) Too easy to fall and die in water areas
- (-) hard to navigate
- (-) spawn far from action

GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



The scale type questions were complemented with ranking questions, and we found that 1/3 of players dislike Pirate Ships.

At the end of the session for Gardens & Graveyards, we ask participants to rank all 5 maps based on their preference and fun. On both questions, the ranking was identical.

It was evident Pirate Ship need more focus than any of the other maps.

Data helped to identify the weakest spot, prioritize where to make changes, and keep pushing the improvement of PvZ GW. This is an example of pivoting focus based on players' feedback. The time left in the project is more than precious. Triangulating responses reinforces the priority of an issue.

Players had trouble navigating the map, were doing easily due to environmental issues, and the spawn locations were seemed too far from the action.

# G&G - Coastal Map – Pirate Ships



GAME DEVELOPERS CONFERENCE® 2015

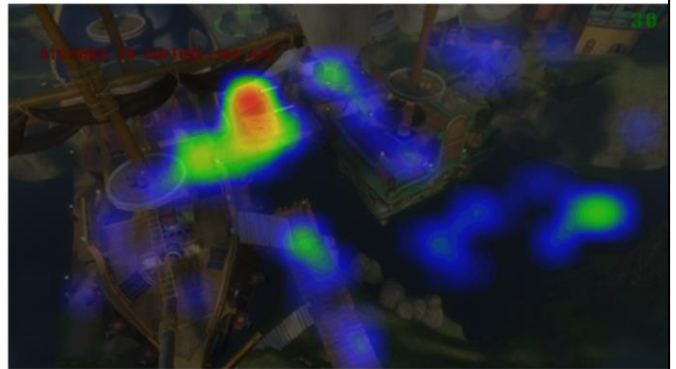
MARCH 2-6, 2015 [GDCONF.COM](http://GDCONF.COM)



This is an 'aerial' screenshot of the Pirate Ship map. The design intention for this map was to 'recreate' a pirate battle and create the tension as if boarding the other faction's ship. Zombies started from the bottom side in the screenshot, Plant's garden is seen at the top (highlight in green)

We used another technique which is Heat Map based on players' preferences. Within the questionnaire they were shown the image and asked to click on their most preferred location and again on their least location.

## Least Preferred Locations



*"I do not like the shoreline which looks like it's giving you an additional path but doesn't. The ship in the middle feels badly placed."*

Main reason for least preferred locations:

- Easy to fall into the water and die



Players' answers created this heatmap. Bear in mind is based on attitudinal data (dislike) and not on telemetry. If the image on the right were a telemetry death map, the heat zones wouldn't be the same, people were also dying a lot on the battle which is fine. For sure those deaths would have appeared in the telemetry but not to the extent to raise this as a red flag and definitely not as 'the frustrating' component. Remember telemetry tells you what but not why. This affective heat map tells as where people had frustrating moments. Navigation wise the shoreline was a deadly trap. Players were also dying a lot falling into the water between the ships and there was nothing preventing that.

The supposed-epic, boarding pirate battle was turned into an involuntarily 'walk off a plank'.

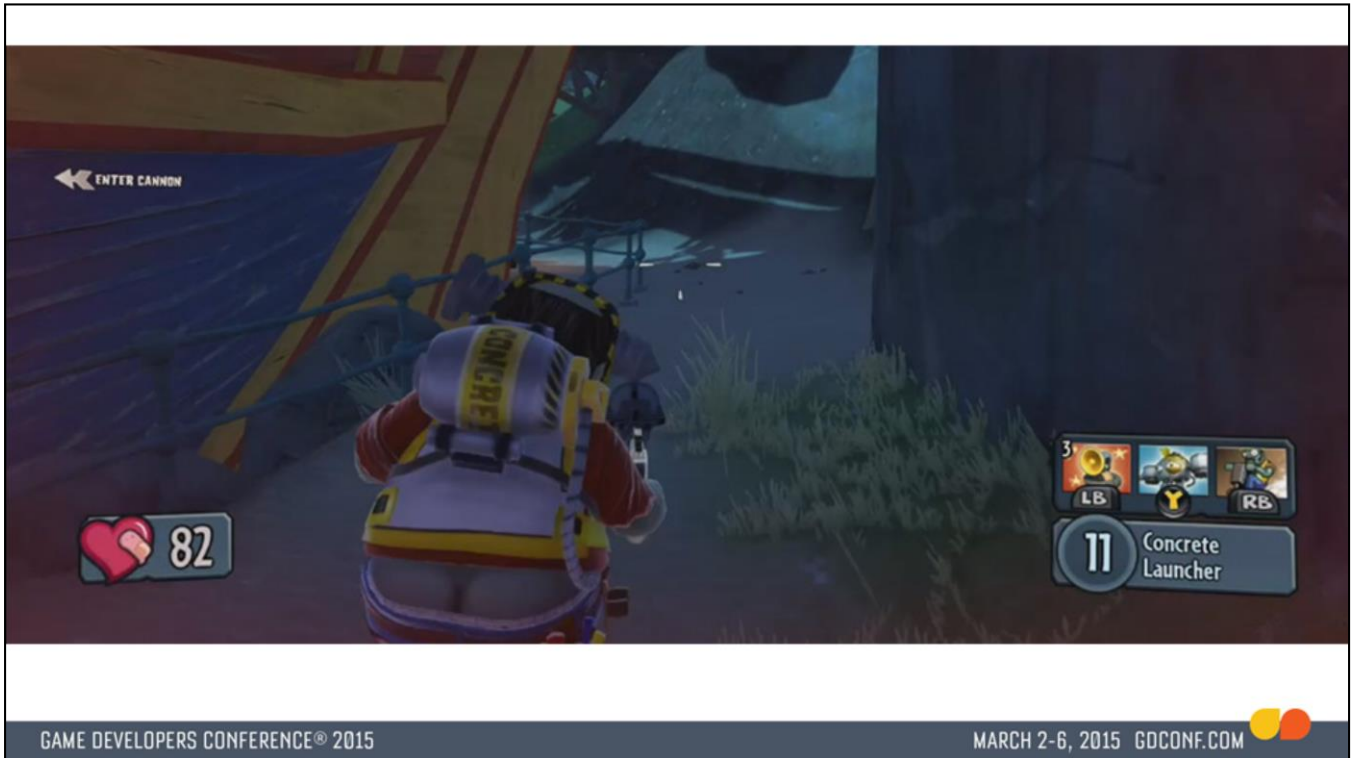


After each session, we sat down with the design team, go over the UR report together, discuss the issues, dig further into the data if needed, start the conversation on what changes should be done.

There were concrete changes in the Pirate Ships map. Ships were rearranged to still offer verticality in the map, but the 'epic' boarding battle was toned down.

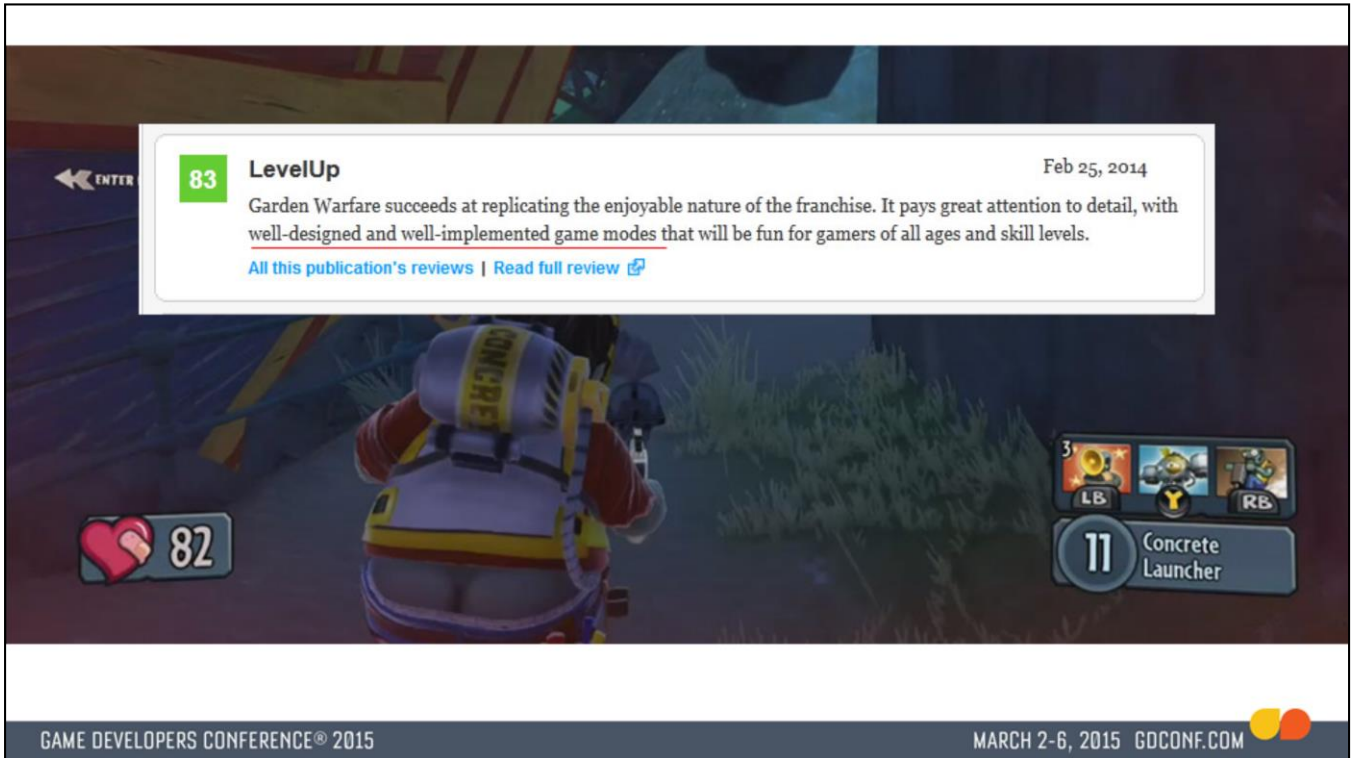
Barriers were added to indicate that shore/water was off boundaries.





Ships were rearranged to still offer verticality in the map, but the 'epic' boarding battle was toned down.

Barriers were added to indicate that shore/water was off boundaries.



We also compared against professional reviews to see how the game was seen and we nailed it!

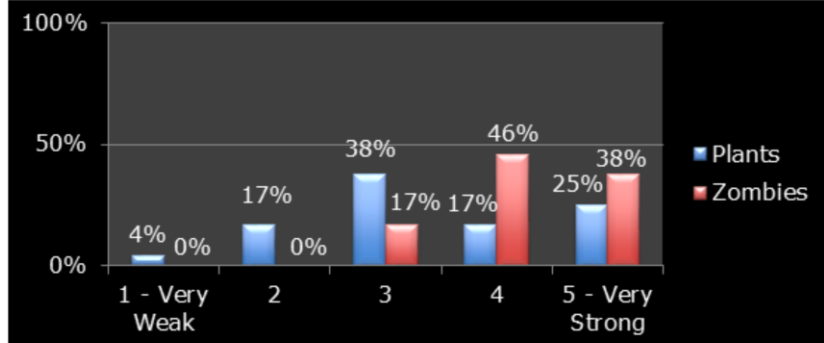


# G&G - Balancing of Factions

## Faction Strength:

(-) Zombies were perceived as stronger than Plants  
(-) Zombies won 28 out of 30 battles

Based on your experience in this mode and map, how would you rate the strength of the two factions?



Another key design aspect was balancing.

Balancing of factions is very important for the feeling of fairness. It's about the skills of the players and not about favoring a faction.

In a mode like Gardens & Graveyards where you have roles like attackers and defenders, it is challenging how the same parameters of health and damage can be perceived differently compared to other modes like Team Vanquish where everyone is even. At the beginning, the zombies were seen as primarily strong.

# G&G - Balancing of Factions

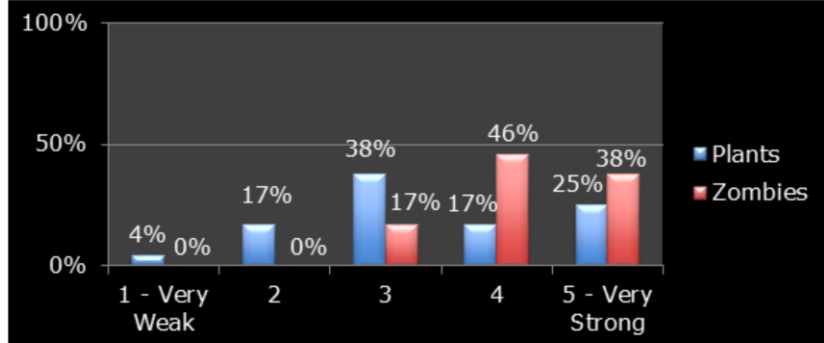
## Faction Strength:

(-) Zombies were perceived as stronger than Plants  
(-) Zombies won 28 out of 30 battles

## FUN (3.42/5):

(+) Teamwork  
(+) Relentlessness action  
(+) AI summoning  
  
(-) balance between factions (zombies found overpowered and plants unable to defend)

Based on your experience in this mode and map, how would you rate the strength of the two factions?



GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



Perceived unbalanced factions had a detrimental impact on Fun.

This illustrates how UR directly impacts improving the overall perception of the game and points out to specific game elements related to key success metrics such as fun.

Tuning of health and weapons was done after this test for Gardens & Graveyards.

# G&G - Balancing of Factions

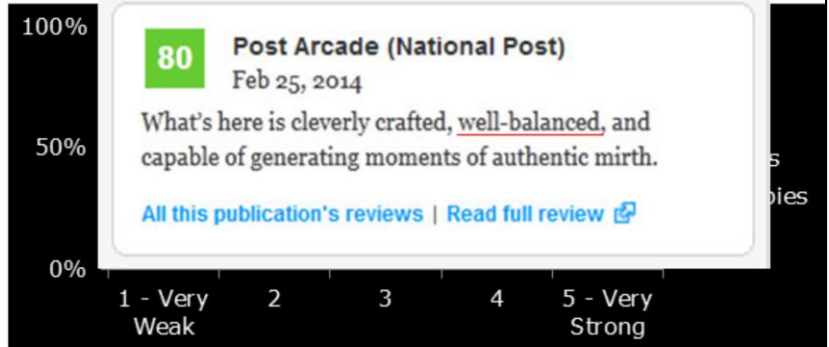
## Faction Strength:

- (-) Zombies were perceived as stronger than Plants
- (-) Zombies won 28 out of 30 battles

## FUN (3.42/5):

- (+) Teamwork
- (+) Relentlessness action
- (+) AI summoning
- (-) balance between factions (zombies found overpowered and plants unable to defend)

Based on your experience in this mode and map, how would you rate the strength of the two factions?



And one more time, we nailed it!

# Characters



GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 [GDCONF.COM](http://GDCONF.COM)



There are 8 core characters (4 per faction) in PvZ GW

# Characters

- Deconstructing Characters
  - Fun
  - Strength
  - Health
  - Abilities
- How zombies perceived plants and vice versa



Characters were mostly fleshed out. There was no room for radical challenges, all abilities would remain.

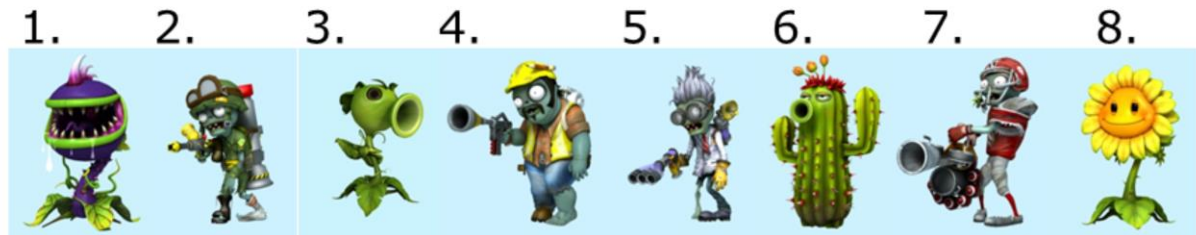
The challenge was to make them the best by fine tuning.

Similarly to maps, we deconstructed characters and we defined 4 metrics to measure each of them.

We also asked how each faction was seen by each other.

# Characters & Fun

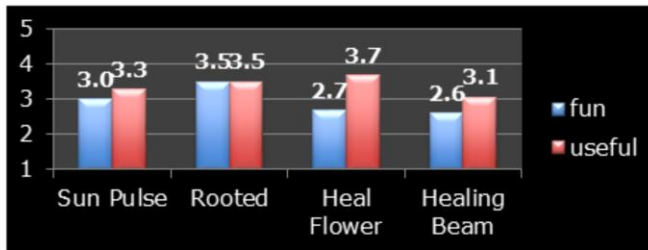
**Ranking** of all 8 characters accordingly to users' fun



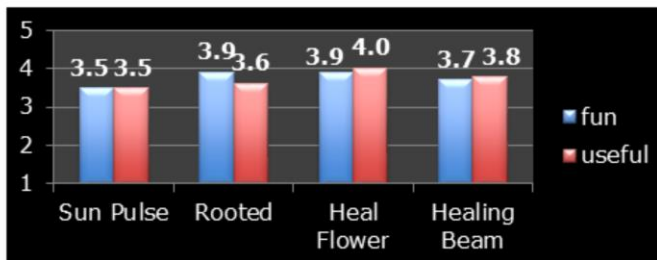
By this session, the priority for tuning was done around sunflower. The healing role, of course, stayed but tweaking the values for health and damage.



Sept  
2013



Feb  
2014



GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



We discussed with the team and debate about the role of a healer and overall reception.

We also had sessions where people could freely play whatever characters they wanted. Other sessions we had people pre-assigned to characters based on best match of their shooter playing styles.

But the weak points still stood around the Sunflower.

The design team action on the tweaking the Sunflower's abilities which are shown in the graphs. The top graph shows metrics from September and the bottom graph metrics from February. We saw an increased on all abilities based on changes.

This example shows you the power of repeated measurements (tracking the same metrics) to be confident the changes were effective and by how much.



After all the revision done to the characters, you would imagine that we were done with them. Woohoo, dance!

But no, there was more work to be done. There were more characters.

Pea Shooter



## Character Subtypes

- Each character has 6 subtypes
- Subtypes have variations on the same abilities



Peashooter



Fire Pea



Ice Pea



Toxic Pea



Commando Pea



Agent Pea

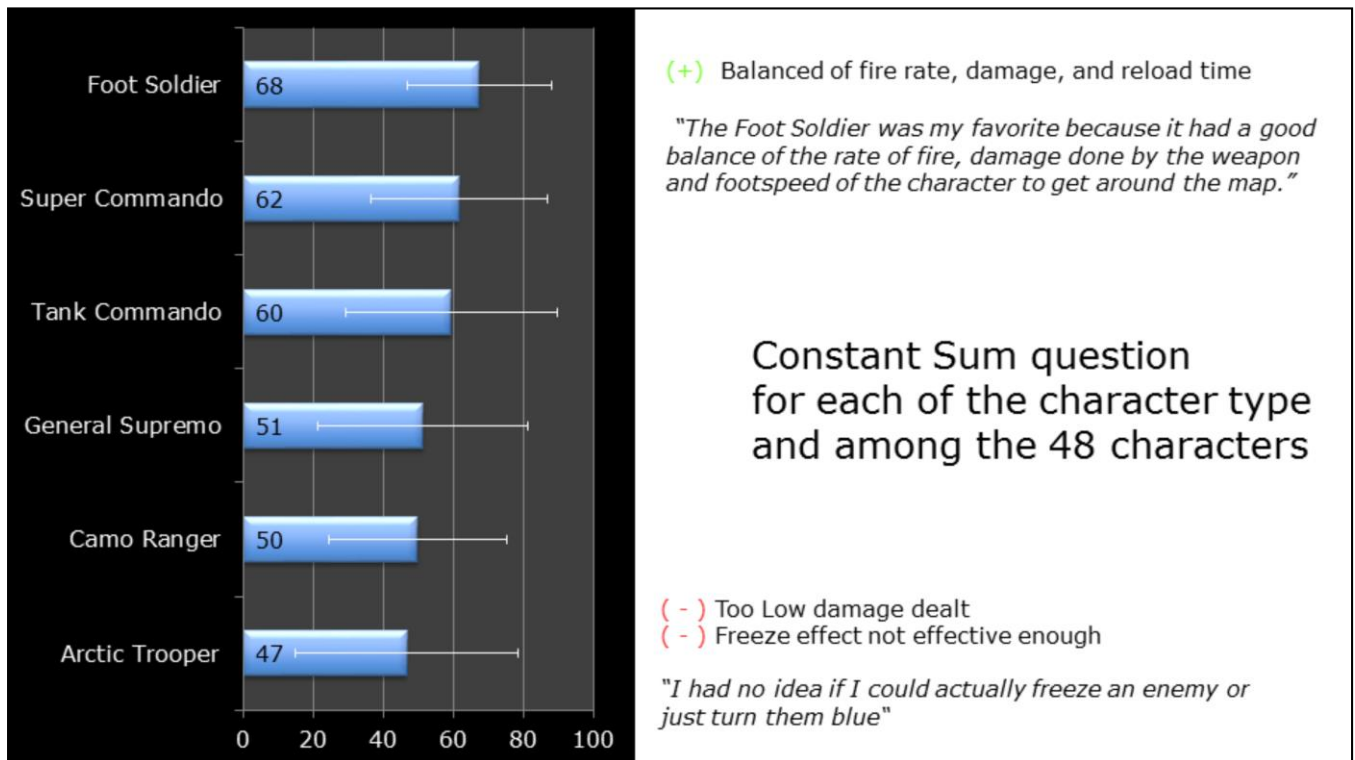
For each character, there were 5 more alts. This means that at launch there were 6 pea shooters, 6 chompers, 6 cactus, and so on.



This totals 48 characters.

That gives perspective about the need to have the 'main' characters very well-rounded.





The challenge was on how to evaluate 48 different characters in a timeline manner and confident about results. This tides into the importance of tailoring methodology. In the previous test we could ask all participants to play 8 different characters and rank them, but 48 is different. We did a big session with 48 players across a couple of days to play multiple rounds with different characters. Similarly we measured for each of the 48 characters: fun, frustration, visual appeal, strength, health, weapon range, and reload speed. Tweaking and tuning was done to match design intention.

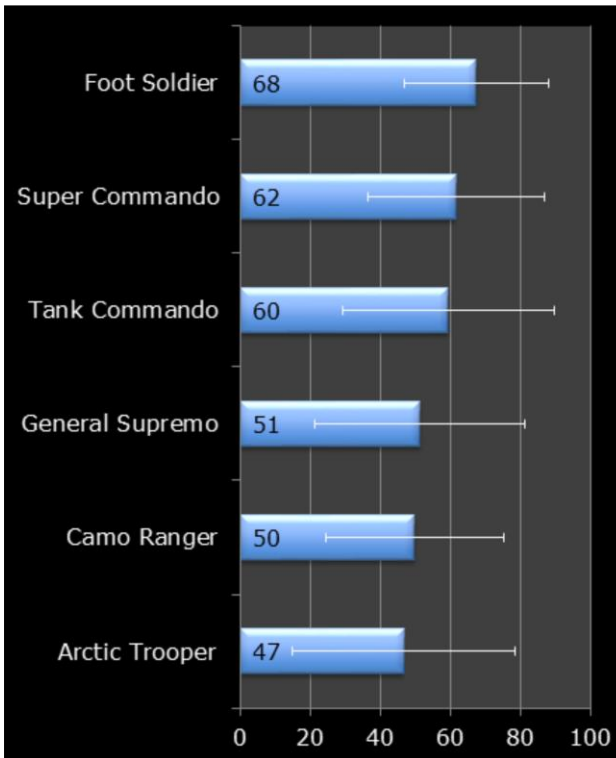
We also employed another techniques which is called 'constant sum', players were asked to give points (up to 100 to each character) based on how much they preferred it. The more points, the most preferred.

This generated a precise number for each character, which allows to compare preferences based on that value. We also compared among each character type, for example among all foot soldiers.

In this Foot Soldier zombie class graph, we have them sort out by the average points they received. There was a noticeable gap with Arctic Trooper. His unique main weapon was not performing as expected for a freeze weapon. The goal was to identified unique aspects that could be detrimental to enjoyment.

This technique allowed us to quickly identified the weak spot and focus on that.

Same analysis for each class and among all characters.



(+) Balanced of fire rate, damage, and reload time

*"The Foot Soldier was my favorite because it had a good balance of the rate of fire, damage done by the weapon and footspeed of the character to get around the map."*

100

Game Revolution

Feb 25, 2014

Sure, you could spend your entire time shooting at both living and dead, but the best gameplay comes from varied and mechanically diverse controllable characters.

[All this publication's reviews](#) | [Read full review](#)

(-) Too Low damage dealt

(-) Freeze effect not effective enough

*"I had no idea if I could actually freeze an enemy or just turn them blue"*

Nailed it!

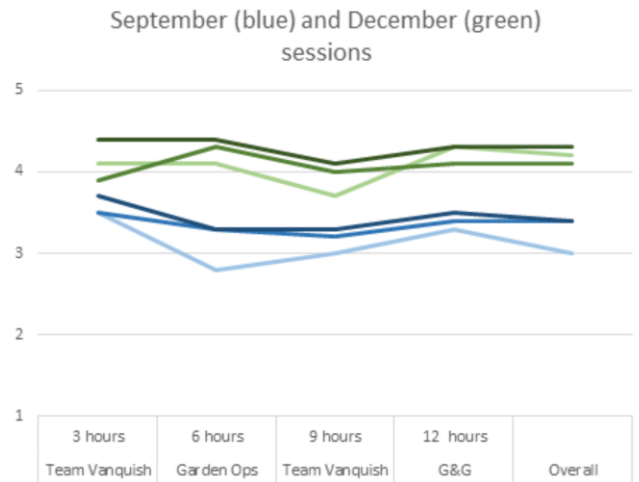


# Moving the Needle #1– PvZ GW

## Longitudinal measurements

Scale question from 1 (Not at all) to 5 (A lot)

Sept	Dec	
		How motivated are you to continue playing PvZ?
		How engaging have you found PvZ so far?
		How much are you enjoying the game?



# Moving the Needle #2– PvZ GW

1- Terrible	0%
2- Bad	0%
3-Poor	0%
4-Budget	4%
5-Average	4%
6-Good	25%
7-Great	42%
8-Amazing	25%
9-Outstanding	0%
10-Flawless	0%

Average: 6.8

1- Terrible	0%
2- Bad	0%
3-Poor	0%
4-Budget	4%
5-Average	0%
6-Good	0%
7-Great	57%
8-Amazing	39%
9-Outstanding	0%
10-Flawless	0%

Average: 7.3



## Metascore

Generally favorable reviews  
based on [65 Critics](#)

[What's this?](#)

**Summary:** Plants vs Zombies: Garden Warfare has players blast zombies, plants, and new characters with inventive weapons across a mine-blowing PvZ world.

GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



We also kept track of overall perception of the game. We used this 10-point scale that is composed by key adjectives from game reviews which closely represent Metacritic's score bands.

This allows us to see how much we are 'moving the needle'.

In an early session, the game averaged 6.8.

By a later session the average was increasing to 7.3, notice that more than half of players were seeing the game as 'great' and almost 40% rating it as 'amazing'.

The Metacritic score for PvZ GW was 76.

We were able to evaluate players' experience closely to what the final Metacritic was, the UX testing helped to provide concrete insights on what to change throughout the last stretch of production.

6.8, 7.3, 7.6: These are exactly the small increases for improving the game, moving the needle to more.

76

### Design intention:

- PvZ tower defense to shooter
- Fun and humor
- Experience but accessible

### Back fire:

- Gradual Content
- Game as a service

70

#### GameSpot

Feb 25, 2014

While the unlockables fit right in with the whimsical spirit of the Plants vs. Zombies franchise, you're left feeling like Garden Warfare needed a little something more to make it a meal rather than a side salad. Still, it's that rare shooter whose world and characters might put a smile on your face, making it a refreshing entry in a genre that usually takes itself really seriously.

70

#### Gamereactor Sweden

Feb 25, 2014

PopCap's lovely tower defense game has been transformed into hectic action with all the features anyone can wish for in a game that can be enjoyed by both younger and older gamers. Only lack of gameplay content prevents this title from higher grades.

GAME DEVELOPERS CONFERENCE® 2015

MARCH 2-6, 2015 GDCONF.COM



You might be thinking, only 76? Most games want to break into the 80's. Are you proud with a 76?

The answer is yes.

When I started this talk I mentioned:

"Plants vs Zombies Garden Warfare is going to take a traditional, loved tower defense game and turn it into a 3<sup>rd</sup> person shooter. The game modes are also all quite unique yet have some familiarity to fans of multiplayer shooters, with a core game mode being Gardens & Graveyards. The goal for Garden Warfare is to immerse players in the world of Plants vs Zombies where they get the opportunity to play as their favorite plant or zombie character and join the battle for Suburbia. Garden Warfare was designed for fans of action/shooter games and made aspirational for kids. A game where parents and kids can enjoy the game together."

PvZ GW design visions was to emphasize the plants and zombies qualities, in a highly cooperative, fun environment,

and a healthy dose of PvZ humor.

All reviews, including those in the low 70's, actually recognized PvZ GW delivering the design intention and that's why we are satisfied with the Metacritic score.

Review on PvZ GW consistently emphasize how the shooter genre was refreshed, how fun the game is, and how accessible the controls were.

"Whimsical spirit", "smile on your face", "refreshing entry", "transform [genre]".

All pillars were achieved.

However, there were a couple of aspect that backfired: gradual introduction of content and the concept of games as a service.

The game was meant to be released with some game modes and then push new modes, more characters, and maps through time. That was not conveyed well in the game, and people evaluated the game based on the general understanding against what comes with in a packaged game. As the review called them, "lack of gameplay content" or as in GameSpot's 70 metacritic, it positively hits on all those core design intentions. But also underlies that content component being closer to a side food size. Obviously the model of "game as a service" that the industry is starting to explore it's still not very well fleshed out, and PvZ GW was not the exception. A better job could have done at the ongoing, increasing content for PvZ GW and how it was meant to support a richer, deep UX.

We are satisfied that the game made it true to its design intentions which players embraced. The solid game mechanics and content presented in the game were polished and consistently called out positively.

# Takeaways

- No time for radical changes
  - Fine-tune the gameplay
- Deconstruct problem areas
  - Design a methodology
    - Scale-type, rank-order, heatmaps, constant-sum questions
    - Longitudinal, repeated measures
    - Large sample size
- Focus on the weak spots
  - Pivot focus based on players' feedback
    - Affective and behavioral data



The takeaways for this talk are:

- There is no room for radical changes. the game portrays the design intention but there are still gaps in user experience. This constrain should be seen as a welcome rule in late production. It creates a great scenario for doing user research with a 'fine-toothed comb' where the mechanics are going to be tested closely and only modified based on precise findings. This is fine tuning. This is the core of 'moving the needle'.

-Deconstructing the game areas into measurable metrics that will be tracked across sessions. I pointed out to the study design and specific techniques employed for the sessions. Remember the examples on different techniques including the scale-type, the rank-order, affective heatmaps, and constant-sum question. Each of them will help towards a comprehensive understanding of the user experience. Remember to triangulate data and to collect it across sessions.


Keep a large sample size for session with quantitative data.

-Remember, you are on the last stretch of production, you want to focus on the weak spots and not being concerned on pivoting focus as players' feedback which should include affective and behavioral data.

These three key learnings will help you to assess the UX, identify concrete design improvements, and make a better game. Moving the needle!

# Thanks!

Veronica Zammitto

 @verozammitto

