Playability Evaluation: Day 1

DOM-E5082 Elisa Mekler

Schedule

- Monday: 10:15am 4pm (or something like that)
 - Overview of GUR methods
- Tuesday : 10:15am TBD
- Christian Guckelsberger: Al-based playtesting
- Wednesday Thursday: Independent work
- Friday: 10:15am 4pm
 - Presentations and discussions

Other Details

- Mostly practical material
- Show up, do the work, and pass
- I don't know what you know
- Let me know if you want to learn any other (related) skills

Take-Aways

- The basis for UX work in games and the core challenges associated with evaluating experiences in games
- The context for the application of user research in game development
- The method space for games UX and the many dangers to result validity.
- The relationship between UX and Analytics (behavioral telemetry)
- Best practices in game UX reporting and playtest structuring

What is GUR?

- Games user research (GUR) is a core part of game development
- Helps games reach their design goals by understanding players.

What is GUR?

- GUR is also an academic area which seeks to better understand what motivates players, how their actions can be explained or predicted, or even just to find new ways to capture and use data about players to help with game design.
- GUR relates to psychology, human factors and ergonomics, user experience design, interaction design, computer science, and many other fields.

Games User Research

• Evidence-driven

- Improves player experience
- Finds weaknesses in game design
- Occurs across all stages of development



Games User Research

- Evaluating (how players interact / feel about) games
 - Observing play
 - Player interactions with game elements of interest
 - Telemetry data
 - Analyse data
- Supports iterative development
- Reflection on design
- Telling (potentially) hard truths to designers



https://taels.net/bentaels/2015/23/05/theusability-of-bloodborne/

A brief history of GUR



A brief history of GUR

1971

Nolan Bushnell installs Computer Space in the Dutch Goose restaurant and watches people play it.



1977

Carol hires Colette Weil, Mary Takatsuno (later Mary Fujihara) & Linda Benzler (later Adam). They are the first Gamer User Research team.



Hubris before the Fall Negotiations for the game rights ended in late July 1982, giving Warshaw just over five weeks to develop it in time for the 1982 Xmas season [...] Atari anticipated enormous sales based on the popularity of the film, as well as the stability of the video game industry in 1982. Due to time limitations, *Atari skipped audience* testing.



Hubris before the Fall • *E.T.* is often cited as one of the worst video games of all time and one of the biggest commercial failures in video game history. It is cited as a major contributing factor to the video game crash of 1983



A brief history of GUR

NAME OF GAME				<u>u</u>			
SYSTEM	7	COMP	PANY				
GAME TYPE		EVAL	UATOR				
DATE	ITE		ISSUE REVIEWED IN NINTENDO POWER				
	2						
C			T				
EVALUATION	POWER PLAYER	POWER METER	Evaluator's Comments				
C EVALUATION B=Graphics & Sound	POWER PLAYER	POWER METER	Evaluator's Comments				
EVALUATION G= Graphics & Sound P= Play Control	POWER PLAYER	POWER	Evaluator's Comments				
C EVALUATION D=Graphics & Sound D=Play Control C=Challenge	POWER PLAYER	POWER METER	Evaluator's Comments				

1997

Bill Fulton starts doing Games User Research at Microsoft. Later in the year Howard Phillips (of Nintendo) hires Bill to run the playtesting department. Bill reformulates Playtesting to be a rigorous in-house survey methodology for measuring emotions.



1997

Michael Medlock, Kevin Keeker and Ramon Romero join Bill Fulton later in 1997. Microsoft now has a Games User Research team.



2000

With the Xbox coming, the Microsoft Games User research team grows from 4 people to 13 people.







A brief history of GUR

Score: 700'000'000'000 pts

Usability vs. Playability

Earn 100'000'000'000'000 pts

Press the button to earn 100'000'000'000 pts

- **Usability:** interactions are effective, efficient, and satisfying (ISO 9241-11 standard)
- **Playability:** the interface is unobtrusive, design intent is clear, and the game is suitably difficult and engaging (Korhonen, 2016)

	UX Usability Goals: Productivity	P	X Playability Goals: Enter tainment
1.	Task completion	1.	Entertainment
2.	Eliminate errors	2.	Fun to beat obstacles
3.	External reward	3.	Intrinsic reward
4.	Outcome-based rewards	4.	Process is its own reward
5.	Intuitive	5.	New things to learn
6.	Reduce workload	6.	Increase workload
7.	Assumes technology need to be	7.	Assumes humans need to be
	humanized		challenged

Games vs. Productivity Applications	Examples
Process vs. results	The purpose of gaming is usually in the process of playing, not in the final result.
Defining goals vs. importing goals	Games (or playerss) usually define their own goals, or how to reach a game's goal. However, in productivity applications, the goals are usually defined by external factors.
Few alternatives vs. many alternatives	Games are encouraged to support alternative choices to reach the overall goal, whereas choices are usually limited in productivity applications.
Being consistent vs. generating variety	Games are designed to provide a variety of experiences. However productivity applications are meant to be consistent in the user experience.

	Games vs. Productivity Applications	Examples	
\ <i>/\/</i> C	Imposing constraints vs. removing or structuring constraints	Game designers intentionally embed constraints into the game loop, but productivity applications aim to minimize constraints.	
	Function vs. mood	Productivity applications are built around functionality, but games set out to create mood (for example, using sound or music to set a tone).	
lity	View of outcome vs. view of world	Players usually play a role in a game world such as race car driver, soldier, warrior, etc. Productivity applications rarely have a point of view.	
	Organization as buyer vs. individual as buyer	Individuals usually buy games, but productivity applications are often bought by organizations.	

Games vs. Productivity Applications	Examples
Form follows function vs. function follows form	Players tend to welcome innovation while users of productivity applications tend to be cautious about adopting innovation.
Standard input devices vs. novel input devices	Games usually explore possibilities to use novel input methods, such as motion capture or biofeedback, in addition to standard input devices. Productivity applications mostly rely on a mouse and keyboard.

GUR Methods

Questions before deciding on method

- Who is the product for? Who are the users?
- Who should the users be?
- What should the product be? What do our users do with our game?
- What do our users wish they could do? What should our business model be? How should the product be made?

Methods Overview



What are you evaluating?



- From top to bottom from the designer's perspective
- But Guresearcher needs to consider from bottom to top

What are you evaluating?



Research Questions

- Current state: How many times do players get lost?
- Compare: Do players perform better with button mapping X, or Y?
- Affinity: How should game objects be grouped?
- Needs: What do players actually want (consciously or not)?
- Generative: Produces new design ideas

	Current State	Compar	Affinity	Needs	Generat
A/B Testing	No	Yes	No	No	No
Benchmark Playtest	Yes	Yes	No	No	No
Card Sort	No	No	Yes	No	No
Critical Facet Playtest	Yes	Yes	No	No	No
Diary/Camera Study	Yes	No	No	Yes	Yes
Ethnographic Field Study	Yes	No	No	Yes	Yes
Extended Playtest	Yes	Yes	No	No	No
ocus Group	Yes	No	No	Yes	Yes
Heuristic Evaluation	Yes	No	No	No	No
nterview	Yes	No	No	Yes	Yes
nitial Experience Playtest	Yes	Yes	No	No	No
Market Segmentation	Yes	No	No	No	No
Narrative Usability	Yes	No	No	No	No
Online Survey	Yes	No	No	Yes	Yes
Personas	Yes	No	No	Yes	No
Review	Yes	No	No	No	No
RITE Test	Yes	No	No	No	No
Felemetry Analysis	Yes	Yes	No	No	No
Jnmoderated Usability Test	Yes	No	No	No	No
Jsability Benchmark	Yes	Yes	No	No	No
Jsability Test	Yes	No	No	Yes	No

ison

Ve

Overview of GUR Methods

• Common methods:

- A/B Testing
- Observation
- Interviews
- Surveys
- Heuristic Evaluation

• 'Common' is relative to the studio size / culture

	Current State	Comparis	Affinity	Needs	Generativ
A/B Testing	No	Yes	No	No	No
Benchmark Playtest	Yes	Yes	No	No	No
Card Sort	No	No	Yes	No	No
Critical Facet Playtest	Yes	Yes	No	No	No
Diary/Camera Study	Yes	No	No	Yes	Yes
Ethnographic Field Study	Yes	No	No	Yes	Yes
Extended Playtest	Yes	Yes	No	No	No
Focus Group	Yes	No	No	Yes	Yes
Heuristic Evaluation	Yes	No	No	No	No
nterview	Yes	No	No	Yes	Yes
nitial Experience Playtest	Yes	Yes	No	No	No
Market Segmentation	Yes	No	No	No	No
Narrative Usability	Yes	No	No	No	No
Online Survey	Yes	No	No	Yes	Yes
Personas	Yes	No	No	Yes	No
Review	Yes	No	No	No	No
RITE Test	Yes	No	No	No	No
Telemetry Analysis	Yes	Yes	No	No	No
Unmoderated Usability Test	Yes	No	No	No	No
Usability Benchmark	Yes	Yes	No	No	No
Usability Test	Yes	No	No	Yes	No

S

Ð

Overview of GUR Methods

A/B Testing

- Players are randomly assigned to play one of two (or more) slightly different games
- The behaviour of interest (e.g., spending, retention) is measured (via telemetry) to compare design performance

• Generates a definitive answer

 Quantifies the influence of design changes

Pros

 One of the few actual experiments (i.e., perceived rigour)

- Requires a large enough player base for confidence in results
- Can be difficult / expensive to set up
- Designing, coding, and testing each condition takes time / labour
- Does not explain why results occurred
- Only finds local maxima

- A researcher observes a tester's play behaviour while taking notes
- Ideally, a second researcher facilitates the session, and the observer is elsewhere (e.g., watching on a separate screen)

Pros

Observation

- See how 'typical' players actually behave in the game
- Can be done early in development
- Can provide specific answers to some questions
- What is happening

- Players are rarely as honest as you might like
- Presence of observers can bias results
- Behaviour requires interpretation -> why things are happening unclear
- Player feelings mostly unknown
- Can be time-consuming
- Outliers may be misinterpreted as trends

Interviews

• A researcher asks individual players about the topic(s) of interest, while taking notes and recording audio

Pros

- Generates contextual data that can stand alone, or explain other results
- Follow-up questions can help generate new insights
- An audio record provides a full account for later reference

- Researcher bias is not always obvious
- Can be difficult to prepare and run without experience
- Considered less rigorous or valid than quantitative data by some

Surveys

- A web form (or rarely, paper) with questions on players' attitudes and experiences
- May include open text questions

Pros

- Relatively easy to deploy
- Can be applied to many topics
- Standard experience measures already exist

- Statistics knowledge is required to interpret quantitative results
- Larger samples (N>100) are typically needed to test hypotheses

Heuristic Evaluation

 One or (ideally) more researchers examine a game and evaluate its compliance with recognised design principles ('heuristics')

Pros

- Comparatively fast and cheap to implement
- Maximises value from more expensive methods
- Comparing your game to conventional design is useful even if you disagree

- Less comprehensive than they seem on paper
- Effectively interpreting and applying heuristics requires some design expertise
- Rigid application of 'authoritative' design principles can rob games of unique qualities

Limit These Types of Questions

- Questions with no actionable findings
- Questions likely to produce ambiguous answers

• Sometimes useful (trust your gut), but resources are finite

DATA \longrightarrow INFORMATION \longrightarrow INSIGHT \longrightarrow ACTION



- Instead of:
 - "How often do players use the sniper rifle?"
- Ask:
 - "How often do sniper rifles win matchups against other weapons?"
- Or:
 - "How often are sniper rifles used in unfavourable matchups?"



Be Specific

GUR criteria

.

Criteria	Description
Representative	Selected methods and recruited participants must correctly reflect user testing needs and outcomes.
Accurate	Results should reflect user testing assumptions and include multiple sources of supporting data.
Specific	Methods selected for conducting the test need to deliver precise and specific results. For example, they cannot state that a game is not good without indicating why or identifying the problems.
Timely	User test findings should be delivered in a timeframe that matches the game development cycle.

GUR criteria

Criteria	Description
Cost-effective	There must be a return on investment or value added to a game that justifies the cost of conducting user tests.
Actionable	Results need to be delivered in an actionable and applicable format. The quality of results is directly affected by the chosen methods and analysis approaches.
Motivational	Presented results should motivate game developers to take action on them. Game developers should believe in and fully understand the results.

Types of GUR Questions
What are you evaluating?



Concept Test

- Which art style do players prefer?
- Which features would players prefer?
- Why do players play games in a certain genre?
- Why did players (not) make IAPs?
- What do players expect a certain object to do?

Competitor Analysis

- Which game has the best onboarding experience and why?
- What is the best way of doing notifications (for retention)?
- What is the best way to present an item in the store?
- What is the best way to signal to the player that the game has more depth in future levels?

Usability Analysis

- Does the game convey how to play the game?
- Is feedback clear?
- Usability evaluation before involving actual players

Usability Playtest

- Are the controls suitable for the target audience?
- Do users navigate through the game as intended?
- Is the onboarding effective?
- Involves players

Large-scale Playtesting • Do players enjoy the game?

• Does the game evoke the intended player experience?

Exercise

 You are the only UX researcher at a small independent game studio. The studio's current project is a single-player shooter with 20 short levels. The team has approached you to conduct some **last-minute playtesting** before demoing the first 5 levels of the game at a convention. Both PC and Xbox builds of the game are available. You only have 2 days to work from start to finish, and the team expects a brief report with your findings. After receiving your report, the team will have a week to make adjustments to the game.

Discussion

• What are *your* challenges? What questions do you want to answer via GUR?

Observation

- A researcher observes a tester's play behaviour while taking notes
- Ideally, a second researcher facilitates the session, and the observer is elsewhere (e.g., watching on a separate screen)

Pros

Observation

- See how 'typical' players actually behave in the game
- Can be done early in development
- Can provide specific answers to some questions

Cons

- Players are rarely as honest as you might like
- Presence of observers can bias results
- Behaviour requires interpretation
- Can be time-consuming
- Outliers may be misinterpreted as trends





• Difficulty

© 2007 Valve Corporation. All Rights Reserved.

Direct Observation

- "Typical" playtest
 - Watch people play the game
 - Observe their gameplay/behavior
 - Simulate at-home experience
- Have a design goal



Direct Observation







VALVE Durphing a Cood Playtact

Running a Good Playtest

Make sure the people responsible for the design and execution are there

- Simplifies evaluation
- Prioritizes
- Motivates
- Simulate the player "in their living room"
 - Don't give them hints
 - Don't answer any questions
 - Don't provide extrinsic rewards

Use external playtesters



© 2007 Valve Corporation. All Rights Reserved.

Roles

- Moderator
 - The **only** person who talks to testers
 - Makes sure testers are comfortable
 - Provides necessary information before testing (and during play – only if testers are **really** stuck)

Moderation

- All the setup (NDAs, consent forms, pre-test survey) should occur before testing
- Always have a script
- Introduce yourself, and anyone else involved in the test
- Small talk; help testers relax in the unusual environment
- Explain the game's basic premise (e.g., genre, any contextual info players would normally have)
- Reminds testers of their expert status (**not** the developers)
- Debriefs afterwards, answers any remaining questions

Observes the tester playing through the game

- Takes detailed notes w.r.t. the design goal
 - Other topics are de-prioritised, but may be possible to fit in
- 'Observation' of:
 - In-game behaviours
 - Facial expressions
 - Exclamations / remarks
 - Emotional reactions
 - Surprising moments
- May have a video record for later review (time-intensive)

Observer

Exercise

- Groups of 3: moderator, observer, tester
- All: Select the game to be tested
 - Your own design, or something from itch.io
 - Discuss design goal(s)
 - Write a report of the aggregate findings (afterwards)
- Moderator:
 - Prepare a script
- Observer:
 - Set up note-taking materials
- Rotate between each role

Heuristic Evaluation

Heuristic Evaluation

 One or (ideally) more researchers examine a game and evaluate its compliance with recognised design principles ('heuristics')

Pros

- Comparatively fast and cheap to implement
- Maximises value from more expensive methods
- Comparing your game to conventional design is useful even if you disagree

Cons

- Less comprehensive than they seem on paper
- Effectively interpreting and applying heuristics requires some design expertise
- Rigid application of 'authoritative' design principles can rob games of unique qualities

Usability vs. Playability

- **Usability:** interactions are effective, efficient, and satisfying (ISO 9241-11 standard)
- **Playability:** the interface is unobtrusive, design intent is clear, and the game is suitably difficult and engaging (Korhonen, 2016)

UX Usability Goals: Productivity			PX Playability Goals: Entertainment			
1.	Task completion	1.	Entertainment			
2.	Eliminate errors	2.	Fun to beat obstacles			
3.	External reward	3.	Intrinsic reward			
4.	Outcome-based rewards	4.	Process is its own reward			
5.	Intuitive	5.	New things to learn			
6.	Reduce workload	6.	Increase workload			
7.	Assumes technology need to be	7.	Assumes humans need to be			
	humanized		challenged			

	Code	Game Usability Heuristics
	GU1a	audiovisual representation supports the game
	GU1b	a view to the game world supports smooth interaction and the camera behaves correctly
	GU2	screen layout is efficient and visually pleasing
	GU3	device UI and game UI are used for their own purposes
Game	GU4	indicators are visible
Usability	GU5	the player understands the terminology
Heuristics	GU6	navigation is consistent, logical, and minimalist
	GU7	game controllers are consistent and follow standard conventions
	GU8	game controls are convenient and flexible
	GU9	the game gives feedback on the player's actions
	GU10	the player cannot make irreversible errors
	GU11	the player does not have to memorise things unnecessarily
	GU12	the game contains help

	Code	Gameplay Heuristics		
	GP1	the game provides clear goals or supports player-created goals		
	GP2	the player sees the progress in the game and can compare the results		
	GP3	the players are rewarded and the rewards are meaningful		
	GP4	the player is in control		
	GP5	challenge, strategy, and pace are in balance		
Gameplay	GP6	the first-time experience is encouraging		
Heuristics	GP7	the game story, if any, supports the gameplay and is meaningful		
(Playability)	GP8	there are no repetitive or boring tasks		
	GP9	the players can express themselves		
	GP10	the game supports different playing styles		
	GP11	the game does not stagnate		
	GP12	the game is consistent		
	GP13	the game uses orthogonal unit differentiation		
	GP14	the player does not lose any hard-won possessions		

Heuristic Evaluation

- No users required in Heuristic Evaluation. Instead, do the initial evaluation yourself, and then combine your results with a team.
- Different evaluators find different problems, with diminishing returns after 5-6 evaluators:



- Form groups of 4
- Agree on a game to evaluate
 - You may need to limit the evaluation to a particular game segment to make this manageable
- In pairs, play through your chosen game (segment), with one person noting any heuristic violations
 - Note all violations even if they're not actually problems
- Discuss your individual findings and decide on the most important issues
- Write up your findings and design recommendations in a report

Application

Heuristic Evaluation Report

L Io	Jnique dentifier	Unique Screen ID	Problem Description	Evidence – Heuristic Violated	Severity (1-5, 5 is high)	Frequency (1-5, 5 is common)	Proposed Solution	
JI	K_1	1.4.A	Required fields during signup are not obvious	Visibility of System Status	3	4	Indicate required fields with a red asterisk (*)	
JI	K_2	1.4.B	Errors during form validation are not uniquely defined	Help users Recognize, Diagnose and Recover from Errors	3	4	Use a verbose description of changes that have to occur when alerting the user of an error	
V								
Us th of ru Joi nu	Use a unique identifier that combines the initials of the evaluator with a running number tally (JK = Jon Kolko, 1 = incident number one)			dentify the neuristic that is iolated	Define a severity and frequency rating to indicate the relative impact of the critical incident and the number of times this incident is likely to be identified by a user			

Interviews

Interviews

• A researcher asks individual players about the topic(s) of interest, while taking notes and recording audio

Pros

- Generates contextual data that can stand alone, or explain other results
- Follow-up questions can help generate new insights
- An audio record provides a full account for later reference

Cons

- Researcher bias is not always obvious
- Can be difficult to prepare and run without experience
- Considered less rigorous or valid than quantitative data by some

- Structured
 - Fully scripted, with no deviations from the question list
 - Guaranteed to probe a particular topic of interest
 - Unusual for scripted questions to be comprehensive
- Semi-structured
 - Scripted question list, with room to deviate / follow-up
 - Flexible (and personally preferred)
- Unstructured
 - No script (beyond an initial question)
 - Largely directed by the participant's interests
 - Not terribly useful for testing a specific research question

Interview Formats

Interview[®] Timing

- Post-play
 - Most common time
 - Can follow up on notes from observation
- Memories of the experience (and related game elements) are most salient
- During play / review of in-game footage (contextual enquiry)
 - Not particularly common for games
 - Investigate strategic decisions, metagaming, 'obvious' behaviours

Other Notes

- It's not a conversation
 - Establish credibility without showing off
 - Don't talk about yourself
- Developing a rapport with participants is essential
- Keep game screenshots / video / the game itself available for reference
 - Participants will sometimes want to *show* what they mean via play

Writing Questions

- Leading questions
 - A question that suggests a possible answer
 - Bad: 'Were you furious when he said that?'
 - Good: 'How did you react when he said that?'
- Double questions
 - 'How old were you when that happened and what effect did it have on you?'
- Abstract language

Writing Questions

- Ground mapping questions
 - Widely framed
 - Open up a topic
 - Followed up with probing questions
 - 'Have you ever...'
 - 'What did you notice most...'

Writing Questions

- Perspective-widening questions
 - Going further into a topic, in a way that suits the research question
 - 'You've said ..., but was there anything you didn't like about [topic]?'
 - 'Are there other cases where your decision would be different?'
Writing Questions

- Content mining (follow-up questions)
 - Get a more complete description
 - Understanding why participants hold a belief or attitude
 - 'Can you tell me a little more about...'
 - 'What gave you that impression?'
 - 'Can you give me an example...'
 - 'What makes you say that?'
 - 'Can you explain why...'
 - 'What do you mean when you say...'

AT: this first part of the interview is about your experience playing mark of the ninja in the experimental play session. we're interested in knowing what's most important to your experience of the game. so first, i'd like to know what kinds of things you noticed as you were playing.

P19: the kinds of things.. just any kind of things?

AT: anything.

Broad Opening Questions

P19: i guess.. i liked the graphics, i don't know, like visually it was guite cool, like, appealing. i.. noticed.. [pause] i don't know, it took me a little while to get the hang of it, so like, figuring out, like, where they could see you, where they couldn't see you, like, you know .. how easy it was -- you know, if you pressed the grappling hook -- how easy it was -- it just latched on, you don't have to aim it or anything like that like that was guite interesting. it was kind of cool that you had a little guide lady.. to start with. and that you could -- you know, you don't have to just sit and read instructions, you can actually just play ... without .. having to, like scroll through stuff, you could -- she just tells you what your goal, which was guite cool. although i was -- i didn't really.. get what the plan was, i was just kind of like, "oh i'm just doing this", but i didn't really realise until like halfway through [unintelligible] when they're like "hey, hey, save this dude, and these other ninjas, you need to free them as well". so.. to start with, i was just getting really excited that i was doing it, and that was kind of figuring out how to use the controller, and then.. eventually i was like "hang on, what am i doing this for, like what's the end goal here, i'm just going through all these little hatches".

AT: how did you feel when you were playing?

P2: how did i feel when i was playing?

AT: mmhmm.

P2: i felt like i was having fun. i was joyful, but with intermittent moments of frustration, if i like fumbled something or didn't understand an instruction and later figured it out, and that was like "uh" because it took a step back from the competency thing i was talking about earlier, where it's just like i didn't feel like i was looking up to the character anymore, and i felt just as bad as everyone else.

AT: so, by kind of the inverse, then.. is the thing that made you feel best about the game the embodiment of the character -- or the correct embodiment of the character?

P2: yes. yes.

AT: ok.

P2: well i mean, i say that because i watched the cinematics, i decided not to skip them, so i learned how this person was making a sacrifice and they were this, they were being affected by this supernatural alchemical substance, and i felt like by playing into that, i'd be rewarded with all the context around the mechanics, and part of that was feeling competent. so when i felt like i wasn't portraying that character properly, it had a negative impact.

Follow-up Questions – Leading(?)

AT: so.. mark of the ninja requires you to do different things.

P8: well you kind of need to interact with the environment and stuff in mark of the ninja to be effective.

AT: ok. in what ways do you mean?

P8: like, obviously if there's an enemy blocking your path, or looking directly at you, you need to cause some kind of distraction before you can walk up to them.

AT: ok. you said that you really enjoyed having a non-violent option. what about that option appeals to you?

P8: it's definitely a bit more challenging, and i feel more rewarding at the end of the game.

AT: ok, so why is non-violence more challenging?

P8: i mean the enemies are going to attack you regardless, i mean it's easier to just cause a distraction and walk up behind them and press X, but with the nonviolence, you have to find ways of getting past them without noticing, or anything like that.

Follow-up Questions

AT: ok. you mentioned that you felt like an idiot when you forgot the controls for, you know, an extended period of time--

P5: [laughs]

AT: --did that just happen the one time, or was that --

P5: --oh it happened a lot, because i don't usually use the controller. so.. and it didn't seem very -- i have a very particular way i set up a controller, and it wasn't super intuitive to me, because i use other things -- other things were very intuitive, but my main issue was the A button. i would prefer to switch the A button to the run button, because i use A to run in literally every other game i play with a controller. and R could have been the open a door, or something like that. or X, i can't even remember what X did [pause] X was kill--

Interruptions + Limits of Audio

AT: --it's like your attack, yeah--

P5: yeah yeah yeah.

AT: ok, so -- so it does display the controller on the screen though, right?

P5: yeah, yeah.

AT: but that wasn't helpful, or..

P5: no it was helpful, like it lets me know, but my motor skills are tuned to something else. so it was -- i wouldn't necessarily call it the game's fault, i would call it.. i wouldn't even call it anyone's fault, i think -- i wouldn't even call it peer game design, but i would just call it "that i wasn't used to using the controller". usually i'm like this, not like this.

AT: for the audio recording's sake, he's miming using a keyboard and mouse, as opposed to a controller.

AT: ok. and that feels.. i mean, how does -- how does that feel?

P13: i guess, like, going back on the whole like role-play of a ninja, like it feels like, you're kind of engaging with like.. yeah like the whole precision of it? like, it's -- it lets you like [unintelligible], it's like a split second, but you've done these multiple precise actions, and you've got like the skillset of -- like, you do everything perfectly? kind of there. yeah.

AT: ok. so how does that kind of tie into what you were saying about wanting to like, sneak around more, or take the pacifism route?

P13: to kind of.. fit into.. the.. like, the world of it, right? so, to kind of.. because like, although -- god, i really sound like a weeb when i keep saying "ninja", like--

AT: [laughs]

P13: --going back to the whole ninja thing, like yeah you do a lot of killing, but also, there's an approach to it where you -- say you have the ability to infiltrate a place but not leave any traces whatsoever, like no-one would ever know that you were there, but even though you've gone in, you've taken a bunch of things, like, you've done stuff but like no-one would ever find out until like -- let's say you killed someone important, they won't find out until the next morning, stuff like that.

AT: ok. so how did you feel, like -- i kind of got the impression that you didn't always succeed at that.

P13: yeah.

AT: so how did that feel?

P13: definitely frustrating. i didn't feel too challenged in the way that you can -you can always rectify your mistake. so let's say if you get spotted, you can just run up and smack the dude a few times. but yeah, it's definitely frustrating, not in the sense that "oh like, oh, i lost the game", but more of like, "damn it, i like, i failed at my own code of trying to be like this perfect ninja".

Repetition for Depth

Application

- Form groups of 3 (moderator, note-taker, participant)
- Collectively, write a short interview script (~3-5 questions)
 - Topic: most recent play experience (outside this course)
 - Plan potential follow-up questions
- Interview each other, rotating between roles
- Compile your group's results into a single document
 - To be analysed later

Data Interpretation & Consolidation

Interpretation Session

- A group meeting consisting of the interviewer plus 2–5 team members
- Is conducted shortly (within 48 hours) after the interview
- As a rule: Lasts approximately the same amount of time as the (Contextual) Interview or inquiry
- Procedure:
 - The interviewer tells the story of the interview
 - Team members ask questions, drawing out details that might have been overlooked and indicate what is important to capture
 - At least person writes affinity notes, the others for example capture design models

- Key practice issues
- Identity and cultural observations
- Tool and activity successes and breakdowns
- Task patterns
- The use of time, place and different devices
- Design ideas
- And any other issues that have relevance to the project



Affinity Notes

Affinity Diagram The process of consolidation is easiest to see in building the Affinity Diagram, but it is similar for all models.

- Grouping the individual affinity notes into a wallsized, hierarchical diagram
- Shows the common issues, themes, and scope of the customer problems and needs in one place
- Acts as the voice of the customer and tells the story of the user's life



Affinity Diagram

	We plan our trip together			
We share the job of researching where to go	We plan the trip as a group	I take responsibility for booking all or just part of the trip		
T01-26 After a conversation about the pros and cons of Victoria and Vancouver, decided Victoria would not work out and returned to their original plan to go to Vancouver	T01-45 The closest friends-in the core team do all the planning and define the date. The second tier (people invited often by core team to come) get added to the email chain to work out details of when to arrive, where to stay, and overall logistics after the date is set	T05-34 Because the AirBnB profile contains her boyfriend's email and personal info, he does most of the research and all of the contact with the owners through the site.		
T01-24 Over several days researched Victoria compared to Vancouver. They emailed each other from work with additional details and links and called each other on the phone after emails to discuss. The pattern of research, share, and talk was repeated when	T01-62 Emails his friends to see if they want to do a ride on the last morning of the trip because if they do he will take the later flight, otherwise he will book an earlier flight that gets	T05-48 Boyfriend had to stay in constand contact with AirBnB owners (using AirBnB website messaging) to make sure they'd have a place to stay in each of the cities they were visiting.		
they were not co-located.	home at a better time.	T01-27 After they decided to		

1. Interpretation Session

• Every interviewer tells the story of the interview, one after the other.

Team members ask questions about the interview, drawing out details that the interviewer might have overlooked and indicate what is important to capture

• Team Members write Affinity Notes on sticky notes

Application 1

Application 2

- 2. Data Consolidation
- Group the sticky notes on the wall, each grouping describes a single issue or a point
 - Keep the groups small, four to six notes in a group
- Label the groups with blue sticky notes to characterize the point made by the group
- Organize the blue labels into larger areas of interest under pink labels

Surveys

Surveys

- A web form (or rarely, paper) with questions on players' attitudes and experiences
- May include open text questions

Pros

- Relatively easy to deploy
- Can be applied to many topics
- Standard experience measures already exist
- Little training necessary
- Efficient, low-cost

Cons

- Statistics knowledge is required to interpret quantitative results
- Larger samples (N>100) are typically needed to test hypotheses

The Importance of Good (Quant) Measures "We've got thousands of game designers in Australia. No problem at all, but we have very, very few experienced project managers, and that's meant most of the attempts [to adapt to post-GFC industry changes] have fizzled out, because if you think of the build-measure-learn cycle, we built, *we didn't quite know what we were measuring, and we learned nothing*."

-George Fidler (Kixeye), 2014 interview with John Banks

Survey Question Types

Open-ended questions

Qualitative analysis

*Try to describe this particular emotionally moving experience as accurately and detailed as you remember, and try to be as concrete as possible. You can use as many sentences as you like, so we can easily understand why this was such an emotionally moving experience for you.

- Closed questions
 - Questionnaires, rating scales, etc
 - Quantitative Analysis

Single-choice questions

How long ago did this experience take place?								
		Choose one of the second se						
<2 months ago	2-11 months ago	1-2 years ago	2-10 years ago	>10 years ago				

Multiple-choice questions

MPAA But of a second of the s

Choose all that apply.

Ranking Questions

Please rank these Game genres from your favorite to your least favorite, 1 = favorite, 5 = least favorite. Please rank these Game genres from your favorite to your least favorite, 1 = favorite, 5 = least favorite. Please rank the questions by clicking on them in order.

Action Game	
Adventure Game	1
Strategy Game	3
Simulation Game	
Role-Playing Game	2

Thinking about your most recent Dark Souls III game-session, please indicate to what extend you agree with each of the following statements.

Thinking about your most recent Dark Souls III game-session, please indicate to what extend you agree with each of the following statements.

	0	0	0	0	0	0	0
	0	0	0	0	0	0	0
	\bigcirc						
	0	0	0	0	0	0	0

Closed Question Types



Rating Questions: Likert Scale ->

Thinking about your most recent Dark Souls III game-session, please indicate to what extend you agree with each of the following statements. Please rate these statements from 1 (strongly disagree) to 7 (strongly agree).

	Strongly disagree			(4)	(5)		Strongly agree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
The game provided me with interesting options and choice	s. 🔘	\odot	0	0	0	\odot	\odot
I felt very capable and effective when playing.	0	\bigcirc	\bigcirc	0	0	0	0
When I accomplished something in the game I experienced genuine pride.	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc
Learning the game controls was easy.	0	0	0	0	0	0	0
I could always find something interesting in the game to do.	0	0	0	0	0	0	\bigcirc
I didn't feel close to other players.	0	0	0	0	Ô	0	0

Closed Question Types

Semantic Differential ->



Survey Question Types Strengths

- Open-ended questions
 - Provides nuanced insights into player experience, not available via quantitative measures
 - **Closed** questions

•

• Quantitative/quantified responses tend to be very persuasive for different stakeholders

• Self-reports of...

What can be asked with surveys?

- Objective facts: Behaviors and facts that are observable
 -> but often not representative of actual behavior
- Subjective states: Feelings, attitudes, thoughts -> Experience

What can you do with surveys?

- Have participants self-report about different aspects of their experience:
 - Attitudes, feelings and experiences: Benchmark, compare subjective experience with objective (behavioral) measures
- Motives: Why are people playing a game at a given time?

What can you do with surveys?

- Personal characteristics: Personality traits, gaming habits, familiarity with design, etc
 - Can be used to create groups post-hoc
- Comparison: Questionnaire data can be used for comparing between designs, groups of people, over time

Pros:

- Focus on user demographics, opinions and motivations
- Reach a large number of people
- Lower bias due to respondent anonymity

Cons:

- Self-reports of behaviour unlikely to be accurate
- Can't easily ask follow-up questions
- Lower hurdle for participants to drop out

When to use online surveys?



• E.g., Platforms such as Google Forms, Webropol, Limesurvey (my recommendation, more flexible and GDPR compliant)

Recruit

participants

Designing a survey: Example structure

- Study information: goal, topic, approx. time to complete, contact person
- Informed consent
- Demographic questions
- Main survey questions
- Debriefing

Choosing survey questions

- What are the goals of the survey? What do you want to find out?
- What types of information do you need to collect from players?

Choosing survey questions

- Draft your own questions
 - Or use questions from other existing surveys
- But: Just because it was used before doesn't mean it is a good question

Good vs Bad Survey Questions

Open Questions • DON'T: "Do you like the game?"

- DO: "Which parts of the game do you like in particular?"
- More open-ended question \rightarrow more useful answers

Avoid doublebarreled questions • DON'T: "I considered the level fun and exciting"

• DO: "I considered the level fun"; "I considered the level exciting"

Make it simple and clear

• Simple questions and wording are easier to understand

With what frequency has your child experienced a raised temperature within the last 30 days?

How many times has your child had a fever in the last month?



Make it simple and clear • DON'T: "To what extent did you experience a sense of flow?"

- DO: "To what extent were you concentrated on playing?"
- Create questions that players have enough information and knowledge to actually respond to

Avoid leading questions

- DON'T: "This game has a Metacritic score of 93. How much fun did you experience with the game?"
- DO: "How fun was it to play the game?"

Other questions to avoid: Positivity Bias

How is YouTube today? $\qquad \qquad \qquad$							
Absolut	Absolutely outstanding						
Extreme	Extremely good						
Very go	od						
Good							
Not goo	d						
		_	_				
Home	Trending	Subscriptions	Inbox	Library			
youtube is so insecure they only give 1 negative option version 4							

positive options 😭
Other questions to avoid

- What players would do / like / want in hypothetical scenarios, e.g., "Over the next month, how frequently will you access the PlayStation store?"; "Which of the following features would make you have more fun with this game?"
- How often players do things (better to ask how many hours they played this week; more concrete and more recent)

Asking about gender: DON'Ts

Gender *

) Female

) Male

Rather not to say

10:36		'III \$ ■)
AA	aaltobusiness.qualtrics.com	C
0%		100%

Firstly we would like to know a little about you as a player.

Are you a beautiful or a handsome?



 <u>Open answer question</u> about gender, so participants can self-identify

Asking about gender: DOs

• "What is your gender: _____"

• "How old are you: ____"

No assumptions made

Bad Survey Practices

	 How long have you been using our software? Less than a month A year
	2. What is most important to you in customer service? Patience Chat service Speed and reaction Other
What did you think of this piece?	 3. Are you satisfied with our customer service? Yes No
Great!	4. What would you improve in out software?
TNOT What I needed	 5. Would you recommend Active Trail to your friends? absolutely! Yes Maybe No
	Submit

Bad Survey Practices

How are you feeling about the coronavirus?





Source (and more great bad examples): https://twitter.com/badsurveyg



Order of the questions

- Survey is similar to conversations
- Start with easy to respond questions
- Group questions of the same topic together
- Ask potentially sensitive questions later in the survey

Things to keep in mind

- How long do participants need to complete the survey?
- Do participants understand the questions?
 - How is the flow of the survey?
- Pilot the survey: with colleagues, friends, potential participants

Sources for creating unbiased surveys

Survey Research in HCI

Hendrik Müller, Aaron Sedley, and Elizabeth Ferrall-Nunge

https://doi.org/10.1007/978-1-4939-0378-8_10

CHAPTER 9

Surveys in Games User Research

FLORIAN BRÜHLMANN, University of Basel

ELISA MEKLER, University of Basel

Highlights

Surveys are an essential method of data collection that can deliver generalizable and actionable insights about the player's experience. In this chapter, we present practice-oriented guidance about when the method is appropriate, what constitutes a good questionnaire, and how to alleviate possible biases and issues with data quality.

<u>https://psyarxiv.com/2csa4/d</u> <u>ownload?format=pdf</u>

Example Surveys

The Appeal of MOBA Games: What Makes People Start, Stay, and Stop

April Tyack Queensland University of Technology (QUT) Brisbane, Australia april.tyack@hdr.qut.edu.au Peta Wyeth Queensland University of Technology (QUT) Brisbane, Australia peta.wyeth@qut.edu.au **Daniel Johnson**

Queensland University of Technology (QUT) Brisbane, Australia dm.johnson@qut.edu.au Aim of the Survey

- RQ1: What draws people to MOBAs?
- RQ2: What motivates people to keep playing MOBAs?
 2a: How does experience change when playing in teams of friends, strangers, or a combination of the two groups?
 2b: How do players develop in-game friendships?
- RQ3: What causes people to stop playing MOBAs?
 3a: What factors affect MOBA players' churn rates?
 3b: What factors can affect players' reasons for returning to MOBAs after long periods of absence?

Survey

- 760 survey responses, age 18 40, 84% men, majority of participants recruited on Reddit, Facebook groups, Twitter
- Survey contained single-choice, multiple-choice, and rating questions
- Originally 924 responses -> some had to be removed because incomplete responses were provided, or participants provided bogus answers (e.g., answering 1 for all rating questions)

What pl	layers va	lue w	hen	р	laying
with str	angers v	s frier	nds		

	Strangers		Friends			
Variable	Mean	SD	Mean	SD	Z	r
Positive Attitude	4.07	1 09	4 21	0.99	-3.46*	0.10
Skilled Play	3.61	0.98	3.08	1.18	11.55**	0.32
Enjoyable	2.35	1.30	3.97	1.11	-20.66**	0.56
Complementary Roles/Characters	3.39	1.14	2.96	1.30	8.34**	0.23
Communication and Coordination	3.89	1.03	3.96	1.06	-1.63	0.04
*n < 0.01 $** < 0.01$	001					

Mood when playing w strangers vs friends

p < 0.01,	** < 0.001
-----------	-------------------

	Strang	gers	Friends		_	
Variable	Mean	SD	Mean	SD	Z	r
Tension	1.89	0.90	1.57	0.76	-9.74*	0.33
Vigour	3.00	0.89	3.34	0.89	7.12*	0.24
Confusion	1.95	0.84	1.51	0.64	-13.69*	0.47
Fatigue	1.89	0.83	1.59	0.72	-11.22*	0.38
Depression	1.79	0.86	1.39	0.65	-12.71*	0.43
Anger	2.35	1.07	1.74	0.81	-13.63*	0.47
.b < 0.001						

Excerpt Results RQ2,

4.78

9.13

20 6.96

59.13

13.19

19.86

10.58

3.76

-47.39



ORIGINAL RESEARCH published: 22 July 2020 doi: 10.3389/fpsyg.2020.01307



Motivational Profiling of League of Legends Players

Florian Brühlmann^{1*}, Philipp Baumgartner¹, Günter Wallner^{2,3,4}, Simone Kriglstein^{5,6} and Elisa D. Mekler⁷

¹ Human-Computer Interaction Research Group, Department of Psychology, Center for General Psychology and Methodology, University of Basel, Basel, Switzerland, ² Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands, ³ Faculty of Business and IT, Ontario Tech University, Oshawa, ON, Canada, ⁴ Institute of Art and Technology, University of Applied Arts Vienna, Vienna, Austria, ⁵ AIT Austrian Institute of Technology GmbH, Vienna, Austria, ⁶ Faculty of Computer Science, University of Vienna, Vienna, Austria, ⁷ Department of Computer Science, Aalto University, Espoo, Finland Aim of the Survey

- Explore how players' motivation relate to their experience and in-game behaviour in League of Legends
- Behavioural data from League of Legends API
- Rating questions from established psychological questionnaires about motivation and player experience

Type of Motivation	Amotivation	Extrinsic Motivation				Intrinsic Motivation
Type of Regulation	Non- regulation (AMO)	External Regulation (EXT)	Introjected Regulation (INJ)	Identified Regulation (IDE)	Integrated Regulation (INT)	Intrinsic Regulation (IMO)

FIGURE 1 | The six types of motivational regulation as posited by Self-Determination Theory. Ranging from the least self-determined (amotivation) to the most self-determined regulation (intrinsic motivation). Figure adapted from Deci and Ryan (2002), p. 16.

Survey

- 750 participants, majority male players, age 18 65, recruited on League of Legends Subreddit
- Several GB of behavioural data from API using Riot-Watcher
 - Over a million individual matches
- The two data sets were matched based on participants' Summoner name

Results

- Latent Profile Analysis to identify patterns in self-reported motivational data
- Intrinsic motivation high overall, but still distinct motivational profiles



Profiles 🚔 Amotivated 늘 External intrinsic 🚊 Autonomous 5. 7 6-Likert-type scale [1–7] ГО I Likert-type scale [1 2-1 -1 ENJ TENS AUT COM REL MastAp MastAv PerfAp PerfAv ΗP OP NA PA IMI PENS ACH_GOAL PANAS PASSION VITALITY

• Distinct player experience based on motivational profile

Results



• Few obvious behavioural differences



Results

 Intrinsic and autonomous motivation groups more likely to assist than amotivated players



ranked*

kda

unranked

Autonomous



Quantitative Questionnaires

Some scientific terminology

- Concept: Description of some specific phenomenon, e.g., "Experience", "Fun", "Challenge", "Flow"
- Operationalization: Turning concepts into something that we can ask about or measure

Some scientific terminology

- Construct: A concept that is operationalized into a questionnaire, i.e., a dimension of a questionnaire
- Item: Describes an individual question in a questionnaire, several items make up a construct
- E.g., construct **fun**
- -> Items may be "I find this game fun to play", "I find this game entertaining", "I had a good time playing this game"

 Construct Validity: Does the questionnaire measure what we think it should measure? Does it adequately approximate the felt subjective experience?

Questionnaire quality criteria

•

Discriminant validity: Do **unrelated** construct items **not** relate to each other?

Test via Factor Analysis (among other statistical analyses)

Questionnaire quality criteria

- Reliability: Is the questionnaire consistent?
 - Internal reliability: Are all items of a (intended) construct measuring the same?
- Test-retest reliability: Does the questionnaire work consistently over repeated use?

Questionnaire quality criteria



VALID BUT NOT RELIABLE

RELIABLE BUT NOT VALID





Questionnaire quality criteria

Should be clear and easy to understand

Easy to read

 Balanced: Enough items for maintaining reliability, not too many to exhaust research participants

Selecting a Questionnaire

Selecting a questionnaire

- It's perfectly legit to select an existing questionnaire
 - BUT: Some commonly used questionnaires may not work as well as advertised
- Pay attention to the scientific rigor with which the questionnaire has been tested

Common Questionnaires

- Player Experience Inventory (PXI)
 - dx.doi.org/10.1016/j.ijhcs.2019.102370
 - Subscales measure experiences (e.g., immersion) and game elements (e.g., ease of control)

Constructs & Items

MEANING:

Playing the game was meaningful to me. The game felt relevant to me.

Playing this game was valuable to me.

MASTERY:

I felt capable while playing the game.

I felt I was good at playing this game.

I felt a sense of mastery playing this game.

IMMERSION:

I was no longer aware of my surroundings while I was playing.

I was immersed in the game.

I was fully focused on the game.

AUTONOMY:

I felt a sense of freedom about how I wanted to play this game.

I felt free to play the game in my own way.

I felt like I had choices regarding how I wanted to play this game.

CURIOSITY:

I felt eager to discover how the game continued.

I wanted to explore how the game evolved.

I wanted to find out how the game progressed.

EASE OF CONTROL:

- I thought the game was easy to control.
 - The actions to control the game were clear to me.
 - It was easy to know how to perform actions in the game.

CHALLENGE:

The game was challenging but not too challenging. The game was not too easy and not too hard to play. The challenges in the game were at the right level of difficulty for me.

PROGRESS FEEDBACK:

The game gave clear feedback on my progress towards the goals.

I could easily assess how I was performing in the game. The game informed me of my progress in the game.

AUDIOVISUAL APPEAL:

I enjoyed the way the game was styled.

I liked the look and feel of the game.

I appreciated the aesthetics of the game.

GOALS AND RULES:

The goals of the game were clear to me.

I grasped the overall goal of the game.

I understood the objectives of the game.

Common Questionnaires

- Ubisoft Perceived Experience Questionnaire (UPEQ)
 - dx.doi.org/10.1145/3235765.3235780
 - Based on self-determination theory (SDT)
 - Competence, autonomy, and relatedness subscales
 - Relatedness toward other players *and* NPCs

Factor	Item				
	I was free to decide how I wanted to [play].		I really like the people I play with.		
	I could approach [the game] in my own way.		I consider players I regularly interact with to be my		
omy	The game allowed me to [play] the way I wanted to.		friends.		
Auton	I had important decisions to make when [playing].		Other players are friendly towards me.		
·	The choices I made while [playing] influenced what happened.	ness	What other players did in the game had an impact on my actions.		
	My actions had an impact on the [game].	lated	I had to adapt my actions to other players' actions.		
	With time, I became better at [playing].	Re			
	My [gaming] abilities have improved since the		I was paying attention to other players' actions.		
ence	beginning.		I felt close to some of the characters.		
Compete	My mastery of the [game] improved with practice.		I was bonding with some of the characters.		
	I was good at [playing].				
	I felt competent at [playing].		I cared about what happens to some of the characters.		
	I felt very capable and effective when [playing].				

Other (not game-specific) Questionnaires

- Positive and Negative Affect Schedule eXpanded (PANAS-X)
 - https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1011&context=psych ology_pubs
 - Standard emotional measure (used since 1999)
 - Overall positive / negative affect scores, or more specific emotions:
 - Joviality, self-assurance, attentiveness, surprise, serenity
 - Fear, sadness, hostility, guilt, shyness, fatigue

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you have felt this way during the past few weeks. Use the following scale to record your answers:

1	2	3	4	5
very slightly	a little	moderately	quite a bit	extremely
or not at all		-	-	
cheerful	sad	ac	ctive	angry at self
disgusted	calm	g	uilty	enthusiastic
attentive	afraid	jo	oyful	downhearted
bashful	tired	n	ervous	sheepish
sluggish	amazed	lo	onely	distressed
daring	shaky	sl	eepy	blameworthy
surprised	happy	ex	cited	determined
strong	timid	h	ostile	frightened
scornful	alone	pi	roud	astonished
relaxed	alert	ji	ttery	interested
irritable	upset	li	vely	loathing
delighted	angry	as	hamed	confident
inspired	bold	at	ease	energetic
fearless	blue	SC	ared	concentrating
disgusted	shy	dı	rowsy	dissatisfied
with self				with self

Other (not game-specific) Questionnaires Intrinsic Motivation Inventory

- Based on self-determination theory
- Long and short forms
- Independently validated*

THE POST-EXPERIMENTAL INTRINSIC MOTIVATION INVENTORY

(Below are listed all 45 items that can be used depending on which are needed.)

For each of the following statements, please indicate how true it is for you, using the following scale:

1 2 3 4 5 6 7 not at all somewhat very true true true

Interest/Enjoyment

I enjoyed doing this activity very much This activity was fun to do. I thought this was a boring activity. (R) This activity did not hold my attention at all.(R) I would describe this activity as very interesting. I thought this activity was quite enjoyable. While I was doing this activity, I was thinking about how much I enjoyed it.
Perceived Competence

I think I am pretty good at this activity. I think I did pretty well at this activity, compared to other students. After working at this activity for awhile, I felt pretty competent. I am satisfied with my performance at this task. I was pretty skilled at this activity. This was an activity that I couldnÕt do very well. (R)

(R)

Effort/Importance

I put a lot of effort into this. I didnÕt try very hard to do well at this activity. (R) I tried very hard on this activity. It was important to me to do well at this task. I didnÕt put much energy into this. (R)

Pressure/Tension

I did not feel nervous at all while doing this. I felt very tense while doing this activity. I was very relaxed in doing these. (R) I was anxious while working on this task.

I felt pressured while doing these.

Perceived Choice

I believe I had some choice about doing this activity. I felt like it was not my own choice to do this task. (R) I didnÕt really have a choice about doing this task. (R) I felt like I had to do this. (R) I did this activity because I had no choice. (R) I did this activity because I wanted to. I did this activity because I had to. (R)

Value/Usefulness

I believe this activity could be of some value to me. I think that doing this activity is useful for _______ I think this is important to do because it can ______ I would be willing to do this again because it has some value to me. I think doing this activity could help me to ______ I believe doing this activity could be beneficial to me. I think this is an important activity.

^Not validated

Other (not game-specific) Questionnaires • User Motivation Inventory (UMI)

- dx.doi.org/10.1145/3173574.3173680
- Based on SDT
- Forms of motivation (as in Elisa's MOBA paper)

	Subscale	Item
	Amotivation	 I use [X], but I question why I continue to use it I use [X], but I wonder what is the point in using it I use [X], but I don't see why I should keep on bothering with it
	External regulation	 Other people will be upset if I don't use [X] I use [X] because others will not be pleased with me if I don't I feel under pressure from others to use [X]
	Introjected regulation	 I would feel bad about myself if I quit [X] I would feel guilty if I quit using [X] I would feel like a failure if I quit using [X]
_	Identified regulation	 Using [X] is a sensible thing to do The benefits of using [X] are important to me Using [X] is a good way to achieve what I need right now
'Good' forms	Integrated regulation	 I use [X] because it reflects the essence of who I am Using [X] is consistent with my deepest principles I use [X] because it expresses my values
of motivation	Intrinsic motivation	 I use [X] because it is enjoyable I think using [X] is an interesting activity Using [X] is fun

Check your sources!!

9.5 Established questionnaires in GUR

An alternative to constructing a new questionnaire is to employ a well-established questionnaire. Ideally, these questionnaires have been previously validated, which allows researchers to compare the results to other studies that have used the questionnaire. An existing questionnaire can be adapted to the specific study context as needed; however, this reduces the comparability between different studies. CUR is a relatively new field in comparison to other disciplines:

therefore some questionnaires in GUR have not been extensively validated, and should be employed with caution (Brühlmann and Schmid, 2015). Some of the most commonly used GUR-related questionnaires are the following:

• Game Experience Questionnaire (GEQ). The GEQ by IJsselsteijn and colleagues (IJsselsteijn et al., 2008) incorporates seven different dimensions of player experience: sensory and imaginative immersion, tension, competence, flow, negative effect, positive effect, and challenge. The GEQ is a self-report measure for a rather multifaceted investigation of game experience and is yet to be validated.

Game Experience Questionnaire (GEQ): A cautionary tale Session: Paper Presentation

CHI PLAY 2018, October 28–31, 2018, Melbourne, VIC, Australia

Systematic Review and Validation of the Game Experience Questionnaire (GEQ) – Implications for Citation and Reporting Practice

Effie L.-C. Law¹, Florian Brühlmann², Elisa D. Mekler² ¹Department of Informatics, University of Leicester; ²Faculty of Psychology, University of Basel lcl9@leicester.ac.uk, florian.bruehlmann@unibas.ch, elisa.mekler@unibas.ch

2. Game Experience Questionnaire – Core Module

Please indicate how you felt while playing the game for each of the items,

on the following scale:

not at all	slightly	moderately	fairly	extremely		
0	1	2	3	4		
< >	< >	< >	< >	< >		

Game Experience Questionnaire (GEQ): Excerpt

1	I felt content
2	l felt skilful
3	I was interested in the game's story
4	I thought it was fun
5	I was fully occupied with the game
6	I felt happy
7	It gave me a bad mood
8	I thought about other things
9	I found it tiresome
10	I felt competent
11	I thought it was hard
12	It was aesthetically pleasing
13	I forgot everything around me
14	I felt good
15	l was good at it

Game Experience Questionnaire (GEQ)

- Intended to assess 7 distinct factors of the player experience: challenge, competence, flow, immersion, tension, positive and negative affect
 - Formal validation never published by original authors (Curran, 2013; Johnson et al., 2018; Norman, 2013)

Study Findings

- Online survey, n = 633 recruited via MTurk
- Original 7-factor structure could not be replicated
- Compromised reliability for negative affect and challenge

 Poor discriminant validity for negative affect, challenge, tension, immersion and competence

MR2 MR1 MR5 MR3 MR4 MR6 MR7 h2 Component 15 I was good at it Competence -.030 .060 -.026 .798 -.006 -.111 .030 .739 02 I felt skillful Competence -.043 .058 .063 .701 -.019 .240 -.051 .451 17 I felt successful Competence -.073 .096 .035 .604 .102 .011 -.092 .697 21 I was fast at reaching the game's targets Competence .114 .042 .122 .593 .005 -.090 .067 .399 10 I felt competent Competence .016 .170 .105 .491 .076 -.042 .115 .650 19 I felt that I could explore things Immersion -.025 -.074 .751 .004 .030 -.065 .032 .697 03 I was interested in the game's story Immersion .002 .050 .722 -.037 -.045 -.049 -.017 .668 .680 .023 18 I felt imaginative Immersion -.037 -.027 .068 .080 -.107 .426 27 I found it impressive Immersion .017 .177 .613 .025 .015 .111 .005 .546 30 It felt like a rich experience Immersion .037 .152 .489 .070 .136 .088 .115 .518 12 It was aesthetically pleasing .269 .337 .094 .009 .018 .521 Immersion -.050 .071 31 I lost connection with the outside world Flow .069 .047 -.010 -.070 .850 -.077 -.023 .529 13 I forgot everything around me .090 Flow -.086 -.128 .065 .110 .724 -.042 .631 25 I lost track of time Flow .086 .033 -.037 .696 -.023 .004 .598 .068 05 I was fully occupied with the game Flow .018 .069 .101 .150 .451 .140 .315 .685 28 I was deeply concentrated in the game Flow -.008 .159 .345 .240 .282 .639 .066 .171 24 I felt irritable Tension .806 .013 -.012 -.014 .054 -.011 -.009 .595 22 I felt annoved -.004 -.007 -.071.004 Tension .800 .048 -.015 .693 .089 .275 .780 29 I felt frustrated Tension .656 -.126 -.131 .035 -.013 23 I felt pressured Tension .393 -.111 -.044 .104 .088 .339 .046 .597 -.007 32 I felt time pressure Challenge .338 -.040 .000 .253 .041 .283 .119 11 I thought it was hard .005 .005 .001 .679 -.088 .583 Challenge .116 -.151 26 I felt challenged Challenge -.032 .134 .049 .081 .059 .661 .085 .658 33 I had to put a lot of effort in to it .107 .109 .034 .253 Challenge -.066 .137 .563 .094 07 It gave me a bad mood .782 -.074 .009 .059 .012 -.078 .075 .470 Negative Affect 09 I found it tiresome .505 -.043 Negative Affect -.240 .113 .063 .013 -.114 .692 16 I felt bored .455 -.213 .015 .089 -.043 -.208 -.212 .732 Negative Affect .102 08 I thought about other things Negative Affect .278 .011 .066 -.241 -.118 -.327 .467 .734 .096 .097 .049 .049 -.184 .401 06 I felt happy **Positive Affect** -.011 04 I thought it was fun Positive Affect -.041 .703 .111 - .006 -.012 .003 .273 .517 20 I enjoyed it Positive Affect -.073 .653 .095 .073 .020 .026 .184 .499 -.097 14 I felt good Positive Affect -.090 .607 .018 .213 .110 .049 .407 01 I felt content Positive Affect -.088 .575 .003 .217 .089 -.017 -.115 .596 After rotation Sums of Squares 3.17 3.70 3.03 2.90 2.612.07 1.44 9.7 11.2 9.2 8.8 7.9 4.4 % of variance explained 6.3

Factor structure

Factor	
structure:	
Close-up	

32 I felt time pressure	Challenge	.338	040	007	.119	.000	.253	.041	.283
11 I thought it was hard	Challenge	.116	.005	.005	151	.001	.679	.088	.583
26 I felt challenged	Challenge	032	.134	.049	.081	.059	.661	.085	.658
33 I had to put a lot of effort in to it	Challenge	.107	066	.137	.109	.034	.563	.094	.253
07 It gave me a bad mood	Negative Affect	.782	074	.009	.059	.012	078	.075	.470

- Not all "Challenge" items load on the same factor / construct
- Poor internal reliability

But don't do this!



Factor structure: Close-up

24 I felt irritable	Tension	.806	.013	012	014	.054	011	009	.595
22 I felt annoyed	Tension	.800	004	007	071	.004	.048	015	.693
29 I felt frustrated	Tension	.656	.089	126	131	.035	.275	013	.780
23 I felt pressured	Tension	.393	111	044	.104	.088	.339	.046	.597
32 I felt time pressure	Challenge	.338	040	007	.119	.000	.253	.041	.283
11 I thought it was hard Challer		.116	.005	.005	151	.001	.679	088	.583
26 I felt challenged	Challenge	032	.134	.049	.081	.059	.661	.085	.658
33 I had to put a lot of effort in to it	Challenge	.107	066	.137	.109	.034	.563	.094	.253
07 It gave me a bad mood	Negative Affect	.782	074	.009	.059	.012	078	.075	.470
09 I found it tiresome	Negative Affect	.505	240	.113	.063	043	.013	114	.692
16 I felt bored	Negative Affect	.455	213	.015	.089	043	208	212	.732
			~ · ·	· ~ -	~	- · ·			•

- Tension, negative affect and 1 item from Challenge load on same factor
- Poor discriminant validity

263

Best practices for validating questionnaires

Intl. Journal of Human-Computer Interaction, 32: 51-62, 2016 Copyright © Taylor & Francis Group, LLC ISSN: 1044-7318 print / 1532-7590 online DOI: 10.1080/10447318.2015.1087664



A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve?

Matt C. Howard Department of Psychology, The Pennsylvania State University, University Park, Pennsylvania

https://doi.org/10.1080/10447318.2015. 1087664 Introduction to Statistics for Questionnaires

- If you want to test if there are statistically significant differences between designs (Null Hypothesis Significance Testing)
- Very pragmatic, don't rely on it for academic/scientific projects! E.g., sample size

QUANTIFYING THE USER EXPERIENCE

PRACTICAL STATISTICS FOR USER RESEARCH

JEFF SAURO / JAMES R. LEWIS