

# Methods Beyond the Screen: Conducting remote player studies for game design research

**Alexander Muscat**

School of Design  
RMIT University  
Melbourne, Australia  
[alexander.muscat@rmit.edu.au](mailto:alexander.muscat@rmit.edu.au)

**Jonathan Duckworth, Douglas Wilson**

RMIT University  
Melbourne, Australia  
[jonathan.duckworth@rmit.edu.au](mailto:jonathan.duckworth@rmit.edu.au), [douglas.wilson@rmit.edu.au](mailto:douglas.wilson@rmit.edu.au)

## ABSTRACT

In this paper we present a qualitative research methodology for conducting scholarly remote player studies, derived from a comparison of player-testing protocols. Within the game design research field, approaches to studying designs are frequently adapted from standardized player-testing techniques. These often focus on measuring player experience so a design may be evaluated. While such methods provide a useful basis for conducting iterative design studies, these present limitations for researchers seeking to interrogate design approaches outside of conventional assessment models or gameplay paradigms. We discuss these issues through a methodological lens, in the study of *WORLD4*, a game designed for experiences of ambiguity. Through a two-stage player experience case study we reveal methodological considerations, insights, and highlight disciplinary questions. In doing so we present a contextually aware, time and resource conscious method for conducting remote player studies, useful for game design researchers working outside of labs or investigating alternate design spaces.

## Keywords

Game design, design, methods, qualitative research, player studies, player experience

## INTRODUCTION

Within the field of game design, the study of player experiences is an industry standard practice. Large-scale commercial studios utilize elaborate player testing or playtesting configurations and deploy a mix of methods: Player behaviors are observed through recording capture, interest levels noted through quantitative surveys, and are gauged through in-game analytics e.g. heat mapping (see Zammitto 2015). In some instances, physiological measurements are deployed to further scrutinize the experience, including heart rate, skin conductance, and eye tracking technologies (see Ambinder 2009).

Such resource-intensive approaches are beyond the reach of most game designers and game design researchers seeking to analyze their designs. Alternatively, game design theorists Fullerton et al. (2004), Fullerton (2008), and Isbister and Schaffer (2008) present formalized approaches more appropriate to smaller groups of designers, which outline qualitative methods for the study of player experiences. Adopting multiple concurrent data-capture methods, these primarily focus on the observation of player

Proceedings of DiGRA 2019

© 2019 Authors & Digital Games Research Association DiGRA. Personal and educational classroom use of this paper is allowed, commercial use requires specific permission from the author.

experience and interviews to ensure a richer ‘picture’ of the characteristics and issues around the game.

These player testing methods share a significant overlap with scholarly approaches in broader games research fields, concentrating on the study of play and players. Historically derived from social science and Human Computer Interaction fields, approaches in capturing the multiple dimensions of player experience methodologically vary, including observational, interview, and diary techniques (Ribbens and Poels 2009), a mixing of qualitative and quantitative methods (Brockmyer et al. 2009; Llanos and Jørgensen 2011; Boellstorff et al. 2012; Lankoski and Björk 2015), analytics (Marczak et al. 2015) and use of physiological measures (Schott and Marczak 2014; Nacke and Lindley 2008).

However, these academically valid approaches present a dilemma for game design researchers seeking to study their designs. Although less resource-intensive than large-scale industry approaches the availability of resources is often assumed, for example, conducting research in an organized team and having a dedicated study environment (e.g. “game experience lab” (Ribbens and Poels 2009)). For researchers working alone, outside of a lab environment, the requirements of such methods raise question to their viability if adopted; a straightforward task like securing a dedicated space or finding an appropriate number of participants may prove significantly more difficult and have greater run-on-effects in planning. For design researchers creating original games this is compounded by time-intensive factors related to game development and production.

Such methods present additional challenges for game design researchers investigating styles whose experience falls outside of well-established genre categories, such as walking simulators (Muscat et al. 2016). Formal, standardized approaches are often part of an evaluative data gathering cycle for a design refinement purpose - to assess the “fun” of a design and aid production (Ambinder 2009) or inform iterations during the design process (Isbister and Schaffer 2008; Fullerton 2008). Within broader games research, study methods often concentrate on the measuring of player experiences within specific analytical frames, for example, the study of gameplay flow or immersion (Nacke and Lindley 2008). As such, many questions surround what can be considered appropriate for game design researchers, particularly those investigating emerging design approaches ostracized from dominant areas of thought and discussion (Polansky 2015).

This paper discusses the questions of appropriate playtesting and player study methods for game design by framing a specific journey undertaken by the principle researcher. Methodological questions in appropriateness and resources manifested over the course of the principle researcher’s PhD within the game design research field. This began in a lab environment and adopted a more HCI influenced game design lens, and over time transitioned away, towards a more arts-focused design practice outside of a lab. During this process disciplinary difficulties were encountered, and we observed how more arts-based approaches to game design (Muscat et al. 2016) challenge how we think about game design research: A common adherence towards the object and its material properties as a focal point of study, without looking at the broader context in which it is played and experienced.

Research presented in this paper is based around a case study of an experimental game design *WORLD4*, which interrogates these disciplinary questions through a methodological lens. Through a discussion of a two-stage qualitative player study of *WORLD4*, we highlight challenges encountered in adopting academically valid, scholarly approaches to provide a ‘rich’ in-depth account of players’ game experiences. Observations reveal limitations of the testing environment, appropriate context of the

designed experience, and in the availability of resources. In light of these limitations we draw upon alternative remote player-testing methods from within industry, presented by resource-limited independent game designers.

Although qualitative remote testing has been utilized and critically evaluated for effectiveness within design research, for example, asynchronous remote testing where “the evaluator is separate from the user both temporally and spatially” (Andreasen et al. 2007, p.1406), we identify a significant difference in methodological context and application to remote testing methods emerging within commercial game design (Daviau and Leacock (2017), De Jongh (2017)). These emerging methods deemphasize ‘user accuracy’ in participant responses and evaluative data, preferencing more holistic captures of naturalistic gameplay experiences for contextually sensitive, rich design insights. Drawing upon these, this paper seeks to contribute a specific methodological perspective relevant to game design research practice, bringing alternative knowledge into a scholarly fold.

Insights from a comparative analysis of player study approaches inform an alternative hybrid-approach to conducting remote player studies that is contextually sensitive and resource-conscious, intended for game design researchers working alone. In doing so we challenge the positioning of the designer as removed from their work and adoption of controlled testing methods in the pursuit of clean or ‘valid’ data. We argue for a naturalistic approach that enables an analysis of the game experience beyond the game artifact, into the context in which it is first encountered and subsequently experienced.

## INTRODUCING WORLD4 AND THE DESIGN STUDY

*WORLD4* is an experimental game whose experience concentrates on the activities of navigating, negotiating, and revealing an explorable, interconnected world fragmented across four simultaneous first-person perspectives; each reveals a partial view of the 3D world (Muscat and Duckworth 2018). *WORLD4*’s design exaggerates and heightens the traits of ambiguity and the activity of looking identified in the ‘walking simulator’ or ‘walker’ game genre (Muscat et al. 2016). Genre exemplars like *Dear Esther* (2012) and *Proteus* (2013) have challenged game design conventions by focusing away from mechanically-oriented gameplay towards more subjective and interpretive spatial-exploratory game experiences. Drawing inspiration from walker games, the design goal of *WORLD4* was to further an understanding of how spatially ambiguous game experiences can be designed.

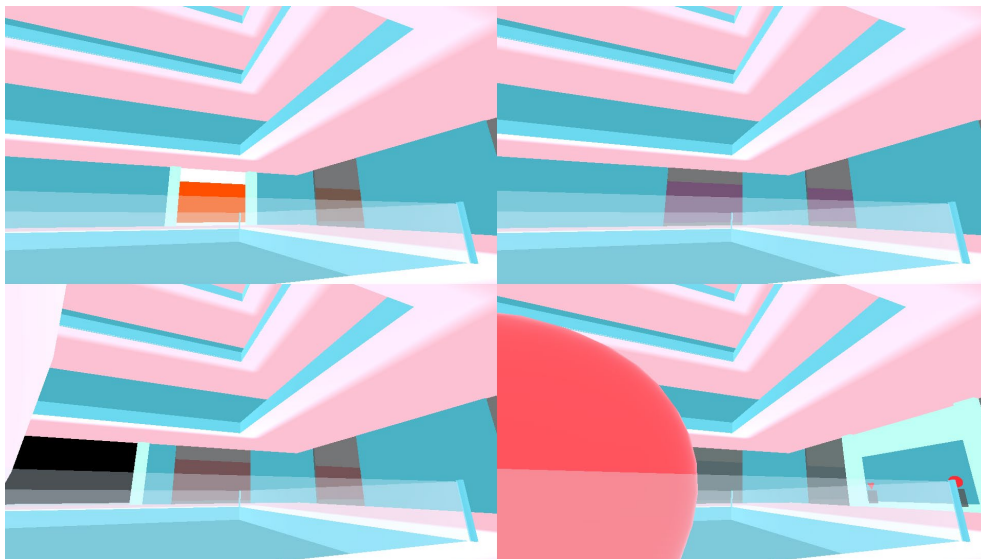


Figure 1: Screenshot, *WORLD4*.

The study objective was not to evaluate *WORLD4* for design iteration, but to understand its exploratory and ambiguous design characteristics, and gain insight into its deliberately subjective traits (i.e. world fragmentation, reduced depth-cues, recurring iconography). This required an approach that could account for how these elements might manifest differently for participants during the game experience. We concluded an in-depth capture and analysis of players' *WORLD4* experience would provide necessary insight, and decided to conduct a close, qualitative observational study. Focus was towards players' in-game behaviors, specifically their navigation, wayfinding, observations, and voiced comments while playing as they might reveal impressions or interpretive, narrative readings of the game world.

The research included a two-stage observational study of 14 participants' player experience, conducted using two alternate techniques informed by playtest approaches. In Study 1 we analyzed 7 participants observed game experiences in-person, adapting grounded theory methods for recorded observation and a conversational interview. Based on observations and limitations noted in the method, the approach was altered for Study 2, drawing upon inspiration from contemporary independent game developer methods. In Study 2 we analyzed 7 remotely recruited participants recorded game experiences and questionnaire results, using an altered grounded theory approach.

## **STUDY 1 APPROACH**

When selecting suitable methods for an initial study of *WORLD4*, our criteria focused on close observation of player behaviors and responses through their in-game actions and vocalized comments. Formalized qualitative player testing approaches were reviewed including Fullerton (2008) (revised from Fullerton et al. (2004)) and Isbister and Schaffer (2008). These provided techniques in capturing 'in the moment' data of players through note-taking, observation, and 'think alouds' so players may vocalize their thoughts as comments (2008, p.66).

*WORLD4*'s intentionally ambiguous, multi-view design suggested 'in the moment' live data capture techniques as ideal for observing exploratory behaviors, as it would allow player participants to vocally specify aspects of their experience and interpretations while playing (Isbister and Schaffer, 2008). The potential impact in the cognitive demand of multi-tasking was considered, but due to *WORLD4*'s reduced temporal pace it was assumed negative drawbacks would be less of an issue in light of worthwhile benefits e.g. narrative readings of the game world. However, Isbister and Schaffer (2008) and Fullerton (2008) specify their approaches as an evaluative tool for iterative design refinement. As *WORLD4*'s usability was not being tested nor its design iteratively refined, instead, more broadly applicable grounded theory methods were opted for, which these qualitative methods are derived from.

Study 1 adapts two observational approaches from grounded theory methods outlined by Charmaz (2006) and LeCompte & Schensul (2010). Both researchers specify a method for recorded observations that utilize multiple data points for comprehensive data capture, translated for the study to capture a participant's onscreen gameplay and vocalized comments through microphone audio, onscreen video capture, and handwritten observational notes taken during in-person observation. To encourage participants to vocalize their thoughts as part of the 'think-aloud' approach, open-ended prompts were adapted Charmaz's (2006 p.24). In addition, a short interview method was adapted to conclude each participant's gameplay session using "ending questions" (Charmaz 2006 p.30) to clarify on-the spot observations noted during the session. To scrutinize data, Charmaz's method for conducting thematic analysis through a multi-stage open coding process was adapted.

## Recruitment

The principle researcher was responsible for conducting the study alone including data gathering and analysis process. To make the study more manageable in time and scope, recruitment was scaled in focus; participants were recruited locally at the university campus and the study conducted in a private office space. Through an open invitation local students were targeted through a poster callout over two weeks, recruiting 7 participants, all university students or recent graduates.

There were no gender or age prerequisites as this was not a point of focus in the study. Existing student–teacher relationships were taken into consideration, however none of the participants had been taught by the principle researcher, with reduced risk of professional conflict<sup>1</sup>. No incentives were offered to participate, instead, *WORLD4* was used as the major draw for involvement in the study. All participants expressed interest in playing *WORLD4* as a major factor in volunteering.

The study area contained a PC and desktop monitor, with speakers and a mouse and keyboard participants could adjust to their preferred settings. Participants were briefed about study prerequisites and *WORLD4*'s controls, but not about the nature of the game. An hour was scheduled for each participant, but all were free to play *WORLD4* for as long as they desired in a single sitting.

## Data Collection

Each session was recorded through onscreen gameplay capture using Open Broadcasting Software and a microphone to capture participants vocal comments. During each session the principle researcher was present, to record handwritten, time-stamped observational notes (LeCompte & Schensul 2010), highlighting noteworthy participant actions, behaviors, and comments. Participants were encouraged to voice their thoughts prior, using open prompts before their gameplay session begun to encourage a 'thinking aloud' voicing of thoughts while playing.

Following the gameplay session, a conversational interview (Charmaz 2006 p.30) expanded upon observations and comments noted. These included intermediate questions specific to game elements and ending questions to prompt reflection towards the game experience. After each session gameplay capture recordings were watched and fully transcribed to pick up on observations and quotes missed, to construct a thorough game experience data capture.

Observations focused on in game spatial-exploratory behaviors and interpretive comments, highlighting specific observations and quotes. Session recordings ranged from 30 to 50 minutes. Similarities were observed across all sessions including comparison and memorization of 3D objects, further investigation of world contents and layout, an overlooking of objects and locations, and returns to locations previously explored.

## Data Analysis

To separate, sort, and study transcription and note data a multi-stage open-coding method derived from Charmaz (2006, p.42) was used for thematic analysis of each participant's recorded game experience, and to identify shared relationships and highlight noteworthy themes. 160 initial codes were generated e.g. 'finding boundaries' and 'surface and object collision' for circumnavigation of rooms and opening

---

<sup>1</sup> Testing was conducted under RMIT University Human Research Ethics guidelines, CHEAN Approval Number: 20227

investigation. Initial codes were subject to an involved coding pass to identify the “most significant and/or frequent earlier codes” (Charmaz 2006, p.57). 50 significant patterns in codes were identified across all sessions, compared, and categorically grouped by likeness.

## **REFLECTION ON STUDY 1 APPROACH**

Study 1 results revealed valuable insights into *WORLD4*’s player experience in codes produced. However, on reflection, unexpected issues in the approach became apparent.

Video and audio recordings were successful in providing detailed account of participant impressions and behaviors. However, multiple capture sources resulted in a substantial volume of datasets to review and transcribe resulting in significant time issues for the principle researcher; For all 7 sessions the transcription process totaled at 34 hours and resulted in over 40,000 words. Data was broad, unwieldy and difficult to scrutinize, requiring additional weeks in the analysis process to thoroughly code data. Software options were considered, but the size and volume of data presented importing issues.

Like walkers that inspired its design (Muscat and Duckworth 2018) *WORLD4*’s ambiguous design lent itself to more solitary and reflective exploration experiences. Consequentially, the controlled study environment and in-person observation did not feel entirely suitable, framing the game experience as a more formal research experiment. On observing recording data the principle researcher’s presence could be felt as an unnatural observer. Although it is difficult to ascertain to what extent this may have influenced each session it introduced an unforeseen social element.

Think aloud prompts were successful in spurring participants to vocalize their thoughts while playing, but a slowing in navigation pace and lingering on objects was observed, raising question to what extent researcher presence was leading. Despite being briefed, four participants casually asked questions about *WORLD4* and its design, for example, whether the environment was randomized or changing over game playtime, highlighting the social dynamic introduced. This also raised question if participants were made to feel compelled play for longer or to express positive, non-critical comments socially acceptable and polite.

In light of these observed limitations the approach was altered in the next stage of study, to test the validity of Study 1 data, and interrogate and crystallize results.

## **STUDY 2 APPROACH**

Alternate methods that removed or downplayed researcher presence and incorporated a more natural game-playing environment were considered. Remote and domestic user testing approaches discussed by design researchers in human–computer interaction (HCI) fields were reviewed, including probes (Gaver et al. 1999; Gaver & Dunne 1999; Hutchinson et al. 2003; Gaver et al. 2004), field testing (Rowley 1994; Nielsen 1998), self-usage (Neustaedter and Sengers 2012), usability (Andreasen et al. 2007), and research through design (Zimmerman et al. 2007) methods. Due to their concentration on data collection across extended timeframes (likely requiring multiple researchers) and usability and ideation purpose, these methods were ultimately unsuited to a game experience study focus.

Instead, asynchronous (a spatial and temporal separation between the researcher and participant (see Andreasen et al. 2007)) remote qualitative player testing methods within the commercial games industry were reviewed. Daviau and Leacock (2017) and De Jongh (2017) present methods conceived in response to widespread industry adoption of analytics-driven game-testing methods (e.g. Zammitto 2015; Ambinder 2009), deemed unsuited to their smaller studio scales and limited resources.

Importantly, both argue analytics driven ‘usability’ focused player-testing testing overlooks naturalistic gameplay experiences.

Daviau and Leacock’s (2017) remote ‘above the table’ testing method created to study boardgame *Pandemic Legacy* involves the sending of physical prototypes to groups of players who record at-home gameplay sessions using a personal phone. Their method seeks to avoid testing ‘blind spots’ found when examining gameplay and describe it as looking for insights in the ‘experience itself’ by allowing less explicit design elements to surface e.g. how the physical components of the boardgame worked and how the rules were read and used. Despite the presence of a recording device, Daviau and Leacock describe their surprise in observing “how relaxed people are in this environment, because people are in their own homes” (2017, t. 36:59).

Logistics factors including time and cost-effectiveness are also considered by Daviau and Leacock, noting their lack of time, budget, and personnel resources. Video recordings are received from participants remotely, observed and annotated using a formatted spreadsheet to identify interesting points for quick categorization, and improve readability and communication between the designers. This method allows designers to ‘internalize and empathize’ with player responses over extended play sessions, suitable to the game’s ‘legacy’ format, designed to be played over multiple sessions like episodes of a serialized television show; unfeasible in-person, taking days to organize, capture, and observe.

De Jongh (2017) presents a time and resource conscious qualitative approach for solo game designers, created to study *Hidden Folks*, informed by observations of game developers and their own experience testing games in private and public settings. De Jongh argues solo designers often adopt resource intensive industry-standard methods unsuited to their designs that do not provide meaningful insight. For example, combined sample groups, analytics, and feedback surveys can produce conflicting and confusing data. Testing games at loud public exhibitions can lead to unsuitable observations, removed from a real-world played context.

De Jongh’s approach extends into recruitment techniques, using the game and designer’s reach online. This involves promoting a playtest via social media, heavily relying upon their established following and sending a version of the game to testers remotely. Testers record themselves playing (on their own device) at home in their own time reducing social pressures e.g. the impulse to offer solutions to the designer, sending captured video and audio commentary data back for analysis.

Both Daviau and Leacock (2017) and de Jongh’s (2017) successes contrasts Andreasen et al.’s conclusions regarding asynchronous methods for instructional usability testing, as time-consuming and unreliable in participant data produced (2007, p.1413). These highlight significantly different methodological objectives to the study of player-experiences for a game design purpose. For Daviau and Leacock (2017) and de Jongh (2017) worthwhile data *is* a holistic impression from a player perspective relative to the designer’s intended experience *rather* than an accurate participant response; by capturing uninterrupted game experiences considering what is ‘beyond the screen’ is no less vital to the methodological context.

Such differences are evident in conclusions towards remote factors, for example, in allowing participants to choose how and when they engage with the study. Although highlighted as a methodological shortcoming by Andreasen et al. (2007), this is a desirable data point to Daviau and Leacock (2017) and de Jongh (2017). Distance allows designers to see whether players will take the initiative to play and for how long,

and allows speak their minds without immediate social or conditional pressures, for example, the impulse of players to offer design solutions.

Based on these impressions de Jongh (2017) and Daviau and Leacock's (2017) methods were adapted for data collection and recruitment techniques, as part of a hybrid approach combined with a streamlined grounded theory observation and data analysis method (LeCompte & Schensul 2010; Charmaz 2006) used in Study 1.

## **Recruitment**

Drawing upon de Jongh's (2017) approach recruitment was focused purely online through an open invitation on social media platform Twitter, not restricted by gender or age. Requirements included the ability to play *WORLD4* remotely and capture onscreen gameplay and microphone audio. Like de Jongh the principle researcher's existing profile was leveraged for recruitment (644 followers). The callout process was more involved than Study 1; recruitment posts were circulated at fixed times each day over two weeks via using popular hashtags to broaden reach to an active global community of game creators (professionals, hobbyists, students) and game-players considered likely participants.

Promotional Twitter posts for the study drew inspiration the posters used in Study 1, centering on *WORLD4* as a distinct and novel experience. Animated GIF images of captured 'movements' from *WORLD4* were used draw attention by providing an attractive glimpse of the game, highlighting the four viewports as a powerful attention-grabbing feature. Posts linked to a sign-up page containing requirement information, estimated total study time (1 hour), privacy and anonymity conditions, and a small optional survey for contextual background information (experience using capture software and video upload preferences, to prepare for data collection).

Sign-up applicants were contacted via email and asked to respond to the test within a two-week timeframe, receiving a reminder email after their first week. All were referred to another page providing step-by-step instructions for conducting the study and download and upload links. 24 applicants were recruited within the two-week period. A falloff in sign-ups to participants was expected (30% as identified by de Jongh (2017)). In total 7 participants were recruited; none were repeats from Study 1 or had played *WORLD4* before.

## **Data Collection**

Participants were recommended to allocate an hour to perform the study with no time restriction on how long *WORLD4* should be played ("play for as long as you desire, quit at any time you choose"). Derived from Study 1's think aloud approach and prompts, participants were asked to voice their thoughts while playing *WORLD4*. Like Study 1 no participants were observed to have difficulty voicing their thoughts while playing, suggesting an initial prompt was enough to encourage voiced commentary.

Following their gameplay session participants were asked to send video files of their recorded screen and microphone capture to the principle researcher via multiple upload options, and to respond to a short open-ended questionnaire and survey (Charmaz 2006, p.36). Although Daviau and Leacock (2017) and de Jongh (2017) discourage survey data, this was adapted as an alternative to Study 1's conversational interview for insights into the testing process and as a space for game experience reflection. Questions were constrained to keep research scope focused on game experience, leaving assumptions about participants' overall game knowledge, age, gender and demographic data, and therefore was not a point of study focus.



Recorded data received was clear and usable, although one participant had audio issues with microphone capture, corrected through video editing software. A partial transcription of each recording was conducted including written observations and notes, adapting Daviau and Leacock's (2017) spreadsheet approach, placing pieces of observation data in categorical columns per player: Time (in video); Context (in-game actions observed); Comments (player comments transcribed); and Additional Notes (researcher thoughts), allowing for a faster recording, varying levels of detail, and clearer formatting of moments and voiced comments observed.

Overall, similarities in participant game experiences were observed in comparison to Study 1 observations. These included comparison of differences across viewports, memorization of recurring 3D objects, further investigation of objects, overlooking locations and objects, and returns to locations previously discovered.

## **Data Analysis**

Transcription and note data was analyzed using a multi-stage open-coding process (Charmaz 2006, p.50) from Study 1, to conduct a thematic analysis of each player's captured game experience. Focus was towards player actions, activities, and comments highlighting exploratory in-game behaviors or game characteristics, resulting in 120 total initial codes across 7 participants, and were subject to an involved coding pass to identify significant patterns. 40 involved codes were identified and placed in groupings categorized by likeness.

Both Study 1 and 2 code data was compared and reviewed to test their viability in being merged. Similarities including in-game behaviors i.e. exploration activities, and descriptive or interesting comments towards game world elements were noted. Study 2 codes supported those already found in Study 1, and furthermore, it was recognized noteworthy observations and quotes in Study 1 had likely affected Study 2 data analysis. While acknowledging problematic aspects, we decided discarding useful data would diminish analysis and discussion, and merged 50 involved codes from Study 1 with 40 from Study 2, totaling at 90 involved codes.

To scrutinize and compare code relationships a visual whiteboard affinity mapping process was adopted. A paper-based approach was preferenced over a software solution like NVivo to determine the viability and ease of a paper-based approach in handling this dataset, following the approach taken in Study 1. Codes were placed as sticky notes on a whiteboard, and gradually clustered by likeness to identify significant patterns, overlap, strong correlations, or those individually distinct. For example, navigation wayfinding, characterization of 3D objects, and investigation of locations were identified. More independent codes were clustered between dominant groupings and their relatability and divergences were discussed. A more general thematic relationship was agreed upon across code groupings, forming seven dominant categorical code groups describing player experience of *WORLD4*, discussed further by Muscat and Duckworth (2018).

## **METHOD REFLECTION AND OBSERVATIONS**

Comparing the tradeoffs, limitations, and advantages in adopting a remote at-home approach, specific methodological considerations were raised, including what we may consider appropriate or 'ideal' when conducting scholarly game design research.

## **Contextually-Aware Player Studies**

Although similarities were observed between the two studies, differences in participant response suggested an advantage in the remote method. Study 1 participants voiced negative comments while playing (e.g. when disorientated), however none expressed strong negative feelings towards the game. Comparatively, Study 2 participants were

observed to express negative comments more frequently, giving more candid commentary and critical thoughts. For example, Player 11 expressed their displeasure and frustration towards *WORLD4* on multiple occasions:

Player 11: “I really hated the colors. They were really uncanny colors”; “As soon as I felt stuck or couldn't find new differences, I stopped being engaged and sort of rebelled against the elliptical language of the game. I'm not stupid, and if a game makes me feel like that, I won't play it any further”.

A 10-minute difference between average gameplay session length was noted: 35 minutes in Study 1, 25 minutes in Study 2. Observations suggest participants may have felt inclined to slow down and direct their comments in the researcher's presence. Without social pressure, participants may have felt less obliged to play for longer durations as their interest declined over time. Furthermore, when given more responsibility and independence in conducting the study, participants may have felt less obligation to meet perceived researcher expectations.

An in-person approach in an isolated room, while useful, could be considered unsuitable for *WORLD4*'s type of game experience: deliberately ambiguous, puzzling and exploratory. As noted in Study 1, participants asked the researcher questions while they played. This type of social discussion and back-and-forth towards and around the game is evidently suited to ambiguous and puzzling game types. In a quiet room environment the researcher had taken on the role of armchair observer part of the social gameplay context and a participant in the game experience.

Isbister and Schaffer (2008) argue by adopting a ‘one way-mirror’ approach physically removing the researcher, this type of social friction and presence can be reduced. Rather than opt for an expensive and potentially awkward lab testing-environment, Study 2 suggests avoiding this by focusing on an intended real-world context. However, although the reduction and removal of the researcher is often seen as an ideal goal, although remote, Study 2 participants were aware of the researcher as an observing presence and directed their comments as if they were physically present ‘in’ the room. For example, sharing personal anecdotes to illustrate their game experience:

Player 9: “This is going to be like a psychoanalysis session ... years ago I stupidly volunteered to be hypnotized ... and that feeling you get when you're hypnotized is about losing your awareness of everything around you, and that disorientation is very similar to the disorientation I am feeling around here”.

Although this type of remote testing can effectively minimize the researcher's physical presence this raises question to the pursuit of ‘removing’ the researcher as a feasible, or even an ideal pursuit. Interrogation of these questions are outside of the study scope, however, observations indicate even remote studies have a bearing on the players' game experience. Appropriate method selection that considers the design and assumed game experience in light of participant comfort and environment may reduce friction, unexpected outcomes, and better withstand methodological scrutiny on reflection.

### **Involved Social-Media Recruitment**

Leveraging online resources proved beneficial in reaching a broader number of potential participants (24). Twitter provided a freely available recruitment resource, raising question to what we may consider suitable positioning of the design researcher to their work. Traditional recruitment methods include mailing lists, social media call outs, and public signups. These do not leverage the researcher or designer's public profile but do leverage the profile of the work associated research lab or group, and university or institute. A researcher without this type of network may have limited

circulation. Profile-based social media recruiting opens an alternate avenue to expand reach, but introduces organizational and contextual considerations.

Having an existing profile and following was essential to recruitment success by increasing call-out circulation for a higher chance of response. For example, the principle researcher's network included colleagues and friends with larger followings who could circulate and 'boost' callouts to a larger audience, creating momentum and traction for postings. This approach reveals parallels with independent commercial game marketing techniques, discussed by professionals like Clinnick (2017) as recruitment required a frequency in call-out posts, content preparation in advance, specific knowledge of social media communities (e.g. hashtags to target) and how to style posts for broad appeal. These added a layer of time required and suggest this type of recruitment is unsuitable to those not active on the targeted platform.



**Figure 2:** Screenshot, web browser. First recruitment callout posted to Twitter (Muscat 2017). Note animated GIF image, URL shortening, and call-out text.

It is important to note much game-development work shared on Twitter is indistinguishable between commercial or non-commercial. For example, a game developer may post both types of project using the same profile and hashtags without clear distinction. Within this context, the personal and promotional are often blurred, an idea reflected in the 'personal brand' marketing concept (see Clinnick 2017), exemplified by designers like de Jongh (2017). In Study 2 the principle researcher's position as a researcher and designer of *WORLD4* was not a neutral one as it required an opening of work up to a broader, partially cultivated audience. Consequentially, this recruitment method is not a one-dimensional method for soliciting player feedback, rather, it can be used for multiple purposes e.g. promoting the project or research, or increasing a following.

We argue these factors are not entirely dissimilar to more traditional or lab-based approaches; although more shielded from the public domain these networks are often occupied by people with similar or shared interests, a connection to the work, researcher, or organization. For these reasons we stress the importance in considering personal biases during recruitment for awareness in how such factors may influence and characterize data produced.

However, research within the public domain, specifically on Twitter does come at a risk for the researcher. Decentralized harassment campaigns targeting women (see Golding & Deventer 2016) highlight dangers in the area of games when engaging with these platforms. A lack of moderation and policing increases the threat of targeted harassment and presents a very real personal risk. As such, we do not wish to glamorize or romanticize these platforms. For those seeking to use this method we ask these risks are properly considered during selection and preparation.

## **RECOMMENDATIONS FOR CONDUCTING REMOTE PLAYER STUDIES**

Derived from the methodological case study analysis, we present the following recommendations aimed at solo, independent game designers and design researchers interested in conducting contextually-aware, qualitative, remote player studies, outside of research lab environments and with limited equipment, space, and time.

### **Recruitment**

Suggestions are specific to Twitter but may translate to other social media platforms. The objective in broader recruitment is to attract sign-ups interested in the project, who will likely commit to the study. Researchers must first establish a clear recruitment time-frame and an ideal, lowest viable number of participants. An estimate fall-off from sign-ups to study participants is 30%; targets will determine the level of circulation required, therefore input and preparation involved.

Before conducting the call-out a clear sign-up form should be prepared (e.g. Google Forms). Succinct language and formatting are essential to reduce friction in the sign-up process. Poor signposting and communication of the nature of the study and ethical information will lead to scrutiny and introduce friction. We suggest using a memorable custom URL when linking to the sign-up form e.g. services like bit.ly.

Content prepared should include promotional text templates and an assortment of attractive media. Posts should include provocations directed at specified disciplines or interests, to engage click-throughs to the sign-up page. Twitter feeds are crowded environments and images are essential to draw attention. We suggest using visually distinct animated GIF images that showcase the project.

Adopting popular hashtags (e.g. #gamedev, #indiedev, #madewithunity) and requesting other users to retweet callouts through post text may boost circulation (see Figure 2). Regional peak-usage times should be considered for stronger visibility to increase chances of receiving sympathetic retweets. Applications like Tweet Deck can automate posts, however, we recommend researchers stay aware of traffic and replies to adjust their approach over the recruitment window.

Targeting additional social media platforms or forums may boost the call-out further, by adding multiple dimensions to communication channels. Researchers must ascertain relevant or potentially interested audiences to not spread their recruitment efforts thin.

### **Data Collection**

To reduce fall-off between applicants and follow-through participants the data collection phase must be planned, streamlined and engaged within a week of sign-ups, to keep the study within their mind's eye. We recommend preparing a template email for contacting applicants, which can be automated if desired. This should clearly establish a study time-frame and deadline so participants can plan ahead, and state that a reminder will be sent if data has not received. In addition, provide a contact channel to the researcher for participant questions.

We recommend separating instructional information from the contact email, so participants are not overwhelmed. In the email, link to a study page outlining the step-by-step process required. Details should include an estimate timeframe, requirements (e.g. hardware and software required), download links, installation and running instructions (e.g. installing and running capture software), study instructions (e.g. think aloud prompts), uploading instructions and links, and any final questions. In advance, researchers must consider providing instructional information to assist participants setting up software e.g. a video tutorial instructing how to install capture software.

To reduce potential difficulties or issues that may surface during the data capture phase, we suggest recommending participants to run a test recording of their microphone and screen capture. In addition, providing participants multiple options for sending their recordings is ideal due to the size of recorded video data and variability of upload speeds. For example, a private Dropbox folder created by the researcher, allowing participants to supply a link to their own personal Dropbox or an unlisted video link to a personal video-service profile (e.g. YouTube).

Flexibility and leeway in data retrieval timeliness can account for scheduling issues and better accommodate for the additional responsibility placed on participants, who may follow through even past the stated deadline.

## **Data Analysis**

A more efficient data analysis process begins during the capturing of recorded data. We suggest partially transcribing received gameplay capture through a method derived from Daviau and Leacock's (2017) spreadsheet process. A categorical spreadsheet method allows for varying levels of detail in partial-transcription, observation, and note-taking data, within the same chronological format. Each piece of observation data should be placed in categorical columns in chronological rows, on a sheet dedicated to each individual player: Time (in video); Context (in-game actions observed); Comments (player comments transcribed); and Additional Notes (researcher thoughts).

This allows faster recording, varying levels of detail, and clearer formatting of participant behaviors, activities and comments observed, in context and time sequence. For example, capturing specific contextual responses such as a participant's voiced comments towards a game world object, and following their in-game actions to interact with it. In addition, columns can be added to the spreadsheet during the coding process for easier tagging, categorical labelling, and data organization. Researchers who find an optimal pace may seek to increase the speed of recording capture playback to x1.2 or x1.4 speed, to reduce overall transcription time spent. Reduce time required and noise detail may allow relevant insights to be better ascertained during data coding.

## **CONCLUSION**

In conclusion, through the two-stage qualitative study of *WORLD4* issues in the appropriateness of formalized industry-standard and academically-valid methodologies have been raised. Such issues are specifically relevant to designers and design researchers seeking to study emerging or arts-based design styles, experimental genres, and game experiences. These reveal difficulties and tensions in conducting contemporary and relevant scholarly game design research, especially relevant to researchers engaging in a game making practices and working alone outside of resourced lab environments.

If design researchers are to overcome such issues within their domain, we suggest they look beyond the design object and consider their game making and research practice within a broader, contemporary game making context. By engaging in emerging and current development contexts and approaches like those shared by commercial

independent designs like de Jongh (2017) and Daviau and Leacock (2017), researchers can ground investigation in relevant game making practices and broaden their reach towards areas of thought and process that would otherwise remain outside the fold of game design research.

Although such shifts are difficult, especially for researchers working alone, an adherence to academically valid or formalized methodologies and approaches present their own risks when studying designs. These include the pursuit of ‘clean’ player experience capture, that assumes universal traits between designs and a fixed environment in which they are experienced. Although such approaches provide focus, relevant design considerations remain outside the frame, such as the game experience ‘beyond the screen’ not always obvious or outwardly expressed or intended in the object itself. Furthermore, an uncritical adoption of these methods as evaluative tools can mislead us in assuming their insights are clearer and reproducible and therefore more valid as scholarly or design research tools.

We posit the role of the game design researcher is perhaps better served in looking towards arts-based practices which make fewer clear distinctions between the object and the artist but considers their positional context. If game design researchers are to produce relevant knowledge and adjust to rapid changes within the field of games design, they must consider their role relative to more intermingled real-world game-making practices and design contexts that designers like de Jongh (2017) engage in. We hope our discussion and contribution will provoke game design researchers to challenge established evaluation techniques, approaches that consider player experiences, and naturalistic contexts more broadly, and directly incorporate and extend qualitative methods beyond the designed game itself.

## ACKNOWLEDGMENTS

We thank all the volunteer participants for their contribution to the study.

## BIBLIOGRAPHY

- Ambinder, Mike. 2009. “Valve’s Approach to Playtesting.” presented at the *Game Developers Conference*, San Francisco, USA.
- Andreasen, Morten Sieker, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. “What Happened to Remote Usability Testing?: An Empirical Study of Three Methods.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1405–1414. CHI ’07. New York, NY, USA: ACM. <https://doi.org/10.1145/1240624.1240838>.
- Boellstorff, Tom, Bonnie Nardi, Celia Pearce, T. L. Taylor, and George E. Marcus. 2012. *Ethnography and Virtual Worlds: A Handbook of Method*. Princeton: Princeton University Press.
- Brockmyer, Jeanne H., Christine M. Fox, Kathleen A. Curtiss, Evan McBroom, Kimberly M. Burkhart, and Jacquelyn N. Pidruzny. 2009. “The Development of the Game Engagement Questionnaire: A Measure of Engagement in Video Game-Playing.” *Journal of Experimental Social Psychology* 45 (4): 624–34. <https://doi.org/10.1016/j.jesp.2009.02.016>.
- Charmaz, Kathy. 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. SAGE.
- Daviau, Rob, and Matt Leacock. 2017. “Board Game Design Day: The Making of ‘Pandemic Legacy.’” presented at the *Game Developers Conference 2017*, San Francisco, USA. <https://www.gdcvault.com/play/1024300/Board-Game-Design-Day-The>.

- Ed Key & David Kanaga. 2013. *Proteus*. PC, Mac.
- Fullerton, Tracy. 2008. *Game Design Workshop, Second Edition: A Playcentric Approach to Creating Innovative Games*. 2nd edition. Burlington, Massachusetts, United States: Morgan Kaufmann.
- Fullerton, Tracy, Chris Swain, and Steven Hoffman. 2004. *Game Design Workshop: Designing, Prototyping, & Playtesting Games*. Waltham, Massachusetts, United States: Focal Press.
- Gaver, Bill, Tony Dunne, and Elena Pacenti. 1999. "Design: Cultural Probes." *Interactions* 6 (1): 21–29. <https://doi.org/10.1145/291224.291235>.
- Gaver, William, and Anthony Dunne. 1999. "Projected Realities: Conceptual Design for Cultural Effect." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 600–607. CHI '99. New York, NY, USA: ACM. <https://doi.org/10.1145/302979.303168>.
- Gaver, William W., Andrew Boucher, Sarah Pennington, and Brendan Walker. 2004. "Cultural Probes and the Value of Uncertainty." *Interactions* 11 (5): 53–56. <https://doi.org/10.1145/1015530.1015555>.
- Golding, Dan, and Leena van Deventer. 2016. *Game Changers*. South Melbourne, VIC, Australia: Affirm Press.
- Hutchinson, Hilary, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, et al. 2003. "Technology Probes: Inspiring Design for and with Families." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 17–24. CHI '03. New York, NY, USA: ACM. <https://doi.org/10.1145/642611.642616>.
- Isbister, Katherine, and Noah Schaffer. 2008. *Game Usability: Advancing the Player Experience*. 1 edition. San Francisco, Calif.: Oxford: CRC Press.
- Jongh, Adriaan de. 2017. "Playtesting: Avoiding Evil Data." presented at the *Game Developers Conference 2017*, San Francisco, USA. <https://www.gdcvault.com/play/1024132/Playtesting-Avoiding-Evil>.
- Lankoski, Peter, and Staffan Björk. 2015. *Game Research Methods*. ETC Press. <http://press.etc.cmu.edu/index.php/product/game-research-methods/>.
- Laurel, Brenda. 2003. *Design Research: Methods and Perspectives*. MIT Press.
- LeCompte, Margaret Diane, and Jean J. Schensul. 2010. *Designing & Conducting Ethnographic Research: An Introduction*. Rowman Altamira.
- Llanos, Stein C, and Kristine Jørgensen. 2011. "Do Players Prefer Integrated User Interfaces? A Qualitative Study of Game UI Design Issues." In *Proceeding of DiGRA 2011 Conference: Think Design Play.*, 12. Utrecht, Netherlands: Digital Games Research Association.
- Marczak, Raphaël, Gareth Schott, and Pierre Hanna. 2015. "Understanding Player Experience Through the Use of Similarity Matrix." In *Proceedings of DiGRA 2015: Diversity of Play: Games – Cultures – Identities*, 16. Lüneburg, Germany: Digital Games Research Association.
- Nacke, Lennart, and Craig Lindley. 2008. "Flow and Immersion in First-Person Shooters: Measuring the Player's Gameplay Experience." In *Future Play '08 Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, 81–88. Toronto, Ontario, Canada: ACM. <https://doi.org/978-1-60558-218-4>.

- Neustaedter, Carman, and Phoebe Sengers. 2012. "Autobiographical Design in HCI Research: Designing and Learning Through Use-It-Yourself." In *Proceedings of the Designing Interactive Systems Conference*, 514–523. DIS '12. New York, NY, USA: ACM. <https://doi.org/10.1145/2317956.2318034>.
- Nielsen, C. 1998. "Testing in the Field." In *Proceedings of the 3rd Asia Pacific Computer Human Interaction*, 285–90. Shonan Village Center, Japan: <https://doi.org/10.1109/APCHI.1998.704437>.
- Polansky, Lana. 2015. "Against Flow." *Sufficiently Human* (blog). 2015. <http://sufficientlyhuman.com/archives/995>.
- Muscat, Alexander. Twitter post. August 7, 2017, 5:17 p.m. <https://twitter.com/alexandermuscat/status/894457415143276544>
- Muscat, Alexander, and Jonathan Duckworth. 2018. "WORLD4: Designing Ambiguity for First-Person Exploration Games." In *The Annual Symposium on Computer-Human Interaction in Play Extended Abstracts*, 341–351. CHI PLAY '18. Melbourne, VIC, Australia: ACM. <https://doi.org/10.1145/3242671.3242705>.
- Muscat, Alexander, William Goddard, Jonathan Duckworth, and Jussi Holopainen. 2016. "First-Person Walkers: Understanding the Walker Experience through Four Design Themes." In *DiGRA/FDG '16 - Proceedings of the First International Joint Conference of DiGRA and FDG*. Vol. 13. Dundee, Scotland: Digital Games Research Association. [http://www.digra.org/wp-content/uploads/digital-library/paper\\_318.pdf](http://www.digra.org/wp-content/uploads/digital-library/paper_318.pdf).
- Ribbens, Wannes, and Yorick Poels. 2009. "Researching Player Experiences through the Use of Different Qualitative Methods." In *Proceedings of the 2009 DiGRA International Conference: Breaking New Ground: Innovation in Games, Play, Practice and Theory*, 5:10. Brunel University.
- Rowley, David E. 1994. "Usability Testing in the Field: Bringing the Laboratory to the User." In *CHI '94 Conference Companion on Human Factors in Computing Systems*, 217–. CHI. New York, NY, USA: ACM. <https://doi.org/10.1145/259963.260374>.
- Schott, Gareth, and Raphaël Marczak. 2014. "<Exploring> the <Cause> of Game (Derived) <Arousal>: What Biometric Accounts of Player Experience Revealed." *Transactions of the Digital Games Research Association* 2 (2). <https://doi.org/10.26503/todigra.v2i2.38>.
- The Chinese Room. 2012. *Dear Esther*. Windows.
- Zammito, Veronica. 2015. "Moving the Needle: User Research at the End of the Production Cycle." presented at the Game Developers Conference 2015, San Francisco, USA. <https://www.gdcvault.com/play/1022491/Moving-the-Needle-User-Research>.
- Zimmerman, John, Jodi Forlizzi, and Shelley Evenson. 2007. "Research Through Design As a Method for Interaction Design Research in HCI." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 493–502. CHI '07. New York, NY, USA: ACM. <https://doi.org/10.1145/1240624.1240704>.