

Using Interviews to Evaluate Location-Based Games: Lessons and Challenges

Nayana Carneiro  [Universidade Federal do Ceará | nayanacarneiro@great.ufc.br]

Ticianne Darin  [Universidade Federal do Ceará | ticianne@virtual.ufc.br]

Mateus Pinheiro  [Universidade Federal do Ceará | mateus@virtual.ufc.br]

Windson Viana  [Universidade Federal do Ceará | windson@virtual.ufc.br]

Abstract Games User Research (GUR) is an interdisciplinary field of study that aims to measure, analyze, and understand players' interaction and experience with digital games. Joining efforts to the advances in GUR, this work focuses on the evaluation of player interaction with location-based games (LBGs), seeking to understand which aspects of this interaction can be explored through the application of interviews and how researchers and practitioners have been using this method. We analyzed 23 studies that applied interviews to this end and conducted an expert opinion survey with these studies' authors. As a result, we presented lessons and research challenges for the use of interviews in this type of evaluation to encourage the conscious and systematic application of this method and guide students, practitioners, and researchers – especially beginners.

Keywords: *Location-based games, Games User Research, Interviews, Evaluation, Interaction, Player Experience*

1 Introduction

Games User Research (GUR) is an interdisciplinary field of research and practice, which ties together Human-Computer Interaction (HCI) and Game Development (Drachen et al., 2018; Abeele et al., 2020). GUR focuses on measuring, analyzing, and understanding players' interaction and experiences with digital games to optimize game designs (Abeele et al., 2020). Thus, GUR experts seek to understand how players experience specific game design choices and how they lead to specific emotional responses, aiming to design games that meet players' expectations and produce actionable insights to guide game development activities (Abeele et al., 2020). To accomplish this mission, the GUR community also works on adapting and improving methods from other areas to compose a toolbox that helps practitioners answer relevant questions about players' behavior and attitudes in different types of games (Nacke et al., 2016; Drachen et al., 2018). During the last years, GUR has obtained many advances in this endeavor.

Seeking to join efforts to GUR advances, this work applies GUR to the study of *location-based games* (LBGs), a subtype of pervasive games that use location technologies to integrate the position of one or more players into their rules as a central element of the game (Kiefer et al., 2006; Ahlqvist, 2018). This information modifies the game state at run time, creating a meaningful connection between the real and the virtual worlds and transforming the physical space in the game scenery (De Souza e Silva and Sutko, 2011). For this reason, LBGs have specific characteristics that distinguish them from other conventional games, such as mobility, spatial expansion, and pervasiveness (Kiefer et al., 2006; Ahlqvist, 2018).

In the last decade, not only LBGs have achieved great success among the public – especially with Pokémon Go (2016) and, more recently, Harry Potter: Wizards Unite (2019) –, but they also have conquered space in industry and academy

(Kasapakis and Gavalas, 2015). Researchers have used these games in studies in different domains – such as health (Chittaro and Sioni, 2012), education (Oppermann et al., 2017), and tourism (Ballagas et al., 2008) – and consequently have faced new challenges, for instance, related to the evaluation of the player-game interaction and the resulting player experience (PX) (Paavilainen et al., 2017; Carneiro et al., 2019b).

In this context, we conducted a three-phased research with the ultimate goal of proposing a guide for qualitative evaluation of PX in LBGs, aiming to provide directions to researchers and practitioners on this task. The first phase consisted of a systematic mapping (SM) study (Kitchenham et al., 2010) in which we investigated how researchers and practitioners have been evaluating the player interaction with LBGs, as reported in Carneiro et al. (2019b). The results showed that PX had been the most commonly evaluated quality of the interaction and most studies applied evaluation strategies heavily based on surveys and questionnaires, mainly ad hoc instruments.

Surveys and questionnaires are extremely useful, as stated in the literature. However, they offer only an overview of the studied phenomenon and provide limited feedback (Lazar et al., 2017; Drachen et al., 2018). Thus, their isolated application should be avoided in studies that seek in-depth investigations, since an approach based on multiple methods is the most desirable path to follow (Lazar et al., 2017; Creswell and Poth, 2018). One of GUR experts' methods to investigate PX deeply and obtain rich feedback from players is interviewing (Drachen et al., 2018). This method offers several benefits for user research, especially in the context of PX evaluation, being useful to compose an understanding of the needs, preferences, motivations, and attitudes of players (Lazar et al., 2017). In GUR studies, interviews are an essential part of a qualitative testing session with users (Drachen et al., 2018).

Thus, this paper reports the second phase of our research, whose goal was to understand how we can use interviews

to evaluate player interaction with LBGs, and how our community has been doing it. For that purpose, we made an additional analysis of 23 works identified in the SM that reported using interviews in their evaluations and conducted an opinion survey with the authors of these papers to investigate how they have used the method in that context and extract lessons from their experiences. This work's contribution is in identifying lessons and challenges to foster discussion and reflection on the use of interviews to evaluate player interaction with LBGs, thus encouraging its conscious and systematic application. These lessons can help practitioners and researchers, especially the novice, plan and conduct LBGs evaluations using qualitative interviews. It is worth to notice that the present paper is an extension of a paper (Carneiro et al., 2019a) published in the *1st Workshop on Interaction and User Research on Game Development* (WIPlay 2019) and brings some additional information about the research and new lessons.

2 Location-Based Games (LBGs)

Location-based games are digital games that use location technologies to integrate one or more players' positions to their rules as a central element in their dynamics (Kiefer et al., 2006; Ahlqvist, 2018). This information about the player's location is applied to modify the game status during its execution, and creates a meaningful connection between the real and the virtual world, turning the physical space into the game scenario (De Souza e Silva and Sutko, 2011; Alha et al., 2019). Therefore, the most distinctive trace between LBGs and most digital games is that, in LBGs, players' actions predominantly occur in a physical environment, usually outdoor, public spaces (Leorke, 2018; Alha et al., 2019).

Thus, the idea of LBGs has two key elements: (i) players' position and movement in real-world spaces, as the game requires considerable locomotion, and (ii) a game dynamic mediated through some position-tracking technology (Ahlqvist, 2018). LBGs use location-aware technologies, wireless networks (e.g., Wi-Fi and 3G/4G), and portable devices (e.g., smartphones) to allow the communication between physical and digital spaces and between players (De Souza e Silva and Sutko, 2009; Leorke, 2018). Therefore, game interactions usually occur while players use their devices to explore the world, as they physically move around and, at the same time, visit points of interest inside the game.

Despite these elements that distinguish LBGs from conventional digital games, any survey on LBGs history can show that, in the last two decades, LBGs have assumed many and sometimes ambiguous forms – see the work of Leorke (2018) to an in-depth discussion of this history. Although there are some definitions concerning these games, previous studies have discussed that it can still be tricky to define LBGs precisely, as there is a diverse range of games that fit into this category (Kiefer et al., 2006; De Souza e Silva and Sutko, 2011; Ahlqvist, 2018; Leorke, 2018). As Leorke states, no simple definition that applies to LBGs and the task of attempting to define these games is also “not helped by the fact that there are almost as many terms used to define them as there are types of games” (Leorke, 2018).

Several researchers have attempted to draw a set of LBGs' core characteristics in the face of this issue. For instance, De Souza e Silva and Sutko (2009) indicate three main characteristics of LBGs: mobility, spatiality, and sociability. The authors point spatiality as an essential characteristic of LBGs since these games expand the *magic circle*¹ by creating a unique way of connecting players with one another and to space, thus defining a new logic for the game space. Another example lies in Ahlqvist (2018) proposal of five key dimensions to identify and characterize LBGs: location, spatial and temporal expansion, representation, and pervasiveness. The author also emphasizes that a game to be considered an LBG should require considerable physical movement of its players.

These studies indicate what can be considered the primary (or more typically addressed) characteristics of LBGs. Although relevant, it is not our intention to exhaust this discussion in this paper. As mentioned above, these studies are just a few examples of how LBGs have been subject to many interpretations by scholars and designers (Leorke, 2018). However, this brief overview helps us clarify the distinction between LBGs and other digital games, an essential notion to our reader, and highlight these games' specificities². Instead of merely focusing on LBGs technological features, we approach these games, emphasizing meaningful aspects of the player-game relation, which can be affected by LBGs characteristics, such as the perception of spaces and the connection between life and game spaces. Building on Leorke's suggestion, we gathered this brief discussion as a way to lay the ground for an analysis that encompasses the diverse and multifaceted dimensions of these games.

3 Interviews in Games User Research

Interviews offer several benefits for user research and are valuable to encourage in-depth investigation of specific topics and answer questions based on the feedback provided by the interviewees (Lazar et al., 2017). In addition to being one of the most used research methods in HCI (Lazar et al., 2017; Shneiderman et al., 2016), interviews are useful in the context of games studies, especially when investigating the PX, because they help to build an understanding of the needs, concerns, preferences, and attitudes of players (Isbister and Schaffer, 2008). They also allow researchers to explore more complex issues than those addressed in other methods, for example, surveys (Schell, 2014).

In GUR, interviews are an essential part of a qualitative testing session with users (Drachen et al., 2018), as they offer one of the only ways to validate observations, discover problems, collect opinions and find causes for difficulties faced by players (Drachen et al., 2018). Interviews can be combined with other methods to enrich the collected data and create a holistic view of the user's thinking and behavior, being

¹ Salen et al. (2004) defines the magic circle as a primary space in which the game action happens.

² By specificities, we refer to LBGs distinctive traces: we mean not only LBGs characteristics but also particular issues that are uncommon to conventional games, but familiar to LBGs, because they are caused by or related to LBGs characteristics, such as excessive physical effort, GPS imprecision, safety and privacy issues.

a primary part of the discovery and understanding of usability problems and setbacks in the player's experience. Thus, interviews may be the most appropriate choice for certain research objectives and types of knowledge desired (Isbister and Schaffer, 2008). For example, Nacke, Drachen, and Göbel indicate using interviews to evaluate the PX and capture the context and its social impact on the individual player's experience with serious games (Nacke et al., 2010).

Since this method presents the potential to reveal emotional dimensions of experience that are not always evident in an individual's behavior (Lamont and Swidler, 2014), several researchers have explored it, devising new techniques and approaches to apply it in different areas and contexts. For example, El-Nasr and colleagues El-Nasr et al. (2015) proposed a formative evaluation method that uses retrospective interviews to investigate how players accept and integrate a game in their lives, posing it as particularly suitable for research on pervasive games in naturalistic settings. Crawford, Monks, and Wells (Crawford et al., 2018) developed a virtual reality-based interview technique to assess candidates for medical emergency residency and identify their communication, problem-solving, and teamwork skills. More recently, Holmes Holmes (2019) has explored the play-based interview method, which allows observing and interviewing young children, taking into account their cognitive and language limitations, through playful activities that promote engagement and direct communication between interviewers and interviewees. These studies are just a few examples of the versatility and potential for using interviews.

However, it is necessary to highlight that planning, interviewing, and analyzing data is an arduous task that requires preparation and rigor, for applying qualitative interviews is not trivial. Questions content and the manner the interviewer poses the questions, for instance, can determine the difference between new insights and a waste of time. Applying this method brings real challenges to the interviewer, requiring specific skills, study, practice, and experience. Nevertheless, it can result in invaluable data. Hence, appropriate and specific instructions can aid the interview preparation process. Literature is filled with materials to help researchers and professionals (especially novice) plan and conduct qualitative interviews in different areas. Although several authors offer orientations and guidelines for using interviews (e.g., Turner III (2010); Blandford et al. (2016); Kvale (2008); Rowley (2012)), more specific issues will still require careful consideration.

Even though the difficulties associated with applying qualitative interviews are widely known, ironically, many researchers and practitioners see in this method an effortless manner to obtain data quickly. As identified by Myers and Newman Myers and Newman (2007), several studies report using interviews inadequately and superficially. Their study analyzed papers in Information Systems (IS), but this deficiency is not exclusive to their area. As we observed in our systematic mapping study Carneiro et al. (2019b), LBGs interaction evaluation still struggles with similar issues, such as the suitability of the chosen methods to the research goals, inadequate application of methods and techniques, and lack of rigor. Nevertheless, to the best of our knowledge, there are no studies focused on the issues of using qualitative interviews

to evaluate LBGs. Thus, this work investigates these topics as an initial step towards the treatment of such problematic practices.

4 Methodology

In the first phase of this research, we conducted a systematic mapping (SM) of the literature to investigate how researchers and practitioners have evaluated the qualities of the interaction between players and LBGs. SM studies consist of a secondary study method that reviews existing primary studies, indicated to construct an overview of a research area (Petersen et al., 2015). The SM followed the approach proposed by Kitchenham et al. (2010), and aimed to answer three questions regarding LBGs interaction evaluation:

- What methods do researchers and practitioners use in LBGs interaction evaluation?
- What qualities of the interaction do they evaluate?
- What strategies and approaches do they use?

We searched for studies on five sources (Scopus, ACM Digital Library, Web of Science, IEEE Xplore, and Science Direct) using the search string shown in **Figure 1**. As a result, we obtained an initial set of 437 papers filtered in a three-step process, applying the inclusion and exclusion criteria summarized in **Table 1**. The filtering resulted in the final set of 51 articles, which we analyzed to extract information to answer our research questions.

```
((("location based game" OR "location based mobile game")
OR (("pervasive game" OR "ubiquitous game" OR "augmented
reality game") AND (location OR locative OR mobile OR
outdoor))))
AND
((user OR player OR gamer)
AND
(evaluation OR assessment OR study))
```

Figure 1. Search string submitted to the sources (Carneiro et al., 2019b)

Table 1. Main inclusion and exclusion criteria

Inclusion	The study focus on LBGs <i>AND</i> It reports an evaluation process of qualities of the player-game interaction <i>AND</i> It is a primary study
Exclusion	The study does not focus on LBGs <i>OR</i> The reported evaluation does not assess qualities of interaction <i>OR</i> The paper is not written in English or Portuguese

According to our results, the PX is the most frequently evaluated quality of the interaction. Regarding methods, the application of surveys and questionnaires was the most used, followed by interviews and interaction logs recording. Although many papers do not provide detailed information about the evaluation process's specifics, it was possible to notice that most studies use multi-methods approaches and

ad hoc evaluation instruments. The complete report of the SM and its results are available in Carneiro et al. (2019b).

The findings obtained with the SM directed our focus towards a more in-depth examination of the use of interviews to evaluate LBGs since our initial analysis indicated a gap regarding the systematic exploration of the method in this context. We also observed that many studies reported interviews with less rigor than other methods, reducing its relevance. Thus, in the second phase of our research, we performed additional quantitative analysis on 23 papers (extracted from the final set of 51) that reported using interviews and conducted an opinion survey with the authors of those articles to understand their use of interviews.

4.1 Expert Opinion Survey

Opinion surveys aim to determine what the participants think about certain concepts (Ozok, 2009). When conducted with experts, they can serve various purposes, being useful for identifying problems, predicting changes, and clarifying relevant issues on a specific topic (Rowe and Wright, 2001; Darin et al., 2019), for example. Therefore, we created an online survey to deepen our understanding of how the authors of the 23 papers applied interviews to evaluate the player interaction with LBGs in their studies, aiming to identify gaps and extract lessons from their experiences with the method.

We invited the authors to participate in the survey mainly by e-mail. In the case of authors whose e-mail addresses we couldn't identify, we contacted them through the social network *Research Gate*³, when possible. We could not find six authors' contact information, so we excluded them from the mailing list, resulting in 81 recipients.

Each author received a personalized message, containing a brief description of our research, the title of her/his article identified through our SM, and a hyperlink to access the survey. Some participants authored more than one paper, so they received a slightly different questionnaire, which mentioned their identified works, but asked them to answer our questions considering their experience as a whole. Ten days after the invitation, we sent reminders to the authors who had not answered the survey yet. The questionnaire was available for approximately two months, and we closed it after the number of responses stagnated.

The survey had 17 questions (eight of them were open-ended questions) divided into three sections: (i) *Researcher/practitioner profile*; (ii) *Your experience applying interviews* – questions about the use of interview reported in the identified paper; and (iii) *Your Opinion* – a poll regarding the author's openness to a proposal of a guide to using interviews to evaluate LBGs. **Appendix A** presents the survey questions.

Fourteen authors, responsible for 11 papers, participated in the survey, representing a response rate of 17.28% (N = 14). Despite the relatively small number of responses, the literature attests to the fairness of this response rate since it advises the use of groups from 5 to 20 experts (Rowe and Wright, 2001). Besides, the fact that the participants have different

backgrounds and responded to the survey independently reduces the potential for biases in answers (Darin et al., 2019).

4.1.1 Analysis of the survey answers

We treated the answers to close-ended questions as quantitative data and performed basic quantitative analysis, while the answers to open-ended questions were analyzed qualitatively. The results were then crossed and combined to address relevant topics – as presented in Section 5.2 – in the light of the survey objective: understand how authors had used interviews to evaluate player-interaction with an LBG.

The goal of qualitative analysis is to turn the unstructured data into descriptions about important aspects of the situation or problem under consideration (Lazar et al., 2017). To this end, we analyzed the responses to the eight open-ended questions by combining two basic qualitative analysis approaches: thematic analysis and data categorizing (Preece et al., 2019). Thematic analysis is an analytical technique that aims to identify, analyze, and report patterns in the data, in which a theme is something important to the study goals. In our case, data categorizing involved inductive analysis to allow themes to emerge from the data itself and use the results to answer the study goals.

Thus, we tabulated the data, grouped it – initially, guided by the questions' themes – and analyzed iteratively to extract descriptions to answer each topic. In a second moment, we coded the data according to topics that emerged from itself. From the codes, we created categories that were systematically analyzed and combined with knowledge and perceptions gained through literature reviews to translate them into the lessons in Section 6. In Lazar et al. (2017)'s words, this application of experience and contextual knowledge is critical for the appropriate interpretation of qualitative data.

5 Results and Discussion

This section presents and discusses the results of the second phase of this research, summarizing the quantitative analyses performed on the 23 papers (identified in the SM study) that used interviews in their evaluations and the expert opinion survey with the authors.

5.1 Use of interviews as identified in the SM Study

The 23 studies reported evaluating 21 LBGs as shown in **Table 2**. Most (80.95%) of the evaluated games were developed for research purposes – only three (14.29%) games were commercial. Eight (38.10%) studies evaluated functional prototypes, ten (47.62%) evaluated versions for testing, and three (14.29%) used versions available on the market. Three (14.29%) LBGs had educational purposes, while 18 (85.71%) aimed at entertainment. In the remainder of this subsection, we discuss issues related to the evaluation performed on these games. However, it is important to emphasize that results related to the interaction qualities under evaluation and sample sizes refer to all the methods applied in each study, not necessarily only in interviews.

³<https://www.researchgate.net/>

Table 2. LBGs under evaluation in the analyzed papers

LBG	Paper	LBG	Paper
Barbarossa	Kasapakis and Gavalas (2017)	Knowledge War	Wang et al. (2016)
Block!Block!	Lankes and Stiglbauer (2016)	LOCUNET game	Diamantaki et al. (2011)
Can You See Me Now?	Benford et al. (2006)	Obscura	Quek and See (2015)
Cutting Corners	Sandham et al. (2011)	Parallel Kingdom	Verdejo et al. (2010)
Destination	Baillie et al. (2010)	Pokémon GO	Pyae et al. (2017)
Epidemic Menace	Lindt et al. (2007)	Spellbound	Sra and Schmandt (2015)
Fruit Farmer	Calvillo-Gómez et al. (2011)	Shhh!	Linehan et al. (2013)
Ghost Detector	Nilsson et al. (2016)	SoundPacman	Chatzidimitris et al. (2016)
GlowPhones	Merritt et al. (2017)	The Songs of North	Ekman (2007)
Growl Patrol	Kurczak et al. (2011)	TimeWarp	Blum et al. (2012); McCall et al. (2011); McCall and Braun (2008)
Horror game	Prasetio et al. (2017)		

5.1.1 Number of participants

The sample sizes used in the studies ranged from 6 to 96 users. The average number of participants was 24.56 users per study (SD = 19.72), and the most frequently used samples consisted of 10 or 24 participants (reported in 3 studies each). The works with the largest samples brought together 36 (Nilsson et al., 2016), 60 (Blum et al., 2012) and 96 (Sandham et al., 2011) participants.

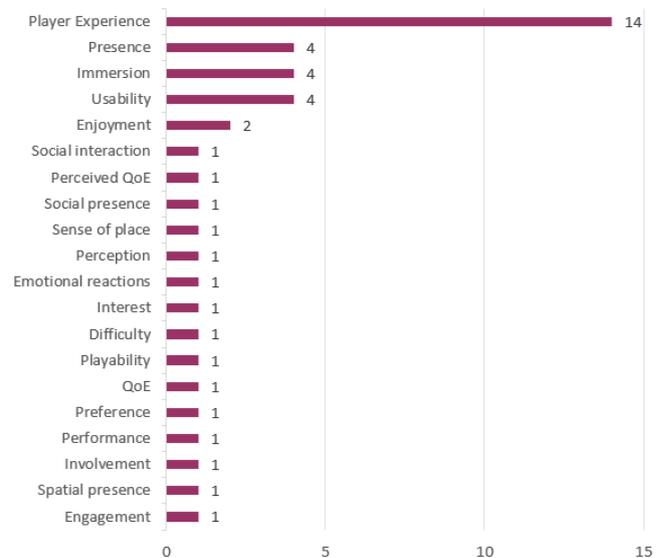
5.1.2 Qualities of interaction under evaluation

We identified twenty qualities of interaction under evaluation, with an average of 1.87 qualities per study (DP = 1.14). PX was the most frequently quality evaluated (indicated in 14 works), followed by usability, immersion, and presence (evaluated in four studies each). Other less evaluated qualities were engagement, playability, and spatial presence, as shown in **Figure 2**. Most studies (12, i.e., 52.17%) focused on evaluating only one quality – in seven of them, that quality was PX. The work of McCall and Braun (McCall and Braun, 2008) evaluated the highest number of qualities (five), namely: usability, PX, presence, sense of place, and social presence. Of the 14 works that evaluated PX, seven also reported evaluating other qualities: usability, engagement, immersion, enjoyment, and presence.

It is worth mentioning that, frequently in the literature, some of these qualities (e.g., presence and immersion) are addressed as PX components, or even part of the set of properties that describe this experience, as stated by Sánchez et al. (Sánchez et al., 2012), for example. However, these works seem to treat PX and the referred qualities as distinct aspects of interaction without discussing this issue.

5.1.3 Evaluation methods and strategies

In addition to interviews, we identified ten other methods applied in the studies, as shown in **Figure 3**. Three (13.04%) studies used a single method (i.e., interview), while the other 20 (86.96%) followed multiple methods approaches (Lazar et al., 2017; Creswell and Poth, 2018). The three studies that used only interviews (Quek and See, 2015; Linehan et al., 2013; Ekman, 2007) evaluated PX in games developed for research – Ekman (2007) also evaluated playability: two of them used functional prototypes (Linehan et al., 2013; Ekman, 2007) and Quek and See (2015) used a test version.

**Figure 2.** Qualities of interaction under evaluation in the analyzed works.

The other 20 (86.96 %) studies combined interviews with at least one more method: surveys and questionnaires were the most common method in the combinations, being used in 18 (78.26%) studies. Observation of use was the second most applied (14 studies, 60.87%), following a common tendency to link interviews to observation (Lamont and Swidler, 2014).

Important to notice that only two (8.70%) studies (Chatzidimitris et al., 2016; Verdejo et al., 2010) used methods to monitor physiological data – both also applied interviews, observation, and interaction logs recording. A possible explanation for this is that the evaluation of an LBG, ideally, is performed outdoors or in environments that simulate mobility contexts, making it challenging to use some equipment types that capture these measures. It is also possible that difficulties linked to the context explain that only one work (Nilsson et al., 2016) used the Think Aloud protocol (Fonteyn et al., 1993). These cases illustrate some of the challenges that accompany the interaction evaluation in LBGs.

5.1.4 Data Analysis

Even though all the studies conducted interviews, fourteen (60.87%) did not inform how (or if) they analyzed the qualitative data. Three of these (13.04%) did not mention performing any analysis, be it quantitative or qualitative – two of them applied exclusively interviews (Quek and See, 2015;

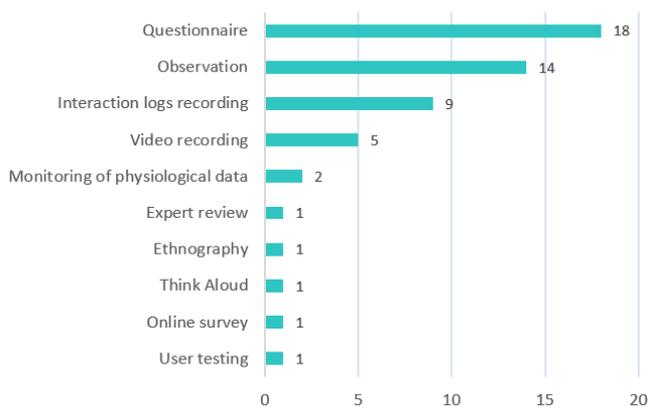


Figure 3. Methods used in combination with interviews in the evaluations reported by the analyzed papers

Ekman, 2007). The remaining nine (39.13%) studies described the analysis process with varying levels of detail and, in some cases, superficially. Only three papers mentioned qualitative and quantitative analysis, while the other six mentioned only qualitative or quantitative analysis (three studies each). Some techniques mentioned were: inductive thematic analysis, coding with ad hoc scheme, and affinity diagram. Two papers (Linehan et al., 2013; Diamantaki et al., 2011) stood out for providing more details on the qualitative analysis procedures of the data obtained in interviews.

5.2 Survey Results Overview

In the remainder of this section, we provide an overview of the expert opinion survey results performed with the authors of the 23 papers that reported using interviews to evaluate player interaction with LBGs.

5.2.1 Participants profile

The survey received responses from 14 authors from eight countries: Germany, Austria, Canada, Denmark, Greece, Indonesia, Mexico, and the United Kingdom. Twelve (85.71%) of them work in the HCI area, seven (50%) in Game Design, and six (42.86%) also work with LBGs. Only three participants (21.43%) listed GUR among their areas of activity. Seven (50%) participants have worked with game evaluation for ten years or more, five (37.71%) have between 1 and 6 years of experience, and two (14.29%) use this type of evaluation in their research only sporadically. Six (42.82%) authors mentioned being familiar with guidelines in the literature for interviews – the works of Steinar Kvale (2008) and Larry Wood (1997) were the most indicated – however, no papers mentioned using them. **Table 3** summarizes the participants' profiles.

5.2.2 Authors' motivations for using interviews

The main reasons mentioned were searching for a deeper understanding of the qualities evaluated and rich and detailed feedback. Another common reason was the awareness that the interviews could complement the data obtained with quantitative methods, especially when the objective was to investigate users' perceptions, behaviors, and emotions. For participant A13, the interviews allowed them to scrutinize

potentially interesting behavior patterns observed during usability tests. Other reasons listed were flexibility, direct contact with users, systematic feedback collection, and the use of interview data as a guide for interpreting quantitative data. Author A11 said that he chose semi-structured interviews because it is a quick and easy method to be applied outdoors. Accordingly, A6 pointed out that interviews allowed them to investigate in more detail critical issues observed in the tests because it offered more freedom to the participants, which generated valuable information that other methods could not generate. Since critical incidents can have real consequences for LBGs players – for example, errors in GPS or a misplaced point of interest can put the player in danger situations – it is crucial to investigate adverse conditions in this type of evaluation.

5.2.3 Elaboration of the interview script

All authors said that they applied semi-structured interviews in the referred studies. The primary approach for composing the script was authors elaborating it themselves, based on the evaluation and the study's objectives. Other common practices were to create the script during discussions between co-authors and to have more experienced colleagues responsible for this activity. Two authors adapted scripts from other studies, and one participant claimed to have used usability consultancy to create the script since game evaluation is not his area of expertise. Only three authors followed guidelines for applying interviews proposed in the literature – however, only one study (Kasapakis and Gavalas, 2017) reported this.

Most authors showed confidence in their previous experiences with the method, in discussions with colleagues, and in using the research objectives as a parameter to create the questions. Only two participants (A11 and A14) reported difficulties in preparing the script. They found it challenging to gather and address all relevant issues for the study. They also mentioned the difficulty in composing a solid script with a right balance between structure and flexibility so that the interviewer can follow the user's mental flow (which often produces unexpected insights), but still get proper comparators between the data.

5.2.4 Difficulties in conducting interviews

Five authors reported difficulties in conducting the interviews, highlighting the adversities inherent to the evaluation context (for example, low temperatures during tests performed outdoors). A common difficulty is that participants are more subject to distractions when the interview is conducted outdoors. The interviewer often has to call them back to the interview questions. In addition to that, A6 reported that, although interviewers get in-depth feedback on PX, users tend to talk a lot more about their views on the game aesthetic, making it more challenging to deal with questions. One of the most experienced authors, A14, pointed out the challenge of defining clear and effective communication between interviewer and interviewee, mainly when the evaluator aims to collect "technical" information from users who have little understanding of the subject, and highlighted the importance of a trained interviewer for situations like this.

Table 3. Survey respondents' profile

ID	Origin	N° of papers	Experience in games evaluation	Areas of expertise	Familiarity with guidelines
A1	Austria	1	≥ 10 years	HCI, Game Design	None of the presented options
A2	Germany	2	≥ 10 years	HCI, Game Design, Mobile Computing, Pervasive Games, LBGs	Not familiar
A3	Canada	1	≥ 10 years	HCI, Game Design, Gamification	Wood (1997)
A4	Mexico	1	≥ 10 years	HCI, GUR	None of the presented options
A5	Indonesia	1	4 - 6 years	HCI, Evaluation methods, Game Design, GUR, Gamification, LBGs	Kvale (2008), Boyce and Neale (2006)
A6	Greece	1	4 - 6 years	HCI, Evaluation methods, Game Design, GUR, Gamification, Mobile and Ubiquitous Computing, LBGs	None of the presented options
A7	UK	1	N/A	Investigative Interviews	None of the presented options
A8	UK	1	≥ 10 years	Information and Geographic Intelligence	Wood (1997)
A9	UK	1	1 - 3 years	HCI, Evaluation methods	Kvale (2008), Myers and Newman (2007), Chauncey (2014)
A10	UK	1	≥ 10 years	HCI, Gamification, LBGs	None of the presented options
A11	Denmark	1	1 - 3 years	HCI, Game Design, Gamification, Mobile Computing, LBGs	Kvale (2008)
A12	UK	1	N/A	HCI	None of the presented options
A13	UK	1	1 - 3 years	HCI, Ubiquitous Computing	Wood (1997)
A14	UK	1	≥ 10 years	HCI, Game Design, Mobile and Ubiquitous Computing, LBGs	None of the presented options

5.2.5 Particularities of interviews to evaluate LBGs

We asked authors if, in their interviews, they had explored questions related to any LBGs characteristics or specificities. Ten (71.43%) of them considered questions related to at least one specificity of the LBG to elaborate the interview script. In general, they indicated location as the characteristic most explored in studies, followed by mobility and social interaction in the game. These choices are in line with the characteristics most commonly attributed to LBGs in the literature – as we mentioned in Section 2, De Souza e Silva and Sutko (2011) point location, mobility, and sociability as the main characteristics of these games. Other specificities considered were context sensitivity, connectivity, pervasiveness, security, privacy, and GPS accuracy. The least explored were temporal and spatial expansion – although both appear in the literature as essential aspects of LBGs (Ahlqvist, 2018) – and level of physical effort.

Regarding the reuse of the interview script applied in the studies, most authors (85.7%) believe that they could not reuse it to evaluate other types of games since the interview's focus depends heavily on the research objectives. They also mentioned that the scripts used in these studies were very specific, applying only to the game and context in question. Two authors (A1 and A6) said they would reuse their scripts, making minor adjustments to include contextual aspects of the LBG and issues particular to the quality of interaction under assessment. A6 pointed out that although “playing an LBG” is arguably different from “playing a conventional game on a smartphone”, it is possible to use similar instruments to evaluate certain aspects in both games (e.g., the degree of realism of the graphics). However, when the evaluation focuses on subjective issues, such as PX, it is essential to consider each game's particularities, as they will dictate the necessary changes in the interview scripts for each case. In general, experts have attested to the benefits of using interviews in their studies.

6 Lessons to Use Interviews to Evaluate Player Interaction with LBGs

The results obtained with the first and second phases of our research allowed us to extract some lessons for conducting interviews to evaluate the player interaction in LBGs.

Together, these lessons form an initial set of guidelines that aims to provide researchers and practitioners, especially novice, with resources for the conscious and adequate execution of such activity. Some of them can be applied to the evaluation of games and interviews in general, while others are specific to LBGs. It is worth noting that the adequacy of the information presented here should be judged based on the particularities of each research and stage of the design process of an LBGs, and does not apply equally to all cases.

6.1 Divide to conquer: in search of better strategies to evaluate LBGs

The survey's results highlighted the importance of combining different methods, as already recommended in the liter-

ature (Creswell and Poth, 2018; Lazar et al., 2017). As expected, the respondents indicated to adopt this approach to evaluate LBGs in their studies, recognizing its value, since the isolated application of a single method does not usually provide the desired data. It was also clear that not all methods apply equally in this type of evaluation – at least not without some adaptation – since the characteristics of an LBG and its context of use can impose several restrictions on the evaluators. It is necessary to consider these issues when planning the evaluation strategies, keeping the focus on the study's objectives, qualities of the interaction to be evaluated, and characteristics of the LBGs. Some lessons related to these issues are:

- L1:** Remember that the specificities of LBGs impact and distinguish the PX in these games. Identify and understand the main characteristics of the LBG under evaluation and use that information to outline strategies that consider and explore them. Choose and adapt the data collection methods keeping in mind that the LBG specificities will influence the conduction of the evaluation and the results obtained.
- L2:** Combine quantitative and qualitative methods, complementing their limitations, but avoid applying methods or equipment that mischaracterize or limit the player's experience with the specificities of the evaluated LBG. For instance, be cautious of using equipment that limits the player's mobility, since it plays a crucial role in LBGs gameplay.
- L3:** Prioritize evaluation in external environments, including and exploring the inherent difficulties of that context (e.g., climatic conditions, distractions, GPS inaccuracy). These factors are part of the actual LBG gaming experience and should be considered in an assessment that seeks a realistic view of PX.
- L4:** When possible, combine laboratory research with real context research to gain a deeper understanding of the PX – if that is what you are aiming for (Drachen et al., 2018).
- L5:** When planning an evaluation, consider the drastic changes that can occur in the context of mobile games (especially concerning LBGs) (Drachen et al., 2018).

6.2 Know to apply: how to enhance the planning and use of interviews to evaluate LBGs?

The use of interviews usually presents different challenges to the evaluators. In addition to that, since an often unpredictable context permeates LBGs, common issues can gain new nuances (for instance, a “minor problem” of delay in map update can take the player to an unwanted space, which can be a big problem). It increases the importance of developing precise and efficient interview scripts that offer both interviewer and interviewee flexibility and freedom to explore unpredicted situations. Thus, to best use this method and ensure the achievement of evaluation objectives, it is necessary to know its potentials and limitations to plan its proper use and execute it in a truly beneficial way.

- L6:** Semi-structured interviews are a good choice to evaluate PX in LBGs. They offer flexibility and freedom for users to express themselves while allowing the interviewer to take advantage of opportunities and further investigate specific points, maintaining some structure to make systematic data collection feasible.
- L7:** When preparing the script for the interview, focus on the objectives of the evaluation, but also prioritize the examination of specific issues of LBGs – such as mobility, spatiality, temporal expansion - and their influences on PX (e.g., How physical effort affects player’s fun?).
- L8:** Identify which evaluation goals can be achieved through other methods and explore in the interview those that require further examination (for example, subjective questions and clarification of doubts). Some of the advantages of interviews to evaluate LBGs are practicality and speed, so be objective and simplify its application.

6.3 Interviews serve multiple purposes: the importance of taking advantage of this when evaluating LBGs

The authors pointed out several motivations for applying interviews in their studies. Versatility, flexibility, and (relative) speed in the application, for instance, make this method valuable for evaluations carried out outdoors. Interviews have proved to be a useful tool for LBGs evaluation, especially when investigating PX, as they serve various purposes and provide rich data that can be used in different ways (for example, to guide interpretation of quantitative data). Some lessons learned are:

- L9:** Use interviews to investigate how the specificities of LBG affect PX in the player’s view. Seek to collect detailed information about how he perceives and deals with the characteristics of the game and its specificities (e.g., security, effort, social image). This practice can assist in obtaining data that, among others, can generate guidelines of design for these games.
- L10:** Combine interviews with user observation, for example, to investigate critical events (including those caused by external factors) and understand how players perceive, interpret, and react to them. These events can modify how the player interacts with the LBG, creating game dynamics that were not foreseen by game designers.
- L11:** Use interviews to investigate how players integrate an LBG into their lives and its impact on the player’s perspective and experience. Since LBGs are played amid daily activities, this kind of information is essential to improve PX and game dynamics.
- L12:** In LBGs, the mix between real everyday social norms and the game world’s rules needs to fit perfectly to support integration between the physical and virtual worlds. Use the interview to investigate issues connected to this relation (Jegers, 2007)
- L13:** LBGs should allow a smooth and fluid transition between different playing contexts and not imply or re-

quire actions from the player that could result in a violation of social norms in everyday contexts. These games should also allow the player to shift his focus between the physical and virtual parts of the game world without losing full immersion in the game. Interviews are a powerful tool to investigate whether a game has these characteristics (Jegers, 2007).

6.4 With great power comes great responsibility: interviews can be expensive, but worth the price

Interviews are an effective method for assessing the interaction between player and LBG - 92.86% of the experts confirmed the method’s effectiveness to achieve their goals. The use of interviews can result in several benefits for a study, such as direct contact with users, obtaining feedback that would not be captured with other methods, a greater understanding of the players’ perceptions, and support in interpreting and verifying the data collected with quantitative methods. However, interviews should not be seen as an effortless way to obtain good results. Its proper use requires effort and rigor, and it comes with some limitations and pitfalls, like any other method. As attested by experts, interviews require planning, preparation, and can result in poor feedback if conducted frivolously. Besides, data analysis is often time-consuming and laborious.

- L14:** The interviewers must be familiar with the characteristics and specificities of the evaluated LBG so they can recognize and explore opportunities in the users’ speeches and actions, allowing them to investigate better the PX offered by the game.
- L15:** A good question should contribute to the production of knowledge and promote good interaction in the interview. In this way, it is crucial to choose the questions to be asked in the interview carefully (Spradley, 2016).
- L16:** LBGs gives players great freedom, which increases the variability of PX. Despite this, data must be collected systematically to ensure useful comparison parameters. When interviewing a user, ask questions about aspects of the experience that are common to all players – the characteristics of the evaluated LBG can generate topics to be consistently explored, for example.
- L17:** When conducting an interview and taking notes, try to categorize your notes into sections during the session – it also serves for observation. This practice provides a useful shortcut to analyzing the collected data and can reduce the time spent reporting the results (Drachen et al., 2018).
- L18:** Combine interviews with methods that capture data during the game and explore gameplay highlights in the interview questions to get more meaningful results. During the analysis, these data can be crossed and used to prove the users’ statements with their actions.

6.5 Practice only makes for improvement: master the craft

Qualitative research can be a tricky task depending on the level of experience a researcher may have with a particular type of methodology (Turner III, 2010). Logically, it also applies to the interview. Contrary to what many people may think, it is not a regular conversation with a predefined topic, and it can even result in frustration and a waste of effort and time. To avoid this risk, the interviewer must appropriately train interviewing techniques to obtain the most detailed and rich data from users (Boyce and Neale, 2006).

- L19:** These techniques include avoiding yes/no and leading questions, using appropriate body language, and keeping their personal opinions in check (Spradley, 2016; Boyce and Neale, 2006).
- L20:** While one of the most important rules about asking questions is to keep quiet and give the interviewee room to talk, it is equally important to show appreciation and interest in what they say. They must make that interviewee comfortable and appear interested in what they are saying. It means that the interviewer should be sensitive to guide the user along with the planned topics without controlling them too much to avoid losing important and unexpected points. Beyond the technical and academic knowledge, only practice will reward an interviewer with the perception and sensitivity necessary to perform this task.

7 Challenges in Using Interviews to Evaluate the Player Interaction with LBGs

During the compilation of the survey's results, we also identified some challenges that reflect research opportunities, especially for the HCI and GUR communities. We present some of them as insights to be discussed and further expanded in hopes of maturing the practice of GUR to study and comprehend player interactions and experiences with LBGs.

- How can we plan evaluation strategies that adequately and systematically explore the characteristics and specificities of the evaluated LBG?
- How to structure the systematic application of interviews for the evaluation of PX with LBG?
- How can an interviewer ensure the systematic collection of meaningful information without compromising the flexibility of the interview?
- How to make the analysis process more practical and straightforward so evaluators can apply interviews in contexts that require higher speed, such as in the industry?
- How to guide novice researchers and practitioners to plan and conduct qualitative interviews to evaluate LBGs considering its specificities?

8 Conclusion and Future Work

Evaluating the player's interaction with an LBG is still a complex task, given these games' peculiar characteristics. In this context, this paper reported the analysis of 23 studies that used interviews to evaluate LBGs and the conduction of an expert opinion survey with the authors of these studies. We aimed to identify common practices and translate them into actionable recommendations to interested researchers and practitioners. Based on the obtained knowledge, we presented lessons and challenges – emphasizing the use of interviews – which point to the importance of considering LBGs specificities and investigating their effects on PX when evaluating the interaction. The information presented here is useful for assisting students, researchers, and practitioners, especially the novice, in this task and encouraging them to reflect and act on these issues.

Concerning the use of interviews, we observed that, although there are plenty of works on this subject, a minimal portion of the experts used these directions in the studies we analyzed. This fact does not allow one to question the quality of the interviews conducted in their studies, but it opens space for reflections about LBGs evaluation practice. Since the scientific community deals with issues beyond practice, it seems relevant to investigate the reasons behind this *modus operandi*. It is possible, for instance, that practitioners judge the existing guidelines inadequate to the actual LBGs context or that the HCI evaluation is treated as something secondary in these studies. Either way, it is necessary to investigate these issues to promote the area's maturing and offer tools and resources capable of performing more accurate and expressive evaluations in this domain.

Therefore, we intend to evolve and expand the lessons presented here to compose a guide for qualitative evaluation of PX in LBGs, to provide directions for planning and conducting these evaluations in a consistent and structured manner. In our survey, we asked participants about such a proposal, and they considered it positive, especially for the benefit of students and beginning practitioners. Thus, we will continue with our efforts in this direction.

Acknowledgements

This paper is an extension of the paper **Uso de entrevistas na avaliação de jogos baseados em localização: Lições e desafios** (Carneiro et al., 2019a) published in the *1st Workshop on Interaction and User Research on Game Development (WIPlay 2019)*. We thank the editors for the invitation to expand the original paper.

References

- Abeele, V. V., Spiel, K., Nacke, L., Johnson, D., and Gerling, K. (2020). Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies*, 135:102370.
- Ahlqvist, O. (2018). Location-based games. *International*

- Encyclopedia of Geography: People, the Earth, Environment and Technology*, pages 1–6.
- Alha, K., Koskinen, E., Paavilainen, J., and Hamari, J. (2019). Why do people play location-based augmented reality games: A study on pokémon go. *Computers in Human Behavior*, 93:114–122.
- Baillie, L., Morton, L., Uzor, S., and Moffatt, D. C. (2010). An investigation of user responses to specifically designed activities in a multimodal location based game. *Journal on Multimodal User Interfaces*, 3(3):179–188.
- Ballagas, R., Kuntze, A., and Walz, S. P. (2008). Gaming tourism: Lessons from evaluating reexplorer, a pervasive game for tourists. In *International Conference on Pervasive Computing*, pages 244–261. Springer.
- Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., and Row-Farr, J. (2006). Can you see me now? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(1):100–133.
- Blandford, A., Furniss, D., and Makri, S. (2016). *Qualitative HCI Research: Going Behind the Scenes*. Morgan & Claypool Publishers.
- Blum, L., Wetzel, R., McCall, R., Oppermann, L., and Broll, W. (2012). The final timewarp: using form and content to support player experience and presence when designing location-aware mobile augmented reality games. In *Proceedings of the designing interactive systems conference*, pages 711–720. ACM.
- Boyce, C. and Neale, P. (2006). Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input.
- Calvillo-Gómez, E., Crisp, J., and Romero, P. (2011). Empirical validation of the involvement component of the pervasive gameflow model. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, page 48. ACM.
- Carneiro, N., Darin, T., and Viana, W. (2019a). Uso de entrevistas na avaliação de jogos baseados em localização: Lições e desafios. In *Anais do I Workshop sobre Interação e Pesquisa de Usuários no Desenvolvimento de Jogos*, pages 59–68. SBC.
- Carneiro, N., Darin, T., and Viana, W. (2019b). What are we talking about when we talk about location-based games evaluation? a systematic mapping study. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, pages 1–13.
- Chatzidimitris, T., Gavalas, D., and Michael, D. (2016). Soundpacman: Audio augmented reality in location-based games. In *18th Mediterranean Electrotechnical Conference (MELECON)*, pages 1–6. IEEE.
- Chauncey, W. (2014). Interview techniques for ux practitioners. *A User-Centered Design Method*.
- Chittaro, L. and Sioni, R. (2012). Turning the classic snake mobile game into a location-based exergame that encourages walking. In *International Conference on Persuasive Technology*, pages 43–54. Springer.
- Crawford, S. B., Monks, S. M., and Wells, R. N. (2018). Virtual reality as an interview technique in evaluation of emergency medicine applicants. *AEM education and training*, 2(4):328–333.
- Creswell, J. W. and Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- Darin, T., Andrade, R., and Sánchez, J. (2019). Principles for evaluating usability in multimodal games for people who are blind. In *International Conference on Human-Computer Interaction*, pages 209–223. Springer.
- De Souza e Silva, A. and Sutko, D. M. (2009). Digital cityscapes: Merging digital and urban playspaces.
- De Souza e Silva, A. and Sutko, D. M. (2011). Theorizing locative technologies through philosophies of the virtual. *Communication Theory*, 21(1):23–42.
- Diamantaki, K., Rizopoulos, C., Charitos, D., and Tsianos, N. (2011). Theoretical and methodological implications of designing and implementing multiuser location-based games. *Personal and Ubiquitous Computing*, 15(1):37–49.
- Drachen, A., Mirza-Babaei, P., and Nacke, L. E. (2018). *Games user research*. Oxford University Press.
- Ekman, I. (2007). Sound-based gaming for sighted audiences—experiences from a mobile multiplayer location aware game. In *Proceedings of the 2nd audio mostly conference*, pages 148–153.
- El-Nasr, M. S., Durga, S., Shiyko, M., and Sceppa, C. (2015). Data-driven retrospective interviewing (ddri): a proposed methodology for formative evaluation of pervasive games. *Entertainment Computing*, 11:1–19.
- Fonteyn, M. E., Kuipers, B., and Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative health research*, 3(4):430–441.
- Holmes, S. (2019). Play-based 7 interview techniques with young children. *Using Innovative Methods in Early Years Research: Beyond the Conventional*, page 92.
- Isbister, K. and Schaffer, N. (2008). *Game usability: Advancing the player experience*. CRC press.
- Jegers, K. (2007). Pervasive game flow: understanding player enjoyment in pervasive gaming. *Computers in Entertainment (CIE)*, 5(1):9–es.
- Kasapakis, V. and Gavalas, D. (2015). Pervasive gaming: Status, trends and design principles. *Journal of Network and Computer Applications*, 55:213–236.
- Kasapakis, V. and Gavalas, D. (2017). User-generated content in pervasive games. *Computers in Entertainment (CIE)*, 16(1):3.
- Kiefer, P., Matyas, S., and Schlieder, C. (2006). Systematically exploring the design space of location-based games. In *4th International Conference on Pervasive Computing*, pages 183–190.
- Kitchenham, B. A., Budgen, D., and Brereton, O. P. (2010). The value of mapping studies—a participant-observer case study. In *EASE*, volume 10, pages 25–33.
- Kurczak, J., Graham, T., Joly, C., and Mandryk, R. L. (2011). Hearing is believing: evaluating ambient audio for location-based games. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, page 32. ACM.
- Kvale, S. (2008). *Doing interviews*. Sage.
- Lamont, M. and Swidler, A. (2014). Methodological plural-

- ism and the possibilities and limits of interviewing. *Qualitative Sociology*, 37(2):153–171.
- Lankes, M. and Stiglbauer, B. (2016). Gazeat: Mobile gaze-based interaction in the context of augmented reality games. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 397–406. Springer.
- Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.
- Leorke, D. (2018). *Location-based gaming: play in public space*. Springer.
- Lindt, I., Ohlenburg, J., Pankoke-Babatz, U., and Ghellal, S. (2007). A report on the crossmedia game epidemic menace. *Computers in Entertainment (CIE)*, 5(1):8.
- Linehan, C., Bull, N., and Kirman, B. (2013). Bollocks!! designing pervasive games that play with the social rules of built environments. In *International Conference on Advances in Computer Entertainment Technology*, pages 123–137. Springer.
- McCall, R. and Braun, A.-K. (2008). Experiences of evaluating presence in augmented realities. *Psychology*, 6(2):157–163.
- McCall, R., Wetzel, R., Löschner, J., and Braun, A.-K. (2011). Using presence to evaluate an augmented reality location aware game. *Personal and Ubiquitous Computing*, 15(1):25–35.
- Merritt, T., Nielsen, C. L., Jakobsen, F. L., and Grønbaek, J. E. (2017). Glowphones: designing for proxemics play with low-resolution displays in location-based games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 69–81. ACM.
- Myers, M. D. and Newman, M. (2007). The qualitative interview in is research: Examining the craft. *Information and organization*, 17(1):2–26.
- Nacke, L. E., Drachen, A., and Göbel, S. (2010). Methods for evaluating gameplay experience in a serious gaming context. *International Journal of Computer Science in Sport*, 9(2):1–12.
- Nacke, L. E., Moser, C., Drachen, A., Mirza-Babaei, P., Abney, A., and Zhenyu, Z. C. (2016). Lightweight games user research for indies and non-profit organizations. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 3597–3603. ACM.
- Nilsson, T., Hogsden, C., Perera, C., Aghaee, S., Scruton, D. J., Lund, A., and Blackwell, A. F. (2016). Applying seamless design in location-based mobile museum applications. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 12(4):56.
- Oppermann, L., Schaal, S., Eisenhardt, M., Brosda, C., Müller, H., and Bartsch, S. (2017). Move, interact, learn, eat—a toolbox for educational location-based games. In *International Conference on Advances in Computer Entertainment*, pages 774–794. Springer.
- Ozok, A. A. (2009). Survey design and implementation in hci. *Human-Computer Interaction: Development Process*, 253.
- Paavilainen, J., Korhonen, H., Alha, K., Stenros, J., Koskinen, E., and Mayra, F. (2017). The pokémon go experience: A location-based augmented reality mobile game goes mainstream. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 2493–2498. ACM.
- Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18.
- Prasetio, Y. L., Wijaya, R., Sjah, M. P., Christian, M. R., and Chowanda, A. (2017). Location-based game to enhance player’s experience in survival horror game. *Procedia computer science*, 116:206–213.
- Preece, J., Sharp, H., and Rogers, Y. (2019). *Interaction design: beyond human-computer interaction*. John Wiley & Sons.
- Pyae, A., Mika, L., and Smed, J. (2017). Understanding players’ experiences in location-based augmented reality mobile games: A case of pokémon go. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, pages 535–541. ACM.
- Quek, A. and See, J. (2015). Obscura: A mobile game with camera based mechanics. In *2015 Game Physics and Mechanics International Conference (GAMEPEC)*, pages 21–25. IEEE.
- Rowe, G. and Wright, G. (2001). Expert opinions in forecasting: the role of the delphi technique. In *Principles of forecasting*, pages 125–144. Springer.
- Rowley, J. (2012). Conducting research interviews. *Management research review*, 35(3/4):260–271.
- Salen, K., Tekinbaş, K. S., and Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT press.
- Sánchez, J. L. G., Vela, F. L. G., Simarro, F. M., and Padilla-Zea, N. (2012). Playability: analysing user experience in video games. *Behaviour & Information Technology*, 31(10):1033–1054.
- Sandham, A., Ormerod, T., Dando, C., Bull, R., Jackson, M., and Goulding, J. (2011). Scent trails: Countering terrorism through informed surveillance. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 452–460. Springer.
- Schell, J. (2014). *The Art of Game Design: A book of lenses*. AK Peters/CRC Press.
- Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., and Diakopoulos, N. (2016). *Designing the user interface: strategies for effective human-computer interaction*. Pearson.
- Spradley, J. P. (2016). *The ethnographic interview*. Waveland Press.
- Sra, M. and Schmandt, C. (2015). Expanding social mobile games beyond the device screen. *Personal and Ubiquitous Computing*, 19(3-4):495–508.
- Turner III, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The qualitative report*, 15(3):754–760.
- Verdejo, A. J., De Moor, K., Ketyko, I., Nielsen, K. T., Vanattenhoven, J., De Pessemier, T., Joseph, W., Martens, L., and De Marez, L. (2010). Qoe estimation of a location-

based mobile game using on-body sensors and qos-related data. In *2010 IFIP Wireless Days*, pages 1–5. IEEE.

Wang, A. I., Forberg, S., and Øye, J. K. (2016). Knowledge war-a pervasive multiplayer role-playing learning game. In *European Conference on Game Based Learning (ECGBL 2016)*. ACPI, Glasgow, Scotland.

Wood, L. E. (1997). Semi-structured interviewing for user-centered design. *interactions*, 4(2):48–61.

A Questions in the survey

This appendix enlists the questions on the survey sent to the 23 papers' authors – the asterisk indicates an open question.

A.1 Researcher/practitioner profile

1. What are your areas of study/research? (Feel free to add other options)
 - As options, this question provides a vast list of areas and research topics (such as HCI, Gamification, Game Design, GUR, Mobile Computing, and Communication technologies) and allows the respondent to add others.
2. For how long have you been researching/working with games evaluation?
 - (a) Occasionally
 - (b) Less than 1 year
 - (c) 1-3 years
 - (d) 4-6 years
 - (e) 7-9 years
 - (f) 10 years or more
 - (g) I don't work with this topic.
3. Are you familiar with any specific guidelines or recommendations from the literature to compose interview scripts and/or to conduct interview sessions?
 - (a) Steinar Kvale's guidelines (2007)
 - (b) Daniel W. Turner's practical guide for novice investigators (2010)
 - (c) Carolyn Boyce's guide for designing and conducting in-depth interviews for evaluation input (2006)
 - (d) Michael D. Myers and Michael Newman's dramaturgical model of the qualitative interview (2006)
 - (e) Chauncey Wilson's interview techniques for UX practitioners (2014)
 - (f) The semi-structured interview techniques of Larry Wood (1997)
 - (g) None of these
 - (h) I'm not familiar with any particular set of guidelines/recommendations to interviews.
 - (i) Other – *allows typing an answer*

A.2 Experience applying interviews

1. What type of interview did you use in this work?

- (a) Fully structured interview - i.e., it uses a rigid script to present questions in a well-defined order.
 - (b) Semistructured interview - i.e., it generally starts with a set of questions, but allows the interviewer to dig through the interviewee's comments, trying to gain additional insights and understanding.
 - (c) Unstructured interview - i.e., it may simply be based on a list of topics or questions known as an interview guide. The interviewer starts with an initial question and the interviewee responds as she sees fit, discussing topics of her choosing.
 - (d) I do not know how to define it.
 - (e) Other – *allows typing an answer*
2. Why did you decide to use interviews in your research?*
 3. How did you compose your interview script?
 - (a) I found a script in the literature and made some adaptations.
 - (b) I created the questions according to the evaluation goals of the project.
 - (c) I have some experience in interviewing users, so I knew what to ask.
 - (d) I researched some important topics and used them as a guide.
 - (e) I discussed with my colleagues and we elaborate the questions.
 - (f) I followed a set of guidelines/recommendations to elaborate the script.
 - (g) Other – *allows typing an answer*
 4. Did you face any difficulties in creating the interview script? If so, which one(s)?*
 5. In your interview, did you make any questions related to specific characteristics of the location-based games or common in-game issues, such as:
 - (a) Mobility
 - (b) Location
 - (c) Connectivity
 - (d) Spatial expansion
 - (e) Temporal expansion
 - (f) Context awareness
 - (g) Pervasiveness
 - (h) In-game social interaction
 - (i) GPS imprecision
 - (j) Excessive physical effort
 - (k) Safety
 - (l) Privacy issues
 - (m) None of these
 - (n) I do not remember
 - (o) Other – *allows typing an answer*
 6. Did you face any difficulties in applying the interviews in this context? If so, which one(s)?*
 7. Would you say that the interview accomplished the research goals you've had established?
 - (a) Yes, it was beneficial for my research results.
 - (b) No, I felt like it was a waste of time.
 - (c) I'm not sure about it.
 - (d) Other – *allows typing an answer*

8. Can you please elaborate on that idea?*
9. If you were to evaluate another type of digital game – a mobile game (not location-based), for example – would you use the same interview script you used to evaluate the location-based game?
 - (a) Yes, I think I could use the same script, because the same questions apply.
 - (b) I would use the same script but with some minor adjustments.
 - (c) No, I think it would be necessary to create a different script.
10. If you chose the 2nd or the 3rd option, can you please tell us why and/or how the scripts should be different?*
11. What are the main advantages and/or drawbacks you've realized in using interviews in this work?*
12. Is there any other consideration you would like to add regarding your experience of using interviews to evaluate location-based games?*

A.3 Your opinion

1. Would you consider [the described] solution useful?
 - 7-point Likert scale
2. Please let us know more about why you gave this answer to the previous question or any other related consideration.*