

First Time User Experiences in Mobile Games: An Evaluation of Usability

Lawrence Barnett^{1†}, Carlo Harvey^{2‡}, Christos Gatzidis^{3†}

[†]*Bournemouth University*, [‡]*Birmingham City University*

Abstract

Unlike most other mobile applications, games are driven by their user experience rather than their functionality. No one wishes to play games that are either frustrating or difficult for the wrong reasons. Usability is an integral part of software development and is about maximizing the effectiveness, efficiency and satisfaction of the user. The delicacy of the user experience and heavy competition it can be argued render usability more important in games than it is in other software. Immersion and engagement are fundamental and core parts of the enjoyment of computer games, and both are dependent on usability. The focus of this article is around a framework for evaluating the usability of First Time User Experiences (FTUEs). Investigating two specific, off-the-shelf games, we demonstrate that the FTUE can affect an element of usability, namely ‘information quality’, when controlling for the guidance and information presented. Despite this, overall usability is unaffected by the presence of the FTUE.

Keywords: Usability, First Time User Experience (FTUE), Game Design

1. Introduction

Design heuristics aim to create and establish a fundamental/native usable system, aiding the visceral and primitive nature of the users’ experience. However, beyond the fundamental design of an application, usability can be aided

¹i7244586@bournemouth.ac.uk, ²carlo.harvey@bcu.ac.uk, ³cgatzidis@bournemouth.ac.uk

through effective guidance and teaching, often referred to as ‘onboarding’ [1]. In this article, we will be exploring the First Time User Experience and, specifically, the use of FTUEs embedded in games on mobile devices. This is towards discovering how, and indeed if, these are effective at increasing usability.

To achieve this goal, we show an evaluation of FTUEs in a mobile gaming context. In particular, this article considers the effect of usability across very distinct game genres and provides an analysis across various scales of usability. Specifically, we make the following contributions:

- We provide a framework for evaluating the usability of FTUEs of mobile games
- We demonstrate that elements of usability are influenced by the guidance and information a player receives
- We make suggestions for designers to adhere to certain usability heuristics as a result of this finding

It should be noted that this article is an extended version of our conference paper, for this see Barnett *et. al.* [2].

2. Related Work

2.1. Usability and Games

Usability, as defined by ISO 9241-11 (Guidance on usability) is termed as “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” [3]. As opposed to tools and software, where production and user productivity is paramount, games are played for a variety of reasons, most of which are rooted in fun and enjoyment. This key distinction arguably changes the weighting of the three areas identified above by ISO 9241-11, from an equal weighting to a hierarchy. Satisfaction needs to be prioritised, with efficiency and effectiveness following. Sauro and Kindlund [4] concluded that effectiveness can be

measured in completion rates and errors, efficiency from time on task, and satisfaction using any of a number of standardised satisfaction questionnaires. This facilitates numerical foundation to ascertain a weighted model under which to conduct usability studies. In the following passage, we contextualise the three areas defined by ISO 9241-11 (effectiveness, efficiency and satisfaction) for our interest in games [3].

First of all, on satisfaction; enjoyment and fun can be seen as the primary and sole motivation for an individual to engage in a computer/video game. Myers' study of Game Player Aesthetics [5], identified "challenge" as "the most preferred characteristic", highlighting balance as an important variable to tune regarding satisfaction. Myers' finding supports and provides strong reasoning for the use of the widely accepted heuristic of creating an interface and control method that can be learned, used and mastered with as little resistance as possible, preparing and enabling the player to enjoy and utilise all available mechanics and, ultimately, strategies [5]. Optimising the complexity and interactions of an interface can aid escapism and support immersion [6]. As for effectiveness; in the context of games, this can be attributed to how accurately and effectively the players can express themselves via the available interface and interactions to achieve specific goals, achievements or desires. In a similar way, efficiency in games usability represents the relationship between the inputs and interactions, plus the success on specific goals, achievements or desires. The inputs may require considerable dexterity in order to enable the player to achieve success, or they may be achievable with comparatively little skill.

Several researchers have investigated the concept of a model-based approach to address elements of usability and suggest meaningful reform in game design. Sweetser and Wyeth [7] presented a model called GameFlow. GameFlow was designed to identify enjoyment within game play. It was shown to be able to successfully identify the elements of strength and weakness and can be used to more generally assess other games. This model was evaluated only on games of the real-time strategy genre. Nacke [8] suggested a hierarchical model of game usability. This model was designed to account for a range of measurable entities,

from concrete to abstract. These entities can be described from theoretical construct to practical implementation. However, this model was not validated in its applicability to game development.

2.2. FTUEs and Onboarding

With usability contextualised to our interest in games, we can begin to discuss the effects of usability in games. As represented in Adams' Story Engine Diagram, the interface is the source of both the input and output [9]. Furthermore, in the Mechanics, Dynamics and Aesthetics (MDA) framework, it can be noted that the aesthetics of a game are the first and foremost of its elements to be experienced by the player [10]. Usability affects the player's immediate and most intimate mechanism, allowing all of the game's elements to function and ultimately be enjoyed. Schell [6] describes and illustrates the importance of designing and building effective interactive systems in games. Schell's recommendations are also echoed in Google's User Experience Principles [1]. When considering FTUEs, the first few minutes of play are especially critical as these minutes of play typically evidence substantial churn rates for new players.

Petersen *et. al.* [11] performed an analysis of the onboarding phase of several mobile games. This was conducted using a study to provide insights for evaluating the user experience of onboarding phases in mobile games. The research made use of objective metrics through the form of physiological measures and from these observations suggested recommendations for design elements that resulted in high arousal. No empirical link was established between high arousal and increased onboarding however. Additionally, the valence (either positive or negative) of a detected event could have been created by external factors, making physiological response data tricky to evaluate in this context.

3. Methodology

3.1. Design

The Independent variable for the research is the following; guidance and information via a first-time user experience, expressed or presented before or

during gameplay. The Dependent variable (Outcome) is the Usability of the mobile game.

This research is to be tested with two groups, control and treatment. These groups represent, respectively, either the presence or absence of guidance via a FTUE in the two selected games (more on these in the next section of this article). The independence of the groups means a participant placed in the control group will play both games with no guidance via a FTUE, and likewise for the treatment group. The approach of independent groups, and between-subject designs, where the participant is only exposed to one condition, was employed in response to the increased bias, and confounding factors presented in the alternative design of within subject design [12].

The null hypothesis is given as \mathbf{H}'_0 , that all conditions are equal under testing ($\mathbf{H}'_0: \pi_i = \frac{1}{2}$). The alternative being that not all the conditions π_i are equal. This is considered to be that guidance and information via a First Time User Experience does not affect the usability of mobile games. A number of alternative hypotheses were considered based on the literature:

H_a: the control and treatment conditions would produce different results.

H_b: guidance and information provided would influence usability.

H_c: various elements of usability would be influenced by guidance and information.

Game usability methods employed by game studios in Northern Europe include the following; gameplay testing, observation of gameplay, usability testing, focus groups, interviews, think-aloud approaches, filmed play sessions, questionnaires, and data logging. Usability Questionnaires, which were found to be utilised by 38% of studios [13], are chosen as the vehicle to facilitate this study. Specifically, we adopt and adapt the IBM PSSUQ [14]. The questions used in our study, as adapted, are as follows:

1. Overall, I am satisfied with how easy it is to play this game
2. It was simple to play this game
3. I could effectively complete the objectives and challenges

4. I was able to complete objectives and challenges quickly
5. I was able to efficiently complete objectives and challenges
6. I felt comfortable using this system
7. It was easy to learn to play this game
8. Whenever I make a mistake in the game, I recover easily and quickly
9. The organisation of information on the game screens is clear
10. The interface of this game is pleasant
11. I like using the interface of this game

As posited by Lewis [14], we investigate across 4 distinct subcategories of usability that can be examined at the various levels. These are the overall variable ‘OVERALL’ (Overall Usability), plus the more specific variables, ‘SYSUSE’ (System Use), ‘INFOQUAL’ (Information Quality) and ‘INTERQUAL’ (Interface Quality).

3.2. Materials

3.2.1. Equipment

Two games were selected to review the effect of guidance upon usability; ‘Super Mario Run’ and ‘Linia’. The games were selected based on their similar yet contrasting interaction complexity, since they can both be controlled with one finger. However, the combinations and precision of interactions, along with other gameplay manipulations such as pace, challenge the player’s inputs past the seemingly simple one-touch interaction.

It should also be noted that these two specific games were chosen as they both are clearly of a different genre. The first is a side-scrolling, platform game whilst the latter is a puzzle game. These are (both) off-the-shelf games, so the FTUEs they both showcase were used as presented in the original, commercial offerings (i.e. no changes have been made to their FTUEs for the purposes of the experiment). The detailed operation of these individual FTUEs for each title are presented in the following section. To further elaborate on the earlier point about the selection of these two games; other than the cross-genre approach and

the similarity in the simplicity of controls, these two games were both released on the same year (2016) yet were derived from very different production values (one is from Nintendo, a well-established large developer with a rich history and the other from an independent developer). So, we feel that for all the reasons detailed above, a contrasting study using these two games, in the context we have chosen, will yield both interesting and also valuable results.

The experiment was conducted on an iPhoneSE. The device's specification was the following; iOS operating system, a 4-inch screen size, pixel density of 326 pixels per inch, an A9 chipset featuring a dual-core 1.84GHz processor with 64-bit architecture and, finally, 2GB of RAM.

3.2.1.1 Games

The controls and interactions found in Super Mario Run adopt an avatar-based interaction model, with an overarching design similar to that of the SEGA Sonic game franchise, with an ever-running avatar, where the player is presented with one input/control method. The one input; touching anywhere on the screen (other than the two UI elements) will action the avatar to perform a jump. The jump is manipulated based on the timing and environment in which it is performed. Information, interface and feedback, consist of a side/scrolling avatar-based camera, following the avatar at a fixed rate and position from a set distance. The user interface consists of a timer, coin collection status, remaining bubbles (extra lives) and, finally, a pause and respawn button, so there is very little to distract and take the attention of the player away from the core gameplay. The difficulty curve of the overall game increases at a steady rate after the initial level, though within the individual levels the difficulty is varied to add drama and tension, as well as challenge, through the manipulation of the frequency, distribution and positioning of obstacles, enemies and rewards. The guidance via a FTUE present in Super Mario Run is intrusive and involuntary. Upon initial entry to the game the player enters a modified, scripted version of Level 1. The modified level uses scripted events, out of the player's control, to

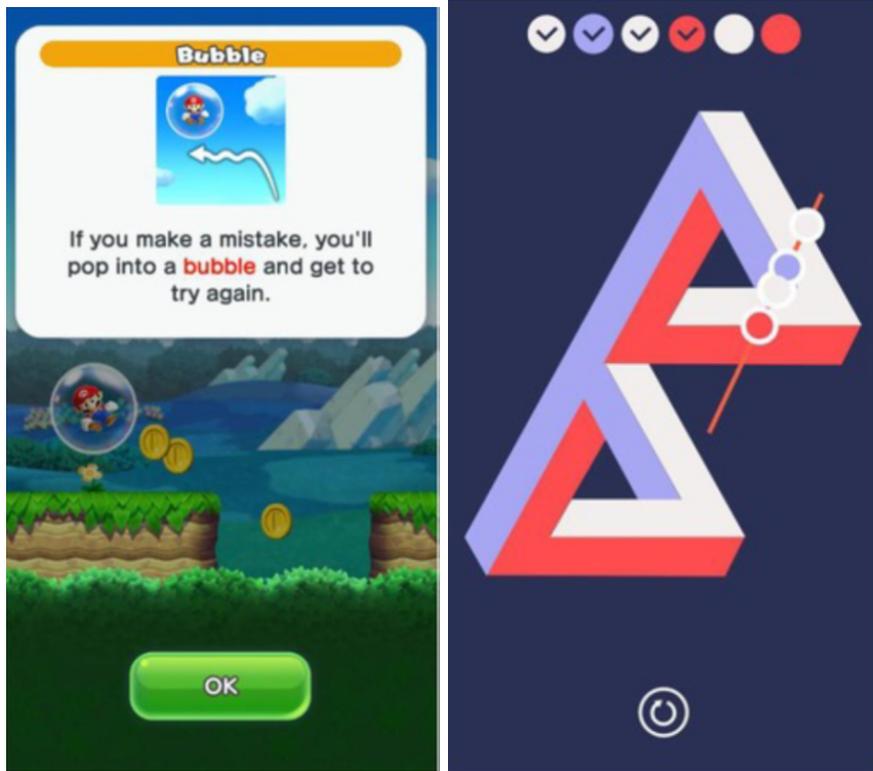


Figure 1: Demonstrating the visual information presented in both Super Mario Run (l) and Linia (r)

force circumstances, events and certain interactions to help educate the player. Through the modified game world, the game highlights individual mechanics, concentrating the player's attention to each sequentially. An example of this; through disabling the jump mechanics, the player inevitably falls to their death, then the respawn/life mechanic is activated (see Figure 1), leading the player back to safety, subsequently requiring the player's input to fall to the ground and continue gameplay. This forced, scripted behaviour may be effective in teaching. However, it strips the player of autonomy and agency in the process, eliminating any opportunity for the player to learn by their own accord.

Linia's controls and interactions adopt a contestant based interaction model, with no avatar present, as the static camera faces towards the specific geometry

elements. The player is allowed to draw a line between a start and end position by touching and dragging on the screen, creating a visible line on the user interface. When the player releases their touch, the line remains drawn, and the intercepted geometry is highlighted (seen in Figure 1), leading to a win or lose state. Linia’s interface, information and feedback employ a minimalist approach. With no avatar present, the aforementioned static camera is used to present the player with the various geometric scenarios. Excluding pausing and quitting, there are no available actions to the player other than the line drawing interaction. The sequence of colours at the top of the screen communicates the correct sequence that must be achieved to complete a specific level. The difficulty of Linia is controlled only by the arrangement and behaviour of the geometry and, secondly, by the complexity of the colour sequence required. Linia’s guidance via a FTUE features a simple and non-intrusive approach to guiding its players in the early stages of the game, utilising graphical overlays on the user interface to provide hints to the correct actions. Linia’s FTUE highlights the interaction and specific motion required to path a line across the geometry, as well as a cue as the colour sequence required and where that information sits on the user interface. The FTUE in Linia leaves the player with full autonomy and agency to control and experiment with the game, using overlay information to guide and inform him/her in a relatively subtle manner.

3.2.2. Setup

The device remained consistent to ensure bias is mitigated where possible. To further ensure consistency and reliability, environmental factors were accounted for where appropriate and practical, i.e. noise levels, physical orientation/position, time of day, social setting and environmental setting. Mental performance is also known to be stronger in the morning than any other time of day, with fatigue and patience more commonly experienced during the later hours of the day [15]. The experiment was conducted within the hours of 9 am and 4 pm, in a quiet, private and comfortable indoor desk environment.

3.3. Procedure

The participants were introduced to the questionnaire with a brief overview of the protocol and events to come. Once the participants had confirmed they were unfamiliar with the games, they were placed in either the control or treatment group (based on a sequential placement). Random counter-balancing was used to determine the first game. Depending on whether the participant was administered guidance and information (Treatment) or not (Control), they would either receive 90 seconds (Treatment) or 60 seconds (Control) to play the game. This time differential exists due to the additional dialogs, cutscenes and other learning and guidance material found present in the Treatment group’s experience. The participants were instructed to try their best at completing whatever goal or objective they believed they should be attempting to achieve. The termination clauses were either the time limit (as outlined above) or the completion of the level/section. Once the session terminated, the participants were asked to complete the previously described adapted IBM PSSUQ, scoring the usability over 11 questions on a 7-point Likert scale [14]. Upon completing the questionnaire, the participants would then be asked to play the remaining game and complete the relevant second questionnaire. This procedure is demonstrated pictorially in Figure 2.

3.4. Participants

The experiment was conducted on 20 participants of mixed gender. The participants were volunteers and no incentive was offered to participate in the trial. This yielded two groups, control π_c and treatment π_t , in a between-participant design of 10 participants each. This group size was based upon the minimal heuristics posited by Winkler *et. al.* and Mantiuk *et. al.* [16, 17].

4. Results

A retrospective power analysis of this experimental design was conducted. For the means π_c, π_t and standard deviations σ_c, σ_t for each group control or

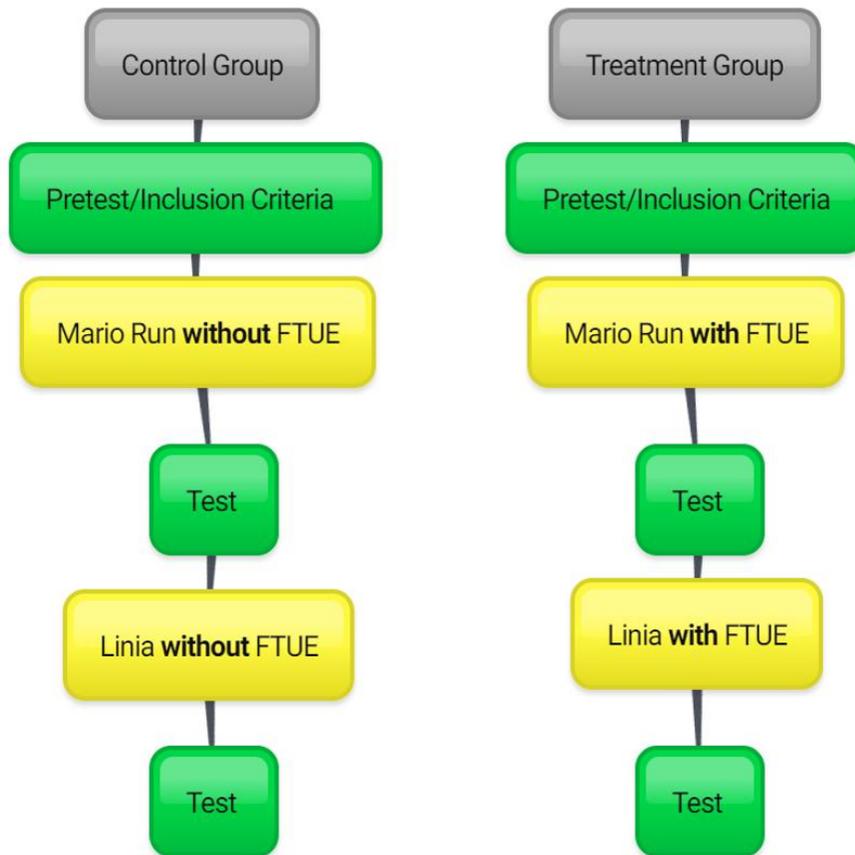


Figure 2: The procedure shown for each game condition. Yellow boxes can be in reversed order in the flow of control based upon the random choice to eliminate experimental ordering bias.

treatment respectively, this elucidated a Power $(1 - \beta)$ of 0.7676. This is based on a Type I error rate α of 0.05, where β is the Type II error rate. This also assumed τ , the number of pairwise comparisons to be made was 11, one for each of the questions being asked in the adapted IBM PSSUQ.

The results, shown in Figure 3 and in Figure 4 display correlations between guidance (existence of FTUE) and usability scores, collected and measured using an adapted (i.e. as outlined earlier in the article with the language contextualised to games) version of the IBM PSSUQ. Combining the groups amongst

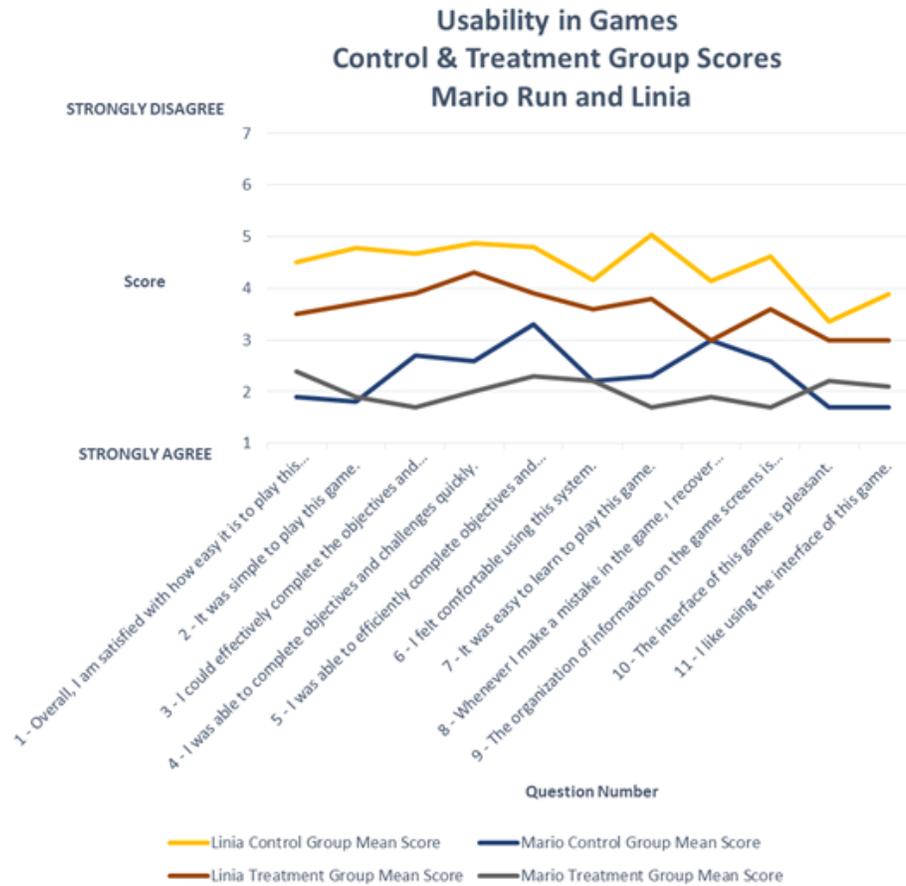


Figure 3: Scores for the Control and Treatment Groups of both Super Mario Run and Linia

Mario and Linia allows for the comparison of Control versus Treatment across both games, providing insights into cross-genre correlations regarding the presence of guidance. Using the non-parametric Mann-Whitney U test [12], the two groups differed significantly in regards to Information Quality (Questions 8 and 9 Av.), reporting $U = 125.5$, $Z = -2.035$ and $p = .043$, shown in Table 1, displaying a positive correlation between the games' usability, specifically the information quality and guidance. We consider that with guidance comes understanding, allowing the player to utilise all available information, from UI elements to in-game mechanics, thus improving usability. In contrast to this,

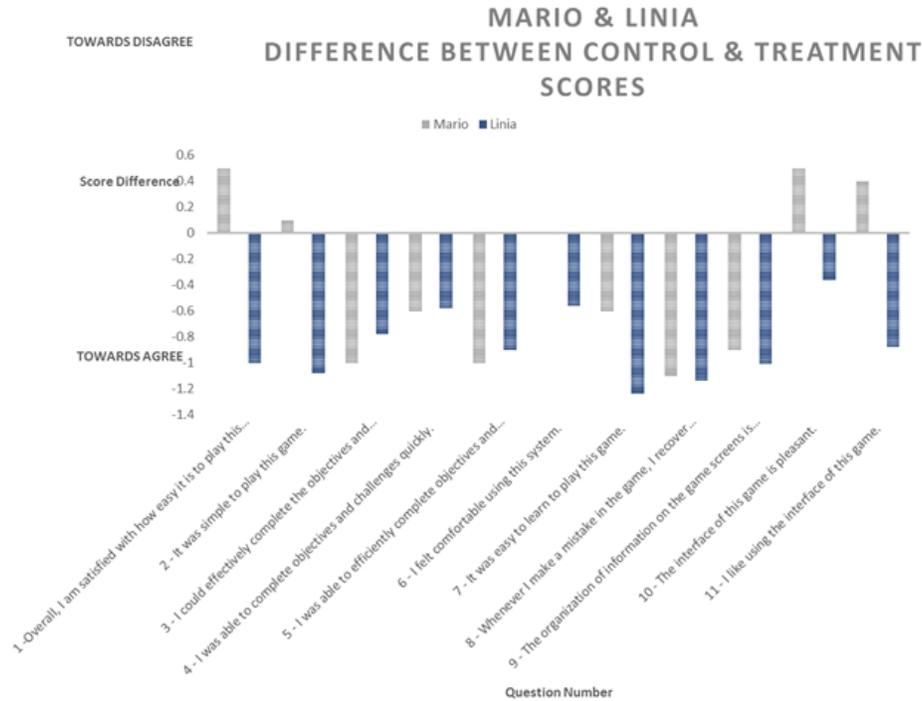


Figure 4: Difference between Control and Treatment scores

Overall Usability (Q1 to Q11) returns $U = 170.5$, $Z = -799$ and $p = .429$, which conveys that there is no significant result for the correlation of overall usability between the Control and Treatment groups. We believe that the design of the intuitive design and interaction model is crucial to usability, with guidance only aiding a game's usability.

As mentioned previously, Linia features a series of colours at the top of the screen that communicates the correct colour sequence the player must achieve to complete the level, contextualised and explained in the FTUE, allowing the player to use the information appropriately. Similarly, in Mario, the player is taught that the bubbles in the user interface represent additional lives, plus the value and use of this mechanic. Though a FTUE might not increase the overall usability of a game, it can be used to help inform the player, ensuring they understand the systems correctly. The absence of positive correlations in the

	U	W	Z	p (2-tailed)	p [2*(1-tailed)]
Question 1	184.50	394.50	-0.430	.667	.678
Question 2	185.50	395.50	-0.400	.689	.698
Question 3	153.50	363.50	-1.283	.200	.211
Question 4	171.50	381.50	-0.784	.433	.445
Question 5	146.50	356.50	-1.476	.140	.149
Question 6	174.00	384.50	-0.718	.473	.495
Question 7	152.50	362.50	-1.314	.189	.201
Question 8	132.50	342.50	-1.864	.062	.068
Question 9	159.50	369.50	-1.120	.263	.277
Question 10	198.50	408.50	-0.043	.966	.968
Question 11	186.50	396.50	-0.379	.704	.718
SYSUSE	166.00	376.00	-0.921	.357	.369
INFOQUAL	125.50	335.50	-2.035	.042	.043
INTERQUAL	192.00	402.00	-0.224	.823	.841
OVERALL	170.50	380.50	-0.424	.424	.429

Table 1: Statistical Significance Test Results from Super Mario Run and Linia Control vs. Treatment Groups. U is the Mann-Whitney U test, W the Wilcoxon signed-rank test and Z is the standard z-score

other usability variables, i.e. SYSUSE, INTERQUAL and OVERALL, stands to serve as an argument for the importance of ensuring fundamental usability of a game or application. As was identified in Section 2, there are usability guidelines and heuristics for improving the usability of games, with evidence and research to prove their worth, as shown in the work of Papaloukas *et. al.* [18].

The above highlights the possibility of a negative correlation between Overall satisfaction of ease of play and guidance via a FTUE. It is our belief that the increase in overall satisfaction can be attributed to the player’s self-discovery of the controls and interface (Control Group), where they are free to learn with full

agency/autonomy. This is the kind of autonomy and agency that is unavailable to the players presented with a FTUE, because of forced scenarios and intrusive dialogs (Treatment Group).

4.1. Limitations

The most significant limitation of the experiment relates to the sample size. It would also be beneficial to assess a different population/sample. Comparing different samples and demographic data might, for example, reveal correlations in age groups and usability ratings. Another limitation, highlighted through reviewing industry practices in usability testing, is the limitation of the single methodology strategy. The value of the discussed usability testing methods, such as gameplay sessions, lies in their ability to collect valuable, qualitative usability concerns and problems present in games.

Regarding the number of games assessed, the small number of games (namely two) across a small genre set is acknowledged to be a limitation of the research. This only allowed discussion and analysis of the findings of the two selected games and thus the two specified genres associated with these (platform and puzzle games). If additional games were employed in the assessment, wider and broader conclusions could have been permitted, strengthening the discussions and conclusions. The results from different genres could have varied, with certain genres more reliant on the acquisition of information and understanding than others, such as the casual/action genre versus the real-time strategy genre.

5. Conclusion and Future Work

The null hypothesis, \mathbf{H}'_0 , i.e. guidance and information via a First Time User Experience does not affect the usability of mobile games is rejected, because elements of usability are evidenced to be affected by the presence of guidance and information. This communicates that FTUEs can increase the elements of usability of a mobile game. However, other areas of usability were observed to not be significant. We can however accept \mathbf{H}_a because the control and treatment results produced different results. Guidance and information influenced

usability, corroborating \mathbf{H}_b . Finally, various elements of usability (in this instance INFOQUAL) were influenced by the guidance and information that was presented, which agrees with \mathbf{H}_c .

From the study in this article it is shown that FTUEs have the power to affect user perception in elements of usability. From a game design perspective, this is impactful. A macro view of this is useful, however, it is yet unclear on the micro scale which influences control this effect. Future work will consider trying to elicit several heuristics to guide game designers in the generation of FTUEs.

We recommend that developers and designers adhere to game usability heuristics and guidelines such as those presented by Papaloukas *et al.* [18] and Isbister and Schaffer [19]. The use and application of FTUEs in mobile games is also recommended, thanks to their evident value in aiding the specific INFOQUAL usability variable, helping to educate players regarding in-game feedback, user interface elements, and appropriate inputs and interactions. This article recommends that developers provide the player with guidance and information in a FTUE in a manner that grants the player agency and autonomy in the game world.

References

- [1] S. Griffiths, Google: Mobile App UX Principles.
URL <https://www.thinkwithgoogle.com/intl/en-gb/advertising-channels/mobile/mobile-app-ux-principles-improving-user-experience/>
- [2] L. Barnett, C. Gatzidis, C. Harvey, An investigation into usability and first time user experiences within a mobile gaming context, in: The 11th International Conference of E-Learning and Games, Edutainment, 2017. doi:10.1007/978-3-319-65849-0.
- [3] Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability, Standard, International Organisation for Standardisation, Geneva, CH (1998).

- [4] J. Sauro, E. Kindlund, A method to standardize usability metrics into a single score, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05, ACM, New York, NY, USA, 2005, pp. 401–409. doi:10.1145/1054972.1055028.
URL <http://doi.acm.org/10.1145/1054972.1055028>
- [5] D. Myers, A Q-Study of Game Player Aesthetics, Simulation and Gaming 21 (4) (1990) 375–396.
- [6] J. Schell, The Art of Game Design: A Book of Lenses, 2nd Edition, CRC Press, 2014.
- [7] P. Sweetser, P. Wyeth, Gameflow: A model for evaluating player enjoyment in games, Comput. Entertain. 3 (3) (2005) 3–3. doi:10.1145/1077246.1077253.
URL <http://doi.acm.org/10.1145/1077246.1077253>
- [8] L. Nacke, From playability to a hierarchical game usability model, in: Proceedings of the 2009 Conference on Future Play on @ GDC Canada, Future Play '09, ACM, New York, NY, USA, 2009, pp. 11–12. doi:10.1145/1639601.1639609.
URL <http://doi.acm.org/10.1145/1639601.1639609>
- [9] E. Adams, Fundamentals of Game Design, 3rd Edition, New Riders, 2013.
- [10] R. Hunicke, M. LeBlance, R. Zubek, MDA: A formal approach to game design and game research, in: Challenges in Games AI Workshop, 2004.
- [11] F. W. Petersen, L. E. Thomsen, P. Mirza-Babaei, A. Drachen, Evaluating the onboarding phase of free-to-play mobile games: A mixed-method approach, in: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '17, ACM, New York, NY, USA, 2017, pp. 377–388. doi:10.1145/3116595.3125499.
URL <http://doi.acm.org/10.1145/3116595.3125499>
- [12] A. Field, Discovering Statistics Using IBM SPSS Statistics, 4th Edition, Sage Publications, 2013.
- [13] M. Rajanen, J. Nissinen, A survey of game usability practices in northern european game companies, Information Systems Research Seminar 8.

- [14] R. Lewis, IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use, in: IBM Human Factors Group, 1995.
- [15] N. G. Pope, How the time of day affects productivity; evidence from school schedules, *The Review of Economics and Statistics* 98 (1) (2016) 1–11.
- [16] S. Winkler, On the properties of subjective ratings in video quality experiments, in: 2009 International Workshop on Quality of Multimedia Experience, 2009, pp. 139–144. doi:10.1109/QOMEX.2009.5246961.
- [17] R. K. Mantiuk, A. Tomaszewska, R. Mantiuk, Comparison of four subjective methods for image quality assessment, *Comput. Graph. Forum* 31 (8) (2012) 2478–2491. doi:10.1111/j.1467-8659.2012.03188.x.
URL <http://dx.doi.org/10.1111/j.1467-8659.2012.03188.x>
- [18] S. Papaloukas, C. Stoli, K. Patriarcheas, M. Xenos, A Survey on How Online Games Usability Affects the Retention of Information, in: 14th Panhellenic Conference on Informatics (PCI 2010), 2010, pp. 10–12.
- [19] K. Isbister, N. Schaffer, *Game Usability: Advancing the Player Experience*, 1st Edition, CRC Press, 2008.