An Investigation into Usability and First Time User Experiences within a Mobile Gaming Context: A Follow-Up and Critique

Ryan Pritchard

University of North Texas, Denton, TX, USA

ryanpritchard@my.unt.edu

# I. Abstract

In this paper, I will discuss the methodology and results of the paper ‘An Investigation into Usability and First Time User Experiences within a Mobile Gaming Context’ by Barnett, Gatzidis and Harvey [1]. I will define ‘usability’ as far as it is significant to the experiment. I will also show the results I achieved when I reproduced the experiment with a smaller sample size. Finally, I will discuss these results and criticize the experiment and results returned by the Barnett, Gatzidis, and Harvey paper.

# II. Introduction

In the Barnett, Gatzidis and Harvey paper as well as this one, the key factor tested is user perception of usability. Therefore, I must start by defining usability.

In short, usability can be summed up as user-friendliness. Usability is defined by ISO 9241-11, 1998 (Guidance on usability): “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” [2]. This definition divides usability into three sections: effectiveness, efficiency, and satisfaction. However, in the context of video games, this is not an even balance, as games are played solely for the purpose of satisfaction, or at least with satisfaction heavily favored over the other two.

Myer’s study of Game Player Aesthetics [3] defines “challenge” as “the most preferred characteristic”. Referencing Myer’s study among others, Barnett, Gatzidis and Harvey drew the conclusion that a lack of direction (instead ‘forcing’ players to discover controls and game mechanics on their own) led to an enhanced perception of usability in a game’s design. Indeed, the paper found a correlation between “Overall satisfaction and ease of play” and lack of guidance; this was attributed to “the player’s self-discovery of the controls and interface (Control Group), where they are free to learn with full agency/autonomy” [1].

Barnett, Gatzidis and Harvey drew this result from a modified IBM Post-Study System Usability Questionnaire (PSSUQ), “with the language contextualised to games” [1]. By using two different video games of different genres, each with control and treatment groups, they intended to find “insight into cross-genre correlations regarding the presence of guidance” [1].

# III. Methodology

In Barnett, Gatzidis and Harvey’s paper, they validated their results via experiments with human subjects. The experiment was laid out as follows:

The original test was performed with *Super Mario Run* and *Linia,* conducted on an iPhoneSE with 20 participants of mixed gender selected from various courses at Bournemouth University. In my retread of the same experiment, the phone was replaced by an Android Alcatel A20, and the participants are current and former students of mixed gender from University of North Texas. There are 6 participants in this experiment rather than the original experiment’s 20.

The games *Super Mario Run* and *Linia* were chosen because of their “similar yet contrasting interaction complexity, since they can both be controlled with one finger” [1], while the interactions are used for entirely different contexts in each game. *Super Mario Run* is a platforming game, while *Linia* is a puzzle game.

A control group and treatment group were both selected (randomly, without the participants’ knowledge) and informed about the events to come. If participants had played either game before, they were ineligible for testing and dismissed; otherwise, the experiment continued. Each participant played one of the games, then took the modified IBM PSSUQ, then played the second game followed by another modified IBM PSSUQ.

The order that each of the two games was played in was chosen randomly for each subject.

## Experiment (Treatment group)

**Super Mario Run session:** Prior to the beginning of the experiment, I skipped all of the introduction tutorials to *Super Mario Run* until the game gives the ‘Tour’ option at the bottom of the screen. I then selected ‘Tour’ and gave the phone to the test subject with directions not to start playing yet. I then read directions for the game off of a script and allowed the subject to begin. The subject was then given 60 seconds to play the game. The subject was allowed to ask questions while he/she played. After 60 seconds, the test subject was told to put the game down and return the phone. The subject was then given the modified IBM PSSUQ.

**Linia session:** I opened *Linia* and gave the phone to the test subject with directions not to start playing yet. I explained the rules of the game by reading off of a script, and then allowed the subject to begin. The subject was given 60 seconds to play the game. The subject was allowed to ask questions while he/she played. After 60 seconds, the test subject was told to put the game down and return the phone. The subject was then given the modified IBM PSSUQ.

## Experiment (Control group)

**Super Mario Run session:** Prior to the beginning of the experiment, I skipped all of the introduction tutorials to *Super Mario Run* until the game gives the ‘Tour’ option at the bottom of the screen. I then selected ‘Tour’ and gave the phone to the test subject. I left the subject alone for 60 seconds. I returned after 60 seconds, and the subject was told to put the game down and return the phone. The subject was then given a modified IBM PSSUQ.

**Linia session:** I opened *Linia* and gave the phone to the test subject. I left the subject alone for 60 seconds. I returned after 60 seconds, and the subject was told to put the game down and return the phone. The subject was then given a modified IBM PSSUQ.

## Modified IBM PSSUQ

After each session, the test subjects were asked to complete a modified IBM PSSUQ. The questionnaire asked subjects to score each game’s usability on 11 questions over a 7-point Likert scale [4]. The modified IBM PSSUQ contained the following questions (the exact same questions used by Barnett, Gatzidis and Harvey [1]):

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 organization of information on the game screen is clear.
10. The interface of this game is pleasant.
11. I like using the interface of this game.

# IV. Results

Figures 1 and 2 show the data gathered for both the control and treatment groups. Figure 1 displays this by participant, while Figure 2 organizes the data by game played.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Control** (Mario/Linia) | | | **#** | **Treatment** (Mario/Linia) | | |
| 7/6 | 6/5 | 4/6 | 1 | 3/7 | 5/5 | 6/3 |
| 6/7 | 7/5 | 7/6 | 2 | 5/5 | 5/6 | 4/5 |
| 6/7 | 6/6 | 3/7 | 3 | 5/5 | 4/5 | 5/5 |
| 6/6 | 6/5 | 3/5 | 4 | 6/5 | 5/5 | 5/4 |
| 6/6 | 6/5 | 3/5 | 5 | 6/5 | 5/5 | 5/4 |
| 7/7 | 7/5 | 6/7 | 6 | 4/6 | 4/4 | 7/6 |
| 7/7 | 7/4 | 7/7 | 7 | 7/7 | 5/5 | 6/5 |
| 6/7 | 6/5 | 6/3 | 8 | 2/5 | 6/5 | 7/6 |
| 6/7 | 5/3 | 2/5 | 9 | 6/7 | 5/7 | 5/6 |
| 7/7 | 5/4 | 6/5 | 10 | 5/7 | 5/6 | 6/6 |
| 7/7 | 5/3 | 6/5 | 11 | 4/7 | 4/6 | 7/6 |

## Figure 1: Results by participant

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Control** | | **#** | **Treatment** | |
| *Mario* | *Linia* | - | *Mario* | *Linia* |
| 7/6/4 | 6/5/6 | 1 | 3/5/6 | 7/5/3 |
| 6/7/7 | 7/5/6 | 2 | 5/5/4 | 5/6/5 |
| 6/6/3 | 7/6/7 | 3 | 5/4/5 | 5/5/5 |
| 6/6/3 | 6/5/5 | 4 | 6/5/5 | 5/5/4 |
| 6/6/3 | 6/5/5 | 5 | 6/5/5 | 5/5/4 |
| 7/7/6 | 7/5/7 | 6 | 4/4/7 | 6/4/6 |
| 7/7/7 | 7/4/7 | 7 | 7/5/6 | 7/5/5 |
| 6/6/6 | 7/5/3 | 8 | 2/6/7 | 5/5/6 |
| 6/5/2 | 7/3/5 | 9 | 6/5/5 | 7/7/6 |
| 7/5/6 | 7/4/5 | 10 | 5/5/6 | 7/6/6 |
| 7/5/6 | 7/3/5 | 11 | 4/4/7 | 7/6/6 |

## Figure 2: Results by game

With this data, I used the non-parametric Mann-Whitney U test to determine correlations across the games and groups. The results are shown in Figures 3 and 4.

Figure 3 takes the sums based on each game, while Figure 4 takes total sums from the control and treatment groups.

Winner (larger sum) gets 1 point, tie is 0.5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Control** | | **#** | **Treatment** | |
| *Mario* | *Linia* | - | *Mario* | *Linia* |
| 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 2 | 0 | 0 |
| 1 | 1 | 3 | 0 | 0 |
| 0 | 1 | 4 | 1 | 0 |
| 0 | 1 | 5 | 1 | 0 |
| 1 | 1 | 6 | 0 | 0 |
| 1 | 1 | 7 | 0 | 0 |
| 1 | 0 | 8 | 0 | 1 |
| 0 | 0 | 9 | 1 | 1 |
| 1 | 0 | 10 | 0 | 1 |
| 0.5 | 0 | 11 | 0.5 | 1 |

## Figure 3: Mann-Whitney U test (sum by game)

MC= Mario control, LC=Linia control, MT=Mario treatment, LT=Linia treatment

UMC=7.5; ULC=7; UMT=3.5; ULT=4

|  |  |  |
| --- | --- | --- |
| **Control** | **#** | **Treatment** |
| 1 | 1 | 0 |
| 1 | 2 | 0 |
| 1 | 3 | 0 |
| 1 | 4 | 0 |
| 1 | 5 | 0 |
| 1 | 6 | 0 |
| 1 | 7 | 0 |
| 1 | 8 | 0 |
| 0 | 9 | 1 |
| 0 | 10 | 1 |
| 0 | 11 | 1 |

## Figure 4: Mann-Whitney U test (combined sums)

C=Control, T=Treatment

UC=8; UT=3

# V. Discussion

The results shown in Figures 3 and 4 display a correlation between usability and lack of direction; that is, users in the control group, on average, tended to score each game higher on the modified IBM PSSUQ. Though combining the sums of both games (Figure 4) gives the control group an indisputable ‘win’ over the treatment group, separating the games (Figure 3) is slightly closer, but still convincingly control group-favored. As suggested in Barnett, Gatzidis and Harvey’s paper, one possible explanation for this is “the player’s self-discovery of the controls […] where they are free to learn with full agency/autonomy” [1]. Indeed, I found a similar negative correlation between “Overall satisfaction and ease of play” and guidance (treatment group).

However, I believe that it may not be this simple. Barnett, Gatzidis and Harvey did not disclose their exact results (just their conclusions mathematically drawn from them), but judging from my own results, there is a strong possibility that outliers may have influenced the results. Some questions received extremely polarizing feedback even across the same groups and games (for example, the scores of 2/6/7 for question #8 in the Mario treatment group, and scores of 7/3/5 for questions #9 and #11 in the Linia control group). While normally such outliers would wash out in larger sample sizes, it could have had a significant impact when only 6 participants were involved.

This leads me to my main discussion point about Barnett, Gatzidis and Harvey’s paper. In their test, they interviewed only 20 participants. While larger than my sample size of 6, it is not significantly larger. While 20 participants would be sufficient for certain populations, Barnett, Gatzidis and Harvey did not specify a direct target audience, and as such limited themselves with a miniscule sample size for no apparent reason.

However, sample size aside, I have more or less verified the results of the original paper. The IBM PSSUQ and Mann-Whitney U test are potent and effective for the task at hand, and now as in the Barnett, Gatzidis and Harvey paper, they show a notable correlation between guidance and user perception of usability on a first-time experience.

# VI. References

1. Barnett, L., Gatzidis, C. and Harvey, C., 2017. An Investigation into Usability and First Time User Experiences within Mobile Games. In: Edutainment, 26-28 June 2017, Bournemouth. England.
2. ISO 9241-11, 1998. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. ISO.
3. Myers, D, 1990. A Q-Study of Game Player Aesthetics. SIMULATION & GAMING, 21(4), 375-396.
4. Lewis, R.J., 1995. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Boca Raton, FL: IBM Human Factors Group. 8