Understanding Player Experience Using Sequential Analysis

Michael Soppitt & Graham McAllister
University of Sussex
Brighton
Sussex
BN1 3QJ
michaelsoppitt@gmail.com g.mcallister@sussex.ac.uk

ABSTRACT
This paper presents a methodology which allows for an objective analysis of video game quality based on player behaviour. A mutually exclusive and exhaustive code of 5 behavioural states is presented based on an analysis of 10 users each playing 3 video games. The coding scheme is verified for inter-coder agreement with resulting Kappa values in the range of 0.74 to 0.91 (good agreement to very good agreement). Results of the game studies presented show that good games allow the player to enter the Engagement state more frequently, and keep them in that state for a longer duration than bad games. In particular, the results show that good games exhibit an overall net positive behaviour from the very early stages of gameplay. The paper concludes with suggestions for future work.

Keywords
Video games, sequential analysis, behavioural coding, immersion, game quality.

INTRODUCTION
Understanding the nature of interactive behaviour is fundamental in illuminating the components of gamer experience. These concepts often centre in notions of flow (Chen et al 1995), engagement, or engrossment (Brown et al 2004). Current methods of assessing the gamer experience include approaches such as interviewing the player after the session, or asking the player to rate the gaming experience at discrete points throughout the gameplay session (Tulathimutte et al 2008). Such measures are often flawed in both applicability (Ermi et al 2005) and suitability (Czaja et al 2005), failing to accurately elucidate the dynamic interactive process of gamer experience (Bakemen et al 1997).

Davis et al (2005) states that usability testing can be an excellent source of behavioural information and can be used to identify barriers which block the user from experiencing the fun of a game. Methods such as questionnaires, interviews and focus groups are the most commonly used approaches to capturing the player's responses. Focus groups can be considered the primary method for extracting rich qualitative information by encouraging participants to discuss and explain their opinions (Morgan 1996). Contrary to usability testing, focus groups help provide perceptual information and thus are commonplace in the concept evaluation stages of game design (Davis et al 2005). However the impact of group social dynamics can have a dramatic impact. Sussman et al

Proceedings of DiGRA 2011 Conference: Think Design Play.

(1991) observed a polarising effect whereby attitudes became more extreme after group discussion.

Outside this realm of social influence are surveys and interviews. Acting at a lower level of granularity, surveys can tap into a player's perceptions of the game itself, informing on the good and bad points of gameplay (Davis et al 2005). The function of a survey is twofold; it must solicit meaningful information as well as provide comparable data about a specific source (Fowler 1995). Thus the questions asked need to be a valid measure of what is to be described. However in achieving this goal we must draw reference to a plethora of other influencing factors including, biases, leading or hypothetical questions, emotional phrases and design layout. Moreover, game experience through survey can only be assessed retrospectively and thus participants are required to rate an experience that may have varied greatly (Davis et al 2005).

The high level objective of any questionnaire is to extract the most accurate data in support of the research goal (Brace 2004). However, such static measures of capturing gameplay experience may only identify conceptual perceptions or how specific features relate to the gaming experience. None of these methods categorically explicates how a player interacts with the game or how it will perform overall. Understanding gaming experience requires an appreciation of the dynamic aspects of interaction as it changes and evolves through time. Bakeman et al (1997) argues that it is "this sequential view that offers the best chance of illuminating dynamic processes of social interaction".

The assessment of complex behavioural interactions through the use of both socially and physically based behavioural codes is well documented. Examples of this vary greatly, with perhaps the most famous being the interactions of marital couples (Gottman 1979). Gottman (1979) has written extensively on marital interaction and developed what is known as the Specific Affect Coding System (SPAFF). Couples were recorded discussing a point of contention in their marriage and the outcome analysed using SPAFF. Gottman was able to predict with a 95% accuracy if a couple would still be together in 15 years. The code acts as the instrument of measurement for observational research much in the same way the survey is the tool of relational analysis. Similarly, as a survey uses questions to extract its data, behavioural analysis uses a code to determine which information is relevant from the passing stream of data. For Gottman, this depicted the distribution of positive and negative emotion throughout the marital discussion. Through this coding he was able to determine the marital footprint that determines the marriages' sustainability. Gladwell (2005) in his acclaimed book Blink compares this to what is known as a 'fist'. It occurs naturally and unconsciously and yet it is inherent in the smallest part of data. Indeed, later research by Carrere et al (1999) illustrated that marital stability could in fact be predicted over a 6 year period from just the first 3 minutes of a marital conflict discussion.

Studies utilising systematic and sequential analysis to quantify naturally occurring behavioural patterns have been well documented. However the focus of these studies has almost exclusively been that of the interactions between humans. Yet as described by Bakeman et al (1997) a "defining characteristic of interaction is that it unfolds in time". Thus it follows that to unpack the variables of gamer experience we must also work to understand this dynamic behavioural axiom. Current academic research has yet to fully examine how this powerful stream of interaction exhibited during gameplay can be used to unravel the components of user experience and inform design. This study attempts to bridge that gap.

BEHAVIOURAL CODES

Systematic interaction has long been considered the most appropriate way to understand behaviour. Yet it is the sequential view of this interaction that perhaps best reveals the dynamic components of social process (Bakeman et al 1997). This distinction whilst simple is essential in understanding why sequential analysis is the appropriate tool in investigating both social process and in this instance that of human computer interaction (HCI). This distinction can be elegantly exemplified in the comparison of the early childhood studies of Parten (1932) and later Bakeman et al (1980). The context of these papers are of course a long way from that of HCI. Yet they typify how sequential analysis could be used to probe the dynamic interactions of gamer experience.

Mildred Parten (1932) conducted a study that aimed to determine how basic demographic information such as intelligence, sex, father's occupation and size of family had on the development of social participation in 42 young children. Parten developed a behavioural coding scheme describing differing states of child interaction during play. These were described as Unoccupied, Onlooker, Solitary, Parallel Activity, Associative Activity and Co-operative Activity. Each child was monitored for an average of 1 minute a day with the order of child sampling varying systematically. Each time sample was categorized based on the code that best described that period of play. What is important here is both the method used and Paten's research questions. Bakeman et al (1997) describes how this methodology allows for the proportion of time spent in each activity to be determined by the summation of the non-sequential time samples. He continues that such a process is adequate for the questions at the heart of Paten's study. The comparison of social participation to the diagnostic metrics summarized above does not require an understanding of how these behaviours were sequenced in time. However such a study cannot determine how these behavioural states are sequenced in relation to each other. Indeed, whilst Parten concludes that Parallel activity was an obligatory play state for children as they develop, she was unable to elucidate the momentary interactions in which such activity is sequenced (Bakeman et al 1980). Thus to probe social interactions it is a sequential approach that must be adopted (Bakeman et al 1997).

Smith (1978) showed that contrary to the research of Parten, parallel play was in fact discretionary and not the obligatory stage through which all children pass. These conflicting views led Bakeman and Brownlee (1980) to consider if "parallel play may arise less from the forces of development than from the needs of the moment". To answer this question an understanding is required of how parallel activity is sequenced in time. A total of 41 were filmed during a period of free play. A code modified from Parten's initial study consisting of Unoccupied, Solitary, Together, Parallel and Group activity was used. Each 15 second interval was then coded dependant on which activity best represented that segment (Bakeman et al 1980). The research allowed Bakeman and Brownlee to determine if the migration between certain states was either probable or improbable. The results illustrated that parallel play acted as an intermediary to group play with a probability that exceeded chance for 32 out 41 of the children (Bakeman et al 1997). Thus contrary to the findings of Parten parallel play was shown to be less a developmental stage but an important transitional state. Without a sequential approach in which time is preserved such a distinction would never have been made. In essence sequential analysis acts as a tool to "determine if and how an action or event is related to another action or event" (Morgan et al 1992 cited McComas 2009). In studies of gamer experience, industry and research alike have long settled for static measures of interaction. Yet the sequencing of behaviour defined through user experience seems to be at the centre of understanding how game quality unfolds itself in time.

CODE CONSTRUCTION

This research focuses on how temporal measures of analysis inferred behavioural states correlate to game quality. This clear proposition helps define what characteristics from the behavioural stream should be extracted. Rosenblum (1978) noted that one of the most effective ways of facilitating this code production was to take a pen and start recording observations. By doing so he claimed that it was possible to put aside any bias or "rigidifying constraints" such to allow for behavioural patterns to "arise as separate entities". The construction of the behavioural code for use in this study was initiated in this manner. In effect observations were made 'cold' with little exposure to any previous literature or priming material.

Pilot study

In order to construct the coding scheme for the main study, an initial pilot study was conducted. Participants were recruited one the basis that they were iPhone owners who regularly played games. Two males aged 37 and 42 were accepted. A total of 3 games were selected for these initial trials all of which would be played by both participants. The games were selected based on the rating provided by pocketgamer.com. These where *Zen Bound Lite* (Chillinqo Ltd, 2010) *Bubble Boom Lite* (Zed Worldwide, 2009) and *Dark Hill Lite* (Nabil Chatbi, 2009) with ratings of 10/10, 5/10 and 3/10 respectively. No specific consideration was given to game genre although the respective categorization by pocketgamer.com was Puzzle, Casual and Adventure Shooter. Each participant was observed in a quiet room and sat directly opposite from the facilitator such that only the participant could view the screen. As the focus of this study was to explore player behaviours relative to the gaming experience, sight of the game screen was not considered a requisite. Users were instructed on game selection and asked to play the game as they normally would. Each participant was filmed for approximately 6 minutes per game.

All 3 games for each player were analysed and a descriptive narrative of the gaming experience constructed. This included every aspect of the gaming experience from slight head movements to verbalizations. From the collation of narratives it was possible to extract common or stereotyped features. Bakeman et al (1997) noted that it is important during code development to accurately articulate why a certain behaviour is classified the way it is. To do this, each sample of video footage was reviewed repeatedly and the reasons for classifying a certain behaviour noted. It is important here to draw reference to the socially based nature of this code. Moreover, it is not entirely defined in terms of physical and intrinsically quantifiable attributes. Instead, socially-based codes are concerned with behaviours that "depend far more on ideas of the mind" (Bakeman et al 1997). Thus accurate articulation of a behavioural process is integral to the replicability and reputability of the code itself. These broad descriptions of play states formed the rudimentary behavioural code of this study.

Refinement through literature

The behaviours extracted from the stream of data elicited between human and game was done so with the subjective bias of the code creator. These behaviours were deemed only in opinion as relevant in helping to elucidate the research question of focus. Thus the behavioural code used in this study is in essence a hypothesis. The quality of its conception is consequently crucial if it is to accurately depict the phenomena it sets out to explain. Bakeman et al describes the code as the lens in which one has chosen to view the stream of behavioural data. If this is poorly conceived then those behavioural sequences that one hopes to describe will simply not emerge. Thus the code at the centre of this study was refined and aligned with the findings of academic gaming research.

Brown et al (2004) conducted a series of interviews to determine gamers' experience of game immersion. Three successive levels of involvement were defined as engagement, engrossment and immersion. Access to each level required that certain barriers be lowered. Due to the nature of iPhone games and the relatively short periods of game play, the opportunity for players to reach an immersive state was seen as unlikely. However, those states of engagement and engrossment drew parallels to those behaviours identified in the initial pilot. The former explicates a state where the player is engaged in the game itself, access to this stage is governed by barriers such as gamer controls and thus requires the investment of time by the gamer. This state therefore requires periods of concentration, a factor considered integral to the foundations of achieving optimal experience (Pace 2004). Such behaviours where categorized during the initial pilot study as a state of 'Continuous Concentration' and where often observed in the initial stages of game play. This categorisation summarized periods where the player appeared entirely focused on the game. Additionally it was noted that such periods often resulted in a reduction of blinking. Research by Jennett et al (2008) supports this, as gamers progress to a state of immersion, their eye movement will reduce as they become fixated on crucial components of the game.

All these features where identified during the initial pilots and are consistent with theories of engagement. Those who appeared to enjoy the gaming experience also often vocalized their pleasure or commented on their actions in relation to the game. This behaviour draws strong parallels to concepts of engrossment. Brown et al (2004) explains how this is governed by game composition requiring that game features "combine in such a way that the gamers' emotions are directly effected by the game". Certainly emotional engagement has been identified as central to concepts of positive experience including concepts of flow (Csikszentmihalyi 1990) and as a dimension of immersion (Qin et al 2009).

Malone et al (1982) describes how challenge is critical in the achievement of fun in games. Specifically this characteristic must be present at a level of equilibrium to induce optimum experience. Snyder et al (2004) concurs that games require a "balance between perceived action capacities and perceived action opportunities". Behavioural traits of both frustration and boredom were identified during the initial pilot. Examples of this included furrowing of the brows or negative verbalizations for the former. Boredom was often characterized by a lack of interest in the gaming experience, with the player attempting to engage in conversation or by attending to items other than the game itself. This also draws parallels to the work of Jennett et al (2008) where eye movements where recorded to increase as players become distracted by items irrelevant to the game itself. Importantly both these states can be seen as barriers to reaching a higher state of engagement. Overcoming such barriers mandate an investment of time. Initial studies clearly illustrated periods of play when gamers struggled with game interaction. This was characterized by dramatic or repeated hand movement. Such behaviour perhaps indicative of usability concerns impacting player's ability to "learn, control, and understand a game" (Pinelle et al 2008).

Curiosity also played a key role in gamer experience. Examples of this included what have been deemed 'eureka verbalisations' such as "I see!" indicating that the player's discovery of something new. Indeed curiosity has appeared in many studies including as a fun heuristic by Malone (1982) and as a key dimension of immersion by Qin et al (2009). Pace (2004) explains how curiosity is crucial not just in ensuring user attention but in the formation of the player's goals determining resultant user experience.

Game / player validation

The footage used for the initial code focused on the behavioural outcome of the gaming experience. The lack of comparable on-screen gaming activity was not an oversight but implemented to focus observation on the gamer. To fully understand if the behaviours elicited correspond to the given taxonomy it is important to understand how they relate to the game itself. With this in mind a game usability company were engaged to help provide dual screen footage of both gameplay and gamer experience. Footage of 5 players playing a single level of an iPhone game was provided. Of the 5 players, 2 where female and 3 male, ages ranging between 21 and 34. Approximately 15-20mins of gameplay was recorded. Players were encouraged to talk about their experience whilst playing the game as well being questioned during the process.

The verbalization of experience helped clarify the cognition behind behaviour. All players were observed to drop the phone away from the body after repeatedly crashing, this was often accompanied by a sigh. Such behaviour was identified during the pilot study. On-screen comparison helps place this behaviour as frustration. Players were often seen to perform rapid and dramatic movements with the iPhone during the initial stages of gameplay. Typically this involved quickly making a motion in one direction then immediately in the other, illustrative of a lack of responsiveness of the game controls. Such behaviours were considered indicative of frustration. Similar examples included moving the game toward the face or repositioning the screen. The features identified during this study fed directly in to the final code.

Inter-observer agreement

Socially based coding schemes, such as that of this study, not only use, but encourage the application of humans' inherent inferential ability. That is, while the code as described above explicates key characteristics of each state it does not attempt to provide an exhaustive description. Bakeman et al (1997) believes observers implementing a social based code should be considered "more as a cultural informant than as a detector". What is important here is that unlike physically based codes, socially based codes rely on the observer to accurately decide which state is being illustrated.

It is therefore important to illustrate accuracy through replicability. The use of a single observer for the main study further enhances this need. Replicability was thus achieved through the analysis of inter-observer agreement. One female aged 26 was recruited. She was naïve to the project and did not play video or mobile games of any description. Considered a secondary observer, she was at no point informed of the intention of the study. She was trained by the primary observer to use the code during a 1 hour tutorial session. During this session she was explained the behavioural states and shown instances of each from the footage acquired through the pilot study. She was encouraged to ask questions and challenge state boundaries.

The first participant of the primary study was used as the subject for comparison. The footage of all three games was first analysed by the primary observer. This coding was then mapped onto a spread sheet depicting the behavioural state against each 1 second unit of time. Once the secondary observer had expressed her understanding of the code and how to apply it, she was asked to select one of the three at random and code the footage. The observer was given as long as required to analyse this and did so alone. This data was then mapped against those observations of the primary observer and the points of disparity highlighted. A point-by-point mapping of code between observers was considered to be too stringent a method of correlation. Agreement was recorded when the codes correlated by ±1 second. The correlation between observers was then translated onto a confusion matrix. The agreement statistic Cohen's Kappa was used to assess significance in preference to simpler agreement percentages as it takes into account the impact of chance data distribution (Bakemen et al 1997). Due to the nature of the code not all behavioural states are were illustrated in the single 3 minute piece of footage. As a result, this initial inter-observer analysis only accounted for the presence of the states Frustration, Curiosity and Engagement. To illustrate the replicability of detection for the behavioural states Engrossment and Boredom, two more samples of footage were analysed for inter-observer agreement at the end of the main study. Each piece of footage was selected based on its presence of the required behavioural state and analysed using the same method as described above. The same secondary observer was used for each additional video.

For the initial inter-observer analysis including the behavioural states Frustration, Curiosity and Engagement, Cohen's Kappa was .78. Cohen's Kappa for the footage including instances of Boredom was .91. For footage with examples of Engrossment Cohen's Kappa was .74. As discussed by Bakeman et al (1997), Kappas of .60 to .75 illustrate good inter-observer agreement and those over .75 as excellent. This illustrates that the code is both accurate and replicable.

BEHAVIOURAL CODE

Positive Behavioural States

Engrossment
Player illustrates a level of emotional investment. Typically characterized by empathy where players seem share the experience of the game. This may be characterized by a physical empathetic reaction to gameplay. Positive movements such as nodding may be observed. Players may also positively vocalize their enjoyment or comment on their actions in relation to the game.

Engagement
Player appears focused and engaged in the game task for a sustained period of time. Characterised by distinct and continued periods of play where the face remains expressionless. Eyes appear focused on screen, often with minimal blinking / movement. The mouth will appear relaxed.

Neutral Behavioural States

Curiosity
Player appears to explore the game and their interactions appearing inquisitive rather than engaged. Actions typically appear laboured. This may involve tapping of screen often followed by slight pause. May be accompanied by eureka verbalizations ("Ah!" / "I see"). Player remains relaxed whilst trying a variety of finger movements and/or hand positions. May involve moving the phone towards the face or angling of the head. Player maintains attention to game often with little change in expression as if waiting for something to happen.

Negative Behavioural States

Frustration

Player will appear frustrated or agitated by the game as if unable to understand there interaction. May often result in rapid or more aggressive taping of the screen. Often negative movements are observed such as shaking of the head or expressed through a deep sigh. The mouth may be seen to tighten or become raised at one side. May be accompanied by a sigh or furrowing of the eyebrows. Player may give a rye smile without any distinct by noticeable change in eyes.

Boredom

Player seemed to become relaxed and un-engaged in gameplay. This may be characterised by the adoption of a more relaxed seating position. There may be prolonged periods without game interaction. Player may become easily distracted attending to items away from the gaming experience. May appear restless but not agitated. Eye movements may also been seen to increase as the player becomes more distracted by non significant items of the games visual display.

METHODOLOGY

With the coding scheme developed, consideration was given to the type of platform and games that should be the studies focus. Mobile games were considered most suitable as they are typically played over short periods of time. Such games aim to ensure that the user can both quickly learn the controls and become engaged. In essence, the lifecycle of each game session is short. Therefore, a range of behavioural states are exhibited over a much shorter time period thus aiding analysis. The iPhone, as a market leader in mobile gaming was deemed the platform of choice.

Game choice was based on two key criteria; genre and rating. It was considered important to ensure that all games assessed where of the same genre. By doing so it was possible to reduce the effect of player preference. Game choice was subsequently determined based on rating. The site pocketgamer.co.uk was used as the source of reference for this distinction, providing independent ratings on a range of mobile games from different mobile platforms. The games *Cops & Robbers Lite* (Glu Gams Inc, 2009), *iDroidsMania Lite* (Artificial Life, 2009) and *Inspector Gadget Lite* (Namco Networks America Inc, 2009) were selected based on their respective ratings of 9/10, 7/10 and 5/10. These ratings were hoped representative of games of high, average and low quality. All games had been reviewed in 2009, reducing any deviation in rating that may have resulted due to disparity in technological capability at the time of development.

Procedure

A total of 10 participants took part in the main study, 8 Male and 2 Female. All participants were English speakers aged between 22 and 46 (average age of 29) and were iPhone owners. Participants were informed that they should be willing to be filmed whilst playing iPhone games. No indication was given as to the focus of this research. Due to the perceived familiarity of platform game, no specific request was made for participants with preference to this gaming type.

All filming took place in a quiet room with only the participant and facilitator present. Participants were neither encouraged nor discouraged from talking during gameplay. Prior to the initiation of filming the facilitator wrote the participant reference at the top of the consent form. This was in the format [FirstInitial] [SecondInitial] _ [Participant#]

with exception of the first subject where [Participant#] was replaced with OBS to denote the use of this data for inter-observer correlation.

The facilitator was also required to later assess the footage and thus great care was taken where possible to ensure that the identity of the game being played was known only by the participant. To ensure this, three slips of paper were produced in the following format:

Game – [GameName]   Selection # [   ]   Rating (1-10) [   ]

All three pieces of paper were laid at random face down on the table such that what was written on them was hidden from the facilitator. These were then shuffled and the participant asked to take one at random and remember the game specified but to not vocalise what it was. The participant was then required to write the number correlating to their selection in the space marked 'Selection #' and place the paper slip in an envelope provided. The participant was instructed to select the game specified in [GameName] from the iPhone. The facilitator remained directly opposite the participant at all times and was thus unable to see what was on screen. All sound on the iPhone was turned off. It is important to acknowledge that this may be deemed to detract from the overall gaming experience. However by doing it was possible to ameliorate the possibility of auditory association and thus ensure the facilitator was kept unaware of the game being played.

To limit any variation of participant aptitude to learn game controls, all subjects were required to read the instructions for each game and inform the facilitator when this was complete. At this stage the subject was then asked to start playing the game, and the recording commenced for a duration of 3 minutes. The footage was then immediately transferred to a laptop and referenced in a manner consistent with the participant ID. This was in the format

[FirstInitial] [SecondInitial]_[Participant#/OBS]_[Selection#]

This process was then repeated for each of the remaining two games. In each instance the participant was required to write the numbers 2 and 3 respectively in the space marked 'Selection #' for each game.

After the participant had been filmed for 3 minutes playing each of the games they were then asked to remove all 3 pieces of paper from the envelope ensuring they were not seen by the facilitator. They were then asked to rate each game in the space denoted 'Rating (1-10)' where 1 was 'Very Poor' and 10 was 'Very Good'. These subjective participant ratings provide a diagnostic of game quality that is inherently bound to the emergent behavioural interactions of gameplay. Once all games were rated, the participant was instructed to place the slips back in to the envelope and seal it. At this point the envelope was handed to the facilitator and marked with the participant reference. Each envelope was then stored in a folder. To achieve maximum rater objectivity these envelopes where not opened until all the footage had been coded.

ANALYSIS

Each game recorded for each player was coded onto a separate spreadsheet within a single Excel document to help facilitate later data manipulation. The top of each page was initially marked with the participant ID and their selection. The code names were ordered on the horizontal, starting with the negative behaviours and progressing to the positive. The onset and offset of a code were noted below with the onset of each subsequent code matching the offset of the previous. Each successive code was recorded on a new line to emphasize the sequential nature of the data.

The sequence of behavioural transitions was noted and the corresponding time spent in each state calculated. State frequency was calculated as the sum of the occurrence of each respective behaviour and the total duration assessed as the sum of the time spent in each state. Two sequence behavioural transitions were documented to create a transitional matrix for each game. Similarly the variance in positive and negative behaviours over time were extracted from the onset and offset data for each 10 second period. This was then aggregated to give respective scores for each 30 second block. Totals were calculated for all metrics allowing the accurate reconciliation of later data aggregation. For each subject, after analysis had been completed for all 3 games the participant envelope was opened and the sections of game name, pocketgamer rating and player rating populated.

Upon completion of the analysis of the entire sample group, the data was aggregated based on the dependent variables identified for exploration. Data was accumulated using Excel formulas and the results verified using the aggregate totals of individual data sets. The data was then transposed to the analytical tool SPSS for repeated measures analysis of variance and subsequent paired t-tests.

Analytical considerations
The data was analysed to assess the validity of three general *a posteriori* hypotheses in relation to gamer behaviour. As with any parametric exploratory research, the risk of accepting H0 as false through unfocused statistical analysis is high. Consequently each hypothesis was considered independently exclusive relative to the next and utilised a discrete set of dependant variables. A repeated measures analysis of variance (ANOVA) was implemented in all instances to assess for any significance of effect. The use of this metric was cautious as it helped reduce the risk of courting type I error through the over application of statistical indices (Bakemen et al 1997). By implementing ANOVA it was possible to concentrate the further application of t to those interactions deemed only to be most relevant in answering the key questions of each hypothesis. A more conservative two-tailed alpha comparison was used in favour of a single to account for any further type 1 incitation (Carrere et al 1999).

Game selection
The games *Cops & Robbers*, *iDroidsMania* and *Inspector Gadget* were selected for this study on the basis of their rating provided by pocketgamer.com. These were 9/10, 7/10 and 5/10 respectively. These scores acted as a key metric in determining game quality prior to assessment. Before the behavioural outcome can be analysed further we must first draw reference to those ratings issued by the sample group. *Cops & Robbers* was rated on average as 6.6 (Range 4-8), *iDroidsMania* 4.1 (Range 1-5) and *Inspector Gadget* 4.7 (Range 3-6). The disparity between the pocketgamer.com ratings and given a part of this study should not be of concern. Indeed, an assessment of game rating across review sites quickly reveals a general variance in perceived quality. The reviews provided by the sample group whilst individually subjective, when aggregated, amount to overall congruent quotient of game quality.

Inspector Gadget was rated independently by Pocketgamer.com as 5/10, however the mean participant rating was 4.7. Thus, contrary to initial indications, participants of this study perceived *Inspector Gadget* to be of comparatively similar quality to *iDroidsMania.* The close correlation of its score to *iDroidsMania* (4.1) resulted in its exclusion from the key analysis of this study. Moreover, by not unnecessarily subjecting the data to statistical analysis, the possibility of inciting type I errors was also reduced.

RESULTS

As discussed in the previous section, the final two games used for analysis were *Cops & Robbers* and *iDroidsMania* with average user ratings of 6.6 and 4.1 respectively. The games were thus the most appropriate for use in determining behavioural differences between games of both high and poor quality. Non sequential metrics were assessed first.

State Frequency

Using a 2 x 5 repeated measures ANOVA, the significance of the frequency of the 5 behavioural states between both games was examined. The result revealed a main effect of state frequency was significant, $F(4,36)=33.0$, $p<0.001$. This suggests that some of the frequencies differed from each other. Importantly, a significant game frequency interaction, $F(4,36)=4.15$, $p<0.01$ was observed. Thus the frequency of some behavioural states differs significantly between games. As is clear from figure 1, whilst playing both games, participants were likely to experience more instances of Curiosity and Frustration whilst playing *iDroidsMania*. Contrastingly, the higher rated *Cops & Robbers* was more likely to induce instances of Engagement. States of Boredom were more apparent for *iDroidsMania* and Engrossment in *Cops & Robbers*, yet both were illustrated with minimal frequency. This result is the probable outcome of sample size and thus these two variables are deemed unlikely to be significant between games.

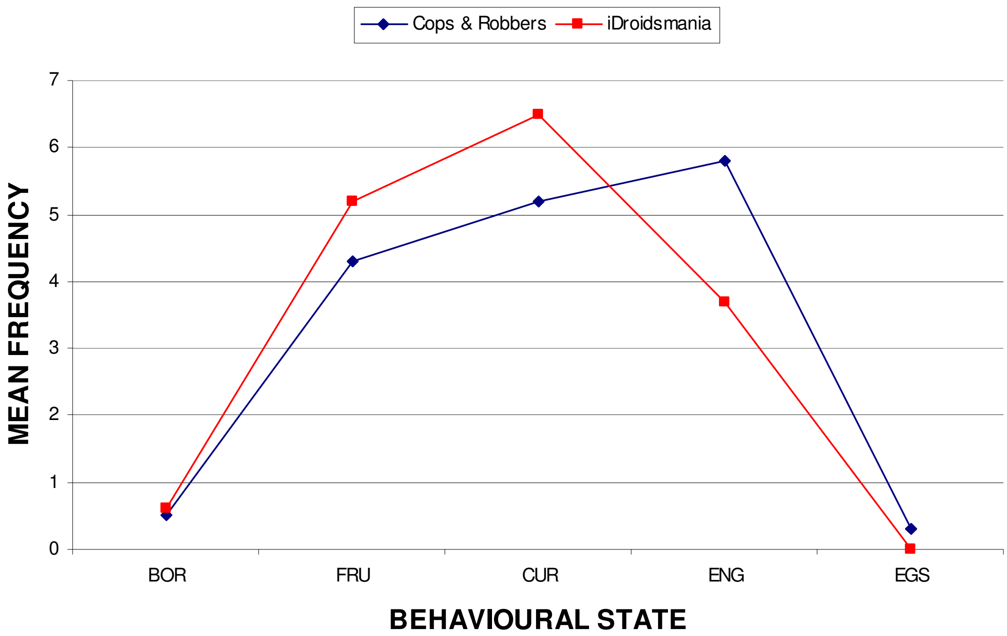

Figure 1: Mean frequency of behavioural states for both *Cops & Robbers* and *iDroidsMania*

To assess the key findings, further paired t-tests were conducted on the key behavioural states. Table 1 provides a data summary of state frequency for the occurrence of Frustration, Curiosity and Engagement for both *Cops & Robbers* and *iDroidsMania*. Both Boredom and Engrossment were omitted from further analysis on account of the minor deviation between games. The mean (M), standard deviation (SD) and paired t value for both games are illustrated.

[Game] Metric	Frustration			Curiosity			Engagement		
	M	(SD)	*t(9)*	M	(SD)	*t(9)*	M	(SD)	*t(9)*
[CR] Frequency	4.30	(3.52)	-1.22	5.20	(1.72)	-1.90	5.80	(1.83)	2.27*
[ID] Frequency	5.20	(1.94)		6.50	(1.57)		3.70	(2.45)	

*p<.05 [ID] = iDroidsmania [CR] = Cops & Robbers

Table 1: Summary of t(9) analysis for state frequency of *Cops & Robbers* and *iDroidsMania*

Whilst figure 1 illustrates clear trends, the higher frequency of Engagement for *Cops & Robbers* was the only state of statistical significance, $t(9)= 2.27$, $p<0.05$. This analysis reveals that a key distinction between good and poor games is how often a player is engaged. Specifically, games of a higher rating induce more instances of Engagement relative to those rated poorly.

State Duration

Similar distinctions are confirmed when the mean overall duration of states was examined. A main effect of behavioural duration ($F(4,36)=38.8$, $p<0.001$) illustrated that some behavioural states were experienced longer than others. The game and behavioural duration interaction was also crucially significant, $F(4,36)=11.8$, $p<0.001$. This indicates that between games some behavioural states were experienced for significantly different periods of time.

Figure 2 clearly illustrates that Engagement was experienced for a longer overall duration in *Cops & Robbers*. Contrastingly both Frustration and Curiosity where present for longer in *iDroidsMania*.

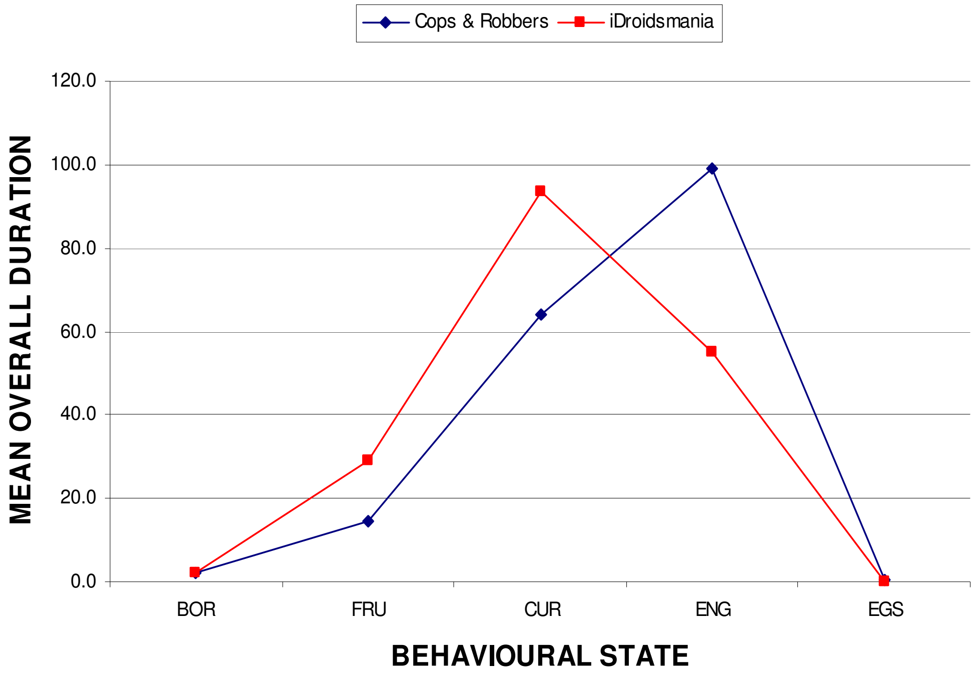

Figure 2: Mean overall duration of behavioural states for both *Cops & Robbers* and *iDroidsMania*

[Game] Metric	Frustration			Curiosity			Engagement		
	M	(SD)	*t(9)*	M	(SD)	*t(9)*	M	(SD)	*t(9)*
[CR] Total Duration	14.40	(12.56)	-2.94*	64.10	(28.78)	-2.73*	98.90	(30.21)	4.18**
[IR] Total Duration	29.10	(19.10)		93.40	(28.11)		55.20	(30.00)	

*p<.05 **p<0.01 [ID] = iDroidsmania [CR] = Cops & Robbers

Table 2: Summary of t(9) analysis for overall state duration of *Cops & Robbers* and *iDroidsMania*

The analysis confirms that both Frustration and Curiosity have a higher overall duration for *iDroidsMania* at a significance of t(9)=2.94, p< 0.05 and t(9)=2.73, p< 0.05 respectively. In line with previous observations the overall duration of Engagement in the higher rated game *Cops & Robbers* was again strongly significant, t(9)=4.18, p< 0.01. This data signifies that games of a higher rating resulted in players being engaged for longer periods of time and Curious and Frustrated for less relative to those rated poorly.

Behavioural Transitions

The analysis this far has focused on static measures of gamer experience. To help understand the dynamic aspects of gamer interaction relative to game quality we must assess how experience "reveals itself unfolded in time" (Bakemen et al 1997). The frequency of two sequence behavioural transitions were transposed into 5 x 5 confusion matrices for each participant and then aggregated for the sample group. A total of 20 possible behavioural transitions were mapped and the correlated transitional probabilities calculated.

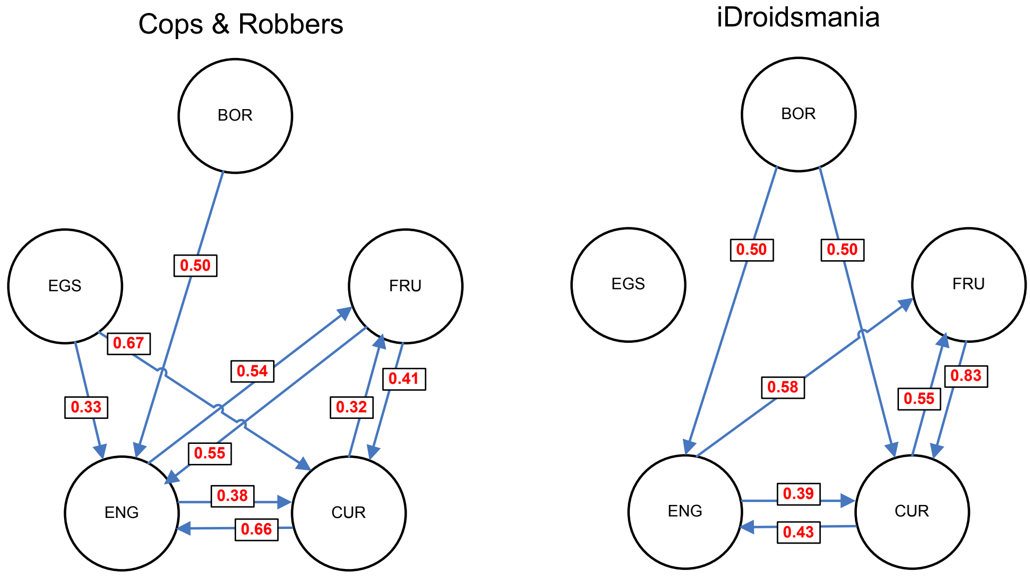

Figure 3: Transitional state diagrams of *Cops & Robbers* and *iDroidsMania* for transitions with a probability >0.25. BOR = Boredom, FRU = Frustration, CUR = Curiosity, ENG = Engagement and EGS = Engrossment

Figure 3 depicts the probability of behavioural transitions for each game, helping to highlight the transitional sequences of importance. Indeed as one would expect, for both games, transitions of proximal states were common whereas transitions between extremes of experience such Boredom to Engrossment were not.

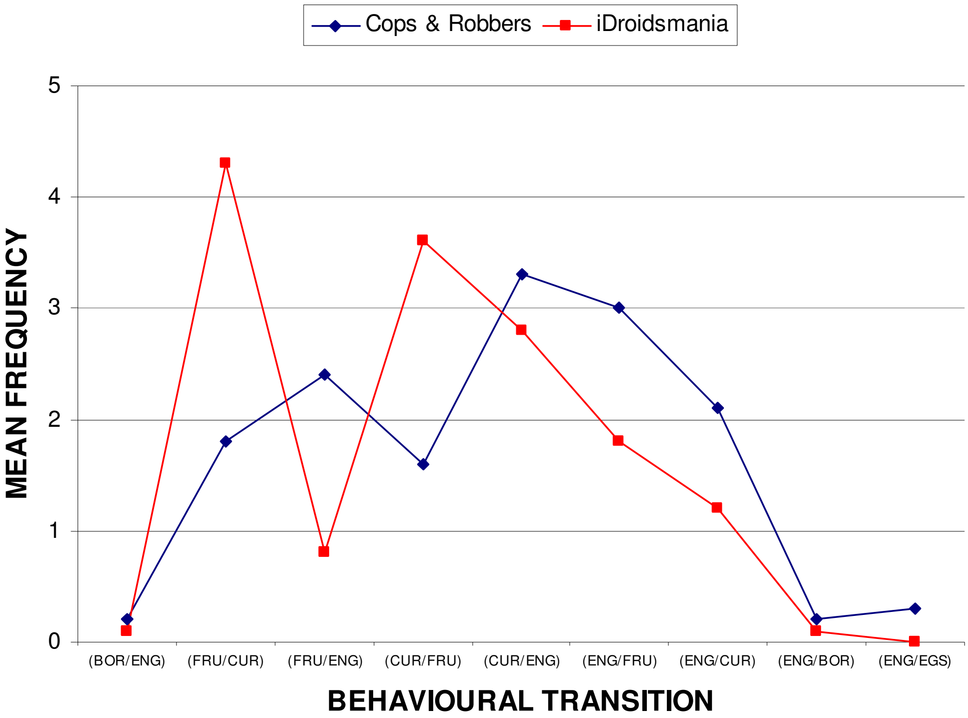

Figure 4: Frequency of Behavioural Transitions for *Cops & Robbers* and *iDroidsMania*

Figure 4 illustrates a main effect of transition ($F(8,72) = 11.2$, $p < .001$) in that for both games some transitions occur more than others. Importantly there is also a significant game transition interaction ($F(8,72) = 4.83$, $p < .05$) demonstrating a difference between games for some of the transitional states. This was explored further with the use of paired t-tests. As Figure 4 illustrates there was little differentiation between games for the Boredom/Engagement and Engagement/Boredom transitions and thus they were omitted from further analysis.

Table 3 provides a data summary of transitional state occurrence for both *Cops & Robbers* and *iDroidsMania*. Paired t-tests demonstrate that the Frustrated to Engaged state transitions were significantly more common in the higher rated *Cops & Robbers*, $t(9)=2.45$, $p<0.05$. Contrastingly Frustrated to Curious and Curious to Frustrated state transitions were more prevalent for *iDroidsMania* with a significance of $t(9)=3.41$, $p< 0.01$ and $t(9)=2.45$, $p< 0.05$ respectively.

Transition	Cops & Robbers		iDroidsmania		
	M	(SD)	M	(SD)	*t(9)*
Frustated / Curious	1.8	(1.40)	4.3	(1.49)	-3.41**
Frustated / Enagaged	2.4	(2.54)	0.8	(0.75)	2.45*
Curious / Frustrated	1.6	(1.91)	3.6	(1.43)	-2.45*
Curious / Engaged	3.3	(1.35)	2.8	(2.23)	0.57
Engaged / Frustrated	3.0	(2.86)	1.8	(1.72)	1.53
Engaged / Curious	2.1	(1.64)	1.2	(1.47)	1.41
Enaged / Engrossed	0.3	(0.46)	0.0	(0.00)	1.96

*p<.05 **p<0.01 1

Table 3: Summary of t(9) analysis for transitional state frequency of *Cops & Robbers* and *iDroidsMania*

Positive and Negative Behaviour over Time

As discussed, this study is primarily focused on how player behaviour for good and poor games differentiates itself over time. To explore this hypothesis further, the positive and negative states of both *Cops & Robbers* and *iDroidsMania* were examined through the course of each 3 minute gaming period. Positive behavioural states of Engagement and Engrossment where given a +1 weighting and the negative states of Boredom and Frustration were rated -1. Curiosity was considered neutral and given a weight of 0. Every 3 minute gaming period was divided in to six 30 second blocks. Each second was then given either -1, 0 or +1 score dependant on the behaviour exhibited. This provided a count of positive and negative behaviour for each block from which a net positive minus negative behavioural quotient could be derived (Table 4).

Transition	Cops & Robbers		iDroidsmania		
	M	(SD)	M	(SD)	*t(9)*
Frustated / Curious	1.8	(1.40)	4.3	(1.49)	-3.41**
Frustated / Enagaged	2.4	(2.54)	0.8	(0.75)	2.45*
Curious / Frustrated	1.6	(1.91)	3.6	(1.43)	-2.45*
Curious / Engaged	3.3	(1.35)	2.8	(2.23)	0.57
Engaged / Frustrated	3.0	(2.86)	1.8	(1.72)	1.53
Engaged / Curious	2.1	(1.64)	1.2	(1.47)	1.41
Enaged / Engrossed	0.3	(0.46)	0.0	(0.00)	1.96

*p<.05 **p<0.01 1

Table 4: Summary of positive (P) minus negative (N) behaviour over six 30 second time blocks

Figure 5 maps the net positive minus negative behavioural means for the each game over the 6 time blocks. A repeated measures ANOVA indicates a main effect of time suggesting that for both games net positive behaviour increases over time (F(5,45) =3.36, p<0.05). Critically there is also a significant effect of game, $F(1,9) = 29.2$, $p < 0.001$. The finding shows that overall positive behaviour is greater for *Cops & Robbers* than that for *iDroidsMania* at all data points. Both main effects of time and game are clearly identifiable in Figure 5.

The interaction of game and time was shown not to be significant ($F(5,45)=92.95$, $p>0.05$). Signifying that there was no appreciable variation in the difference of net positive values between games over time. Moreover the net positive behaviour for *Cops & Robbers* was greater than *iDroidsMania* at all data points. This lack of interaction negates the need for further analysis through t-tests. The data illustrates that the higher rated game demonstrated a net overall positive behaviour over time that was greater than that for the game rated poorly. This disparity between games at all data points illustrates not only that this is a consistent trait but it is one that emerges in the very early stages of gamer experience.

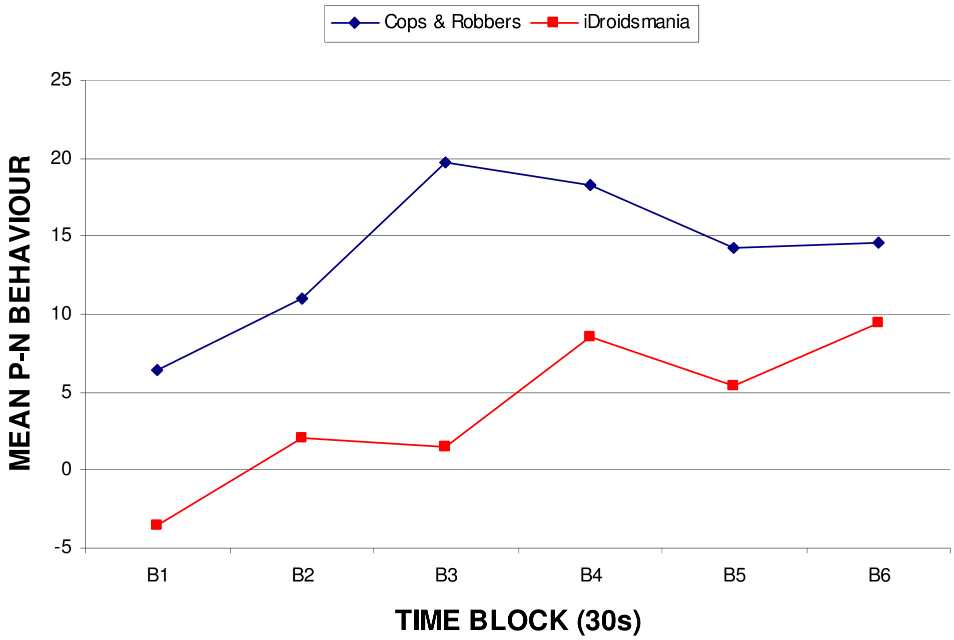

Figure 5: Positive (P) minus negative (N) behaviour over six 30 second time blocks for *Cops & Robbers* and *iDroidsMania*

CONCLUSIONS AND FUTURE WORK

This research has illustrated how both behavioural and sequential analysis can be applied to accurately unravel dynamic characteristics indicative of game quality. A mutually exclusive and exhaustive code of the behavioural states exhibited during gameplay was created through the analysis of user interactions with iPhone games.

Actively identifying the differentiating characteristics symptomatic of game quality, engagement was shown to be crucial in games of high quality, with both a higher instance of occurrence and overall duration. Curiosity and frustration by contrast were characteristically experienced for longer in poor games. Frustration was further illustrated to be a critical component in the interaction of games of both high and low quality games.

However, the sequential data demonstrated that as behaviour unfolds in time, it is in fact the transition from frustration that is most indicative of quality. Games rated poorly showed a significant occurrence of transitions between Frustration and Curiosity. This fluctuation is indicative of players' attempts to lower the barriers of access. This finding helps characterise curiosity as a behavioural period where the player is focused on overcoming the barriers of access required to achieve a higher state of engagement. Moreover, for bad games, players were shown to spend more time in this percussive state. In contrast for higher rated games, Frustration transitions to states of Engagement, indicating that for good games once barriers of access have been lowered, frustration can be overcome in a way that allows the player to return to Engagement.

Future work must address the limitations of this preliminary study; increasing sample size, game volume and cross platform applicability. Furthermore, this study only focussed on the initial moments of gameplay. Future studies may wish to examine how experience varies over extended periods. What is clear is that for this method to become truly relevant future research must focus on relating these behavioural states to elements of the game design. However, perhaps most compelling is the possibility of true regression correlation. By identifying the relation of game quality to one or more behavioural variables, a game's overall success in the market place could be predicted from just a very small amount of data. The impact of which would be dramatic for both the gaming industry and research alike.

REFERENCES

Artificial Life Inc. (2009). *iDroidsMania Lite.* [iPhone]. Artificial Life Inc. United Kingdom: Played 2009.

Bakeman, R. & Brownlee, J.R. (1980). The Strategic use of parallel play: A sequential analysis. Child Development, 51, pp 873-878

Bakeman, R. and Gottman, J.M., Observing interaction: An Introduction to Sequential Analysis, Cambridge University Press, Cambridge, 1997.

Brace, I., Questionnaire Design: How to plan, Structure, and write survey material for effective market research, Kogan Page, London, 2004.

Brown, E. and Cairns, P., A Grounded Investigation of Game Immersion, Proc. of CHI 2004, ACM Press (2004), 1297-1300.

Carrere, S. and Gottman, J., Predicting Divorce Among Newlyweds from the First 3 Minutes of a Marital conflict discussion. Family Process, 38, 3 (1999),293-301.

Nabil Chatbi. (2009). *Dark Hill Lite.* [iPhone]. FBS Software. United Kingdom: Played 2009.

Chen, H., Wigand, R. T., and Nilan, M. S., Optimal experience of Web activities, Computers in Human Behaviour, 15 (1999), 585–608.

Chillinqo Inc. (2010). *Zen Bound Lite.* [iPhone]. Chillinqo Inc. United Kingdom: Played 2010.

Csikszentmihalyi, M., Flow: The Psychology of Optimal Experience. Harper & Row. New York, 1990.

Czaja, R. and Blair, J., Designing Surveys: A Guide to Decisions and Procedures, Pine Forge Press, London, 2005.

Davis, J.P., Steury, K. and Pagulayan R.. A survey method for assessing the perceptions of a game: The consumer playtest in game design, Int. Journal of Computer Game Research. 5, 1 (2005).

Ermi, L. and Mayra, F., Player Centred Game Design: Experiences in Using Scenario Study to Inform Mobile Game Design. Int. Journal of Computer Game Research. 5. 1 (2005).

Fowler, J. F. Improving Survey Questions; Design and Evaluation, Sage, 1995.
Gladwell, M., Blink: The Power of Thinking without Thinking, Penguin Group. London, 2005.
Glue Games Inc. (2009). *Cops & Robbers Lite.* [iPhone]. Glu Games Inc. United Kingdom: Played 2009.
Gottman, J. M., Marital Interaction: Experimental Investigations, Academic Press, New York, 1979.
Gottman, J. M. and Levenson, R. W., Marital Process Predictive of Later Dissolution: Behaviour, Physiology and Health, Journal of Personality and Social Psychology, 63 (1992), 221-223.
Jennett, J., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T. and Walton, A., Measuring and Defining the Experience of Immersion in Games, Int. Journal Human-Computer Studies, 66 (2008), 641-661.
Malone, T. W., Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games, Proc. of the ACM and National Bureau of Standards Conference on Human Factors in Computer Systems, 1982.
McComas, J.J., Moore, T., Dahl, N., Hartman, E., Hoch, J. & Symons, F. (2009). Calculating Contingencies in Natural Environments: Issues in the Application of Sequential Analysis. Journal of Applied Behaviour Analysis, 42, Number 2, pp 413-423.
Morgan, D.L., Focus Groups, Annual Review of Sociology, 22 (1996), 129-52.
Namco Networks American Inc. (2009). *Inspector Gadget Lite.* [iPhone]. Namco Networks American Inc. United Kingdom: Played 2009.
Pace, S., A Grounded Theory of the Flow Experiences of Web Users, Int. Journal of Human-Computer Studies, 60, 3 (2004), 327-363.
Parten, M.B. (1932). Social participation among preschool children. Journal of Abnormal and Social Psychology, 27, pp 243-269
Pinelle, D., Wong, N. and Stach. T., Heuristic Evaluation for Games: Usability Principles for Video Game Design, Proc. of CHI 2008, ACM Press, 1453-1462.
Qin, H., Rau, P. P. and Salvendy, G., Measuring Player Immersion in the Computer Game Narrative, Int. Journal of Human-Computer Interaction, 25, 2 (2009), 107-133.
Rosenblum, L. A., The Creation of a Behavioural Taxonomy (in G.P. Sackett -Ed.), Observing Behaviour, Vol. 2, Data Collection and Analysis Methods, Baltimore, University Park Press, 1978.
Smith, P.K. (1978). A Longitudinal study of the social participation in preschool children: Solitary and parallel play re-examined. Developmental Psychology, 14, pp 517-523
Snyder, C. R. and Lopez, S. J., Handbook of Positive Psychology. Oxford University Press. USA, 2004.
Sussman, S., Burton, D., Dent, C. W., Stacy, A. W., and Flay, B. R., Use of focus groups in developing an adolescent tobacco use cessation program: Collective norm effects, Journal of Applied Social Psychology, 21 (1991), 1772-1782.
Tulathimutte, T. and Bolt, N., The science of fun: one-to-many moderated game research, Ext. Abstracts CHI 2008, ACM Press (2008), 2033 - 2038.
Zed Worldwide. (2009). *Bubble Boom Lite.* [iPhone]. Zed Worldwide. United Kingdom: Played 2009.