Hookham, Geoffrey; Nesbitt, Keith; Kay-Lambkin, Frances “Comparing usability and engagement between a serious game and a traditional online program”, Published in Proceedings of the Australasian Computer Science Week Multiconference Article No. 54 (Canberra, A.C.T. 2-5 February, 2016). (2016)

Available from: http://dx.doi.org/10.1145/2843043.2843365

Accessed from: http://hdl.handle.net/1959.13/1347627
Comparing Usability and Engagement between a Serious Game and a Traditional Online Program

Geoffrey Hookham  
University of Newcastle  
Callaghan, 2308,  
NSW, Australia  
+ 61 (02) 49215206  
Geoffrey.Hookham@newcastle.edu.au

Keith Nesbitt  
University of Newcastle  
Callaghan, 2308,  
NSW, Australia  
+ 61 (02) 49854519  
Keith.Nesbitt@newcastle.edu.au

Frances Kay-Lambkin  
University of New South Wales  
Sydney, 2052  
NSW, Australia  
+ 61 (02) 49216023  
f.kaylambkin@unsw.edu.au

ABSTRACT
A frequent motivation behind the development of serious games is the notion that game play provides a greater level of motivation or engagement that encourages players to perform better or complete more difficult challenges. There are in truth many aspects to concepts such as engagement, immersion, flow or other terms used to motivate the use of games in place of more traditional didactic training approaches. In this paper we discuss some of the issues of usability and its relation to engagement. This discussion is framed in relation to a usability study of a serious game that is being developed to support psychological counseling of individuals with comorbidity of depression and alcohol use disorders. We compare issues of usability and engagement in the serious game called Shadow directly with a more traditional online intervention program called SHADE. The study uses previously established surveys, namely, the System Usability Scale, the Perceived Usefulness and Ease of Use, and a modified survey for measuring Engagement to compare the two approaches. Results from a pilot study with 10 participants frame a discussion of the role usability plays in designing for engagement in serious games such as Shadow.

Categories and Subject Descriptors
Human-centered computing → Laboratory experiments  
• Human-centered computing → Laboratory experiments  
• Software and its engineering–Interactive games • Applied computing–Computer games

Keywords
Perceived Usefulness; Ease of use; Engagement; Usability; System Usability Scale; Serious Games

1. INTRODUCTION
The computer games industry continues to grow as an important segment of the entertainment market. As of 2015, there is also a significant growth in academic interest surrounding the use of such games for purposes that might be considered more serious than entertainment alone. “Serious Games”, as these applications have been called, have been developed to support domains as diverse as military training [72], finance [60], education [61], health [9] and even scientific discovery [23].

Terms like engagement, immersion, fun and flow are commonly used to motivate the development of serious games both for education and more generally. The premise is that by providing greater levels of engagement and enjoyment, players will become immersed and thus more committed to the experience. In education, for example, this suggests that greater levels of engagement will provide a significant educational advantage over more traditional educational approaches [5,11,15,36,70,82]. In the context of health, understanding how serious games can be designed to better engage players and applying this to health-related interventions has the potential to increase treatment completion rates, the translation of strategies to real world contexts, and the expected health outcomes [10,19]. Indeed, there has been a long history of developing serious games for psychological purposes (see recent review by Rodriguez et al. [63]).

However, engagement is a complex concept, encompassing varying components. Unfortunately, the term itself is often used without clear definition, which can lead to difficulties when it comes to measuring how engaged one is with an activity. This paper will begin with a discussion of the concept of engagement, describing key concepts related to the general notion of engagement, namely cognitive engagement, behavioural engagement and affective engagement. We shall further distinguish between the related concepts of immersion, presence, flow and absorption. Finally, we relate more traditional usability concepts and evaluations back to these discussions of engagement and serious games.

The paper also directly investigates notions of usability in relation to the development of a serious game called Shadow [39]. Shadow is a serious game prototype, developed from an existing online intervention program, called SHADE that uses a more traditional, didactic approach for psychological counselling. SHADE is a web-based intervention program for helping to treat binge drinking and depressed mood [44]. SHADE consists of information and interactive components, including case vignettes and in-session exercises. It provides tips for reducing alcohol consumption and improving mood. In this study the focus was on the “Mood Monitoring” and “Taking Charge of Your Thoughts” module of the SHADE program (See Figure 1). This module uses
cognitive behaviour therapy (CBT) and motivational interviewing techniques to encourage behavioural and cognitive change related to a person’s mood, and misuse of alcohol/other drugs. While the SHADE program has demonstrated efficacy in reducing depression and alcohol use over time [44], treatment completion rates are 36%, with average attendance at 5-6 out of the full 10 sessions. Sessions 5-6 within the SHADE program correspond to a move from behavioural tasks to more complex cognitive strategies, which seem to be difficult to relay in the more traditional interface style used with SHADE.

To try to improve completion of the SHADE program, and the ability to encourage program users to understand the cognitive tasks associated with SHADE treatment, the serious game called Shadow is being developed and evaluated as an alternative teaching method. The Shadow program was designed and constructed to be a “game” version of the SHADE program [39]. Shadow provides game players with several scenarios that they need to navigate through successfully using a range of CBT techniques. These scenarios are based on travelling to a party, and interacting with several people and situations along the way (See Figure 2). An additional interactive component of the game, focusing on “Mindfulness” allows players to manage their thoughts and control their mood by categorising harmful thought patterns (See Figure 3). These strategies are characteristic of the cognitive tasks associated with typical CBT programs, and the SHADE computer program in particular.

For research purposes, Shadow provides an in-depth case study into the design of a game where a key motivation is to better understand how serious games engage the user. We consider Usability to be central to this quest for engagement and have just completed a pilot study using mixed methods to investigate various issues of engagement, usability and affect. These methods include a think-aloud study [31], startle reflex modulation measures [14] and three different self-report questionnaires collected under experimental conditions. This particular paper reports on the results from the System Usability Scale [7,17] questionnaire, the Perceived Usefulness questionnaire [27] and a specific Engagement questionnaire [16]. These questionnaires we completed by 10 users who experienced both the SHADE and Shadow programs.

2. CONCEPTS OF ENGAGEMENT
The Oxford English Dictionary defines the term “engagement” as stemming from the verb “engage”, within multiple contexts meaning to bind, attach, or entangle, often within a voluntary action [58]. This definition highlights a potential confusion as the term “engagement” as it can be used within multiple contexts. Indeed, engagement is a multifaceted term referring to a person’s involvement in a specific activity. Stemming from examinations on pupil engagement in school-work [32], one succinct description of engagement suggests it is “a complex meta-construct with behavioural, affective and cognitive components that vary both situationally and dispositionally” [53]. We suggest that differentiating engagement into these components, “Behavioural Engagement”, “Affective Engagement” and “Cognitive Engagement” allows for more targeted measurement of the different aspects of engagement and also suggests a variety of approaches that might be required to better quantifying the various components of engagement.

Behavioural engagement is defined as focused activity on a task, with a typical measurement being time on task [4]. Behavioural engagement might thus be influenced by both motivational characteristics of situation and personality.
Cognitive engagement we define as mental activity associated with presented content, and is measured by successfully achieving the desired goal of the game, or by pre and post testing of outcomes. Other, more objective, tests related to physiology, such as eye-blink and scanning, have also been suggested to measure this aspect of engagement. Once again individual characteristics might impact on this type of engagement and dynamic difficulty or adaptive stair casing approaches have been used to try and ensure that challenges meet the playerschanging ability levels. This a key relationship often attributed to players being in a high zone of performance.

Affective engagement relates to emotional responses of players to game content. Affect is in itself a complex concept that can be further differentiated into emotion, cognitive and affective processing elements [76]. Affective processing might be measured in response to a simple emotional cue, and can be positive or negative. More complex emotional states such as curiosity, fun [45], interest and excitement can be measured using more traditional subjective feedback or approaches such as think aloud. Affect in interface design is primarily measured in terms of valence and arousal. In simple terms, valence refers to the positive or negative nature of the affect, and arousal refers to the intensity of the affect. Typically valence is objectively measured using the Startle probe [54] while arousal is measuring using a range of objective physiological measures such as skin conductance or heart rate that are known to vary with changes in arousal [76].

The concept of engagement in interface design has further been related to the notions of immersion, presence, flow and absorption [16]. Presence is a term often considered in Virtual Reality as it relates to how aware the user is of their real world environment while also feeling or present in the virtual one [8]. In terms of gaming, an immersive experience relates to the game’s capacity to induce the feeling of actually being a part of the game environment [79].

Like engagement and presence, the concept of immersion is also used in game research literature. The dictionary definition of immersion refers to the “dipping or plunging into liquid... and transference into other things“ [59]. In the context of video games, immersion refers to becoming absorbed in the game as though the player were in the game.

The two concepts of engagement and immersion are often used interchangeably in the literature. However a key distinction lies within how actively game content is experienced [29]. Immersion is a process through which one becomes drawn into, or deeply involved in the material, gameplay or story. Presence can then be seen as playing a part in inducing immersion. Both presence and immersion might be experienced in other, less interactive media, such as books, music and film. By contrast engagement implies a more active participation with the material [29]. The active engagement with material, whether solving puzzles, understanding concepts or overcoming challenges [29,81] links directly to the design of user interactions within video games.

The immersive and engaging capabilities of video games can lead the player to experience what is described as the ‘Flow’ state [25]. It is generally agreed that the concept of Flow is an identifiable and key element in the concepts of engagement and immersion. The Flow state is referred to, in common vernacular, as the feeling of being “in the zone”, with a limited awareness of one’s surroundings as attention is given entirely to the focus of the activity. The Flow state is an intrinsically rewarding state of extreme concentration or absorbed focus that can lead to being unaware of the passage of time.

Achieving the Flow state, combined with challenge, progression, and skill, can result in an optimal experience [24,25,42]. These optimal experiences might be considered fun, and intrinsically motivating, when a balance can be achieved between the challenge of the game and the skill of the player [40].

Although Flow is often presented as a continuum of experiences [24], it is also considered by some to be binary in nature [29]. That is, one can either be in the zone, or not in the zone [24]. Regardless of the binary nature of the Flow state there is generally a notion that immersion, engagement and flow can be related in some continuum of experience. For example, one model suggests three stages of immersion, where the first stage is Engagement; the second stage Engrossment; and the final stage relates to total immersion or the Flow state [24,42].

In examining the psychological notion that some individuals are more prone to experiencing engagement a further state, called absorption is highlighted [16]. In this model of behaviour, engagement is considered a continuum of experiences that begins with immersion, passes through presence and Flow and ends in Absorption [16]. Absorption describes total engagement in the current experience [41]. Like the Flow state, this suggests an altered state where consciousness of current thoughts and feelings is less accessible [35]. Like Flow, Absorption is a specific activity can considered as a binary state [62]. One distinction between Flow and Absorption is the intrinsically motivating or self-rewarding attribute associated with the Flow state. This might contrast with social rewards or goal achievement and task performance that serve as more outcome based motivations underpinning Absorption. Negative affects related to stress or frustration, for example with the user interface would be expected to block entry to the Flow state [33] but may not necessarily prevent Absorption.

In the next section we discuss the links between various aspects of usability and interface design to some of the concepts of engagement introduced in this section. As we have noted these terms are not always used consistently in the literature. In our further discussion we focus on the role of usability in one Immersion model that describes three stages of Engagement, Engrossment and Flow [18,24].

3. ENGAGEMENT AND USABILITY

The defining characteristics of a game are a goal, rules, feedback and voluntary participation [2,51]. A goal provides the motivation to the player; the rules define how that player can approach the goal; feedback informs the player of their performance; and voluntary participation implies the choice the player makes to enter the game world and interact with that world. The ease by which a player can enter the game world is tied to the game's design. The elements that comprise games, include the narrative, game mechanics and interface. The interface provides the means for players to interact with the game world and receive feedback from the changes they make.

Usability is a term that is applied to describe how easy an interface is to use and might be measured directly using criteria such as: learnability (how easy a task is to learn), efficiency (how quickly tasks can be performed), reliability (how error-free tasks may be), utility (the range of tasks that can be accomplished) memorability (how easy it is to remember how to perform tasks) [55]. Other more subjective criteria may be associated with the general notion of user satisfaction (how pleasant it is to use the interface). The satisfaction criteria may be particularly relevant for game acceptance and player engagement. However, games often require the player to perform tasks that allow them to
achieve the goal of the game so measures of efficiency and reliability can be just as relevant for maintaining player interest.

One way to consider the overall level of player interest is to use an existing Immersion model [18,24]. This model suggests three stages of immersion, namely Engagement, Engrossment and total Immersion (Flow). In the follow sections we relate the significant usability barriers that must be overcome to transition through these three levels of immersive interaction.

3.1 Engagement – Access and Time
To achieve the engagement stage of immersion it has been suggested that general barriers of access and time must be overcome [18,24]. The amount of time required will depend on the complexity of the game and how quickly the interface can be learnt. Access directly relates to both player’s genre preference and the type of game interface being used. The physical link between the player and the game needs to be as usable as possible to ensure that this particular barrier is breached.

For example, the natural motion control mechanism associated with the Nintendo’s Wii console illustrates of how usability in terms of interaction (motion control) and feedback (game play) can lend itself to easy adoption and enjoyment of games. The Wii has been used as a tool to examine social engagement [46] using video observation and questionnaires. This study found that though there was no performance gain, participants reported higher engagement using natural motion controls. In this instance the access barrier is mitigated by using familiar motion controls. Wii Sports [56] has also been compared with Angry Birds [64] in terms of intellectual, sensory, physical, emotional, narrative, and social engagement [67]. Using an adaptation of the Product Reaction Card sorting exercise [12] and semi-structured interviews it was found that Wii Sports primarily prompted physical, social and engagement, whereas Angry Birds triggered intellectual and physical engagement.

3.2 Engrossment - Game Construction
The next major barrier to the user developing a deeper immersion in the game, moving from Engagement to Engrossment, has been suggested as being more directly related to issues of game construction [18,24]. Engrossment, requires a deeper form of player involvement with the game content. With ‘good’ game construction, the game can trigger elements of presence in the player. The player can begin to feel as though they are ‘in’ the game and may become less aware of their physical surroundings [18,48]. However good game construction hinges on many usability elements. Game features such as visuals, tasks and plot [18] impact upon both the player’s attention and their desire to continue playing. This relates to the player’s emotional investment in the game, its characters or its story, otherwise known affectionate engagement. Players’ investment in games can hinge on bridging the gap between their expectations and what they’re allowed to do.

Various usability guidelines or heuristics have been developed to support designing game interfaces. For example, heuristics for enjoyable user interfaces [49] and the PLAY Principles for evaluating and developing higher quality games [28]. The PLAY Principles were evaluated using self-report measures and in industry, finding that these principles were a “useful structure for both the design teams and for HCI practitioners as a foundation for thinking about the user experience needs of gamers” [28].

The need to consider player needs in design forms the core of player-centric game design [3]. Player centric design is based on designing a game using a representative model of a target player. The designers’ duty is then to empathise with the player to create a game that entertains [3]. While most player-centric design practices occur before and during game development, an alternative approach is to design games that adapt to players. These games adapt difficulty, the game’s structure, and player progression and preferences based on player modelling [20]. Maintaining an engaging experience is the aim of these exercises. With affective and cognitive engagement tied to a game’s construction, breaking games into components may illuminate both how to measure engagement and how usability principles can overcome barriers.

There have been a number of efforts to try and define what elements of games are important for engagement [43,65]. Affect and engagement have been examined in the game-based learning environment called Crystal Island [66]. Pre, mid and post study questionnaires were completed by participants. The pre-study questionnaires gauged personality [50], goal orientation [30], emotional regulation strategies [34] and a curriculum test to measure prior knowledge. The use of this range of tests indicate the complex contextual issues that may vary significantly between individuals and impact on the report of affect related to engagement. The mid study questionnaire assessed cognitive and emotional state through the use of simple emoticons related to a range of positive (curious, excited, focused) and negative emotions (anxious, bored, confused, and frustrated). The post study questionnaires related to the students’ feelings of immersion. This study noted that the distribution of measured affective states indicated that “a carefully constructed game environment may serve to encourage positive affect during learning activities” [66]. Positive affect was considered to have a positive correlation with learning while negative affect was suggested to produce the contrary outcome [66].

3.3 Flow – Empathy and Atmosphere
The final barriers to total Immersion or the Flow state are related to empathy and atmosphere [18,24]. At this stage in the immersion model, the players' experience should be such that the game is all that matters. Csikszentmihalyi [26] describes nine attributes of the Flow experience:

1. There are clear goals every step of the way.
2. There is immediate feedback to one’s actions.
3. There is a balance between challenges and skills.
4. Action and awareness are merged.
5. Distractions are excluded from consciousness.
6. There is no worry of failure.
7. Self-consciousness disappears.
8. The sense of time is distorted.
9. The activity becomes autotelic (an end in itself)

The requirement of empathy with game characters and events and also the atmosphere created by the game world highlight the importance of narrative and also the need for a high quality interface. The player should have a clear goal where action is merged with concise feedback, there is no thought of failure and no distractions are present in the interface. Well-built games using appropriate usability heuristics may assist in improving the likelihood that a player would experience Flow. Regardless the Flow experience is fleeting and this highlights the difficulty of measuring such experiences.

4. Measuring Engagement
Much of the research into engagement and immersion has relied on qualitative or subjective methods. Jennett’s [42] work on
measuring and defining immersion provides a good example of this approach. Such subjective measures of engagement revolve primarily around questionnaires; asking what the participant perceived, thought, felt or experienced during the testing process. In a further example, Coller and Shernoff [22] studied student engagement in a game designed for Mechanical Engineering education, using the Experience Sampling Method (ESM) developed by Hektner, Schmidt and Csikszentmihalyi [37]. Participants completed the Experience Sampling Survey at random intervals throughout a week to gauge participant activity, and their affective and cognitive experiences. In this part of our study, we used three types of questionnaires, the System Usability Scale [7,17] questionnaire, the Perceived Usefulness questionnaire [27], and a specific Engagement questionnaire [16]. More detail on each is provided below.

4.1 System Usability Scale
The System Usability Scale (SUS) is composed of 10 questions that are scored on a 5-point Likert scale [17]. Questions alternate between positive (the interface is good) and negative statements (the interface is bad). The final score (between 0-100) is calculated from a simple formula that takes into account this reverse logic and combines and scales the ordinal Likert responses to fit into this range. Questions in the original survey instrument [17] refer to a generic “system”. These questions were later modified slightly, with a few word changes and so the questions referred to a “product” [7]. We adapted this later version of the questionnaire for our study by referring to a “game” rather than a “product” (See Table 1).

The SUS tool is a generic, quick and simple to administer questionnaire. While the tool measures usability across only a single, high-level dimension, it has been widely used across a variety of systems including games [7,52,68,77] and has the advantage that it allows for a broad comparison of usability even where the interfaces are not directly comparable [7].

In this study we slightly modified the SUS questions to determine any broad usability differences between SHADE and Shadow. We do not necessarily expect to note any differences in usability between the two systems but rather highlight possible problems that may interfere with engagement in either system.

4.2 Perceived Usefulness and Ease of Use
The original “Perceived Usefulness” and “Ease of Use” questionnaires were developed to better measure and predict user acceptance of new computer technology [27]. These two dimensions underpin the Technology Acceptance Model [2,27], one of a number of user acceptance models that was later updated to include social influence [73] and then formed part of a later, unified model of User Acceptance [74]. The original Technology Acceptance Model (TAM) discounted other influences related to the usage, or acceptance of technology into measures of a persons belief that the technology would enhance their performance in a task (Perceived Usefulness) and their belief about how much effort (Perceived Ease-of-use) was required to use the technology.

| TABLE 1: Questions for System Usability Scale (SUS) (Adapted from Brooke, 1996; Bangor & Miller, 2008) and Perceived Ease of Use (PEU) (Adapted from Davis, 1989) |
|-----------------|-----------------|
| **SUS 01** I think that I would like to use this game frequently | **PEU 01** I would imagine that most people would like to use this game frequently |
| **SUS 02** I found the game was unnecessarily complex | **PEU 02** I think that I would need the support of a technical person to be able to use this game |
| **SUS 03** I thought the game was easy to use | **PEU 03** I felt very confident using the game |
| **SUS 04** I thought that the various functions in the game were well integrated | **PEU 04** I thought the game was easy to use |
| **SUS 05** I thought that there was too much inconsistency in this game | **PEU 05** I thought the game was unnecessarily complex |
| **SUS 06** I felt that I would need the support of a technical person to be able to use this game | **PEU 06** I found that the various functions in the game were well integrated |
| **SUS 07** I would imagine that most people would like to use this game very quickly | **PEU 07** I thought the game was easy to use |
| **SUS 08** I found this game awkward to use | **PEU 08** I found that the various functions in the game were well integrated |
| **SUS 09** I felt very confident using the game | **PEU 09** I would imagine that most people would like to use this game very quickly |
| **SUS 10** I needed to learn a lot of things before I could get going with this game | **PEU 10** I thought that the game was unnecessarily complex |

The Technology Acceptance Model was arguably most influential as a management tool during the 1990s when many large organisations were changing over to new enterprise-wide IT systems and needed to understand the risks of user acceptance with various technology options. It has criticised for its limited predictive and explanatory power [21] and because in most parts it has ignored core social issues and changing attitudes that impact on technology acceptance [6]. Despite these problems the TAM provides a well-validated measurement tool that has been broadly used and empirically validated [2,38,69,71].

Perceived usefulness is measured in six questions (See Table 2) using a 7-point Likert scale. The questions are related to how the technology being considered might, allow the user to work more quickly (PU01), perform the task better (PU02), increase productivity (PU03), be more effective at the task (PU04), do the task more easily (PU05) or if the technology is generally considered useful (PU06). The questions used in our study have been adapted to refer to usage of either SHADE or Shadow.

Perceived Ease of Use is also traditionally measured as part of the Technology Acceptance Model using five 7-scale Likert questions. However, since we already ask relevant questions in the System Usability Scale questionnaire (See Table 1) we relate these back to the 5 core elements of the Perceived ease of use scale. That is, how easy is the technology to learn (SUS07, PEU01), is it clear and understandable (SUS02, PEU02), is it easy to become skilful (SUS10, PEU03), is it easy to use (SUS03, PEU04) and is it controllable (SUS08, PEU05).

While there is some argument that this does not fully replicate the original perceived ease-of-use survey, the intention of this work is not to replicate or validate the relationships between ease of use, usefulness and technology acceptance. In this study we are interested in trying to understand if there are any obvious variations in user perceptions of usefulness and ease of use and how these might impact on engagement in either SHADE or Shadow.
TABLE 2: Questions for Perceived Usefulness (Adapted from Davis, 1989)

<table>
<thead>
<tr>
<th>PU 01</th>
<th>Using Shadow/SHADE in my therapy would enable me to learn strategies more quickly.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU 02</td>
<td>Using Shadow/SHADE would improve the effect of my therapy.</td>
</tr>
<tr>
<td>PU 03</td>
<td>Using Shadow/SHADE in my therapy would enhance my skill (mindfulness and mood monitoring) development.</td>
</tr>
<tr>
<td>PU 04</td>
<td>Using Shadow/SHADE would enhance the effectiveness of my treatment.</td>
</tr>
<tr>
<td>PU 05</td>
<td>Using Shadow/SHADE would make it easier to participate in my treatment.</td>
</tr>
<tr>
<td>PU 06</td>
<td>I would find Shadow/SHADE useful in my therapy.</td>
</tr>
</tbody>
</table>

4.3 Game Engagement Survey
The Game Engagement Questionnaire (GEQ) adapted for use in this study consists of nineteen 5-point Likert scale questions. These questions were originally devised as a potential measure of an individual’s predilection to experiencing the stages of engagement [16]. This was intended to be used to try and predict the likelihood of individuals exposing themselves to violence in games [16].

The questions in this questionnaire were associated with different phases of engagement, beginning with immersion, through presence, flow and ending in absorption as defined in this model (See Table 3). These phases and related questions are suggested to move from lower levels of engagement, such as presence and immersion, to the more complex stages of flow and absorption.

In this study, the GEQ is used as a measure to indicate, in concert with usability measures, how engaged a participant is with the content of Shade or Shadow. The GEQ primarily relates to affective engagement, as it measures player states rather than their behaviour or thought patterns.

5. METHOD
Ten participants, 5 male and 5 female, within the ages of 18-30 were recruited for the study using poster and word of mouth. The participants consisted primarily of students at the University of Newcastle. No history of depression or substance abuse were required for participation, however participants were screened for suicide risk with provision of crisis information if required. Participants were required to have normal or corrected to normal vision. All participants were informed through a participation information statement about the intention and methods to be used in the experiment, including the fact that occasional startling noises would be played during the experiment.

All included participants completed two phases of the experiment, an evaluation of Shade and an evaluation of Shadow. These evaluations were carried out in two separate sessions at least a week apart. The order of evaluation was randomised so that five participants experienced Shadow, the game, followed by Shade, the online program. The other five participants experienced Shade first followed by Shadow.

The evaluation process was complex with a range of measures being taken including Startle reflex modulation [14] and think aloud [31] over a 10 minute period. Neither of these parts of the study are reported here. At the conclusion of each condition participants were asked to complete three questionnaires. Two of these were standard interface usability assessments, the System Usability Scale [7,17] (see table 1 for questions) and an adapted Perceived Usefulness survey based directly on the Perceived Usefulness and Ease of Use survey [27]. The third was a specially designed survey typically used to profile engagement behaviour to measure different aspects of user engagement [16]) (see table 3 for questions). Something about finding a correlation between usability and engagement? Or research questions of 1) is there a difference in usability and engagement between Shadow and Shade, and 2) does usability impact engagement. This is based on the theory above and provides some level of reason for testing.

TABLE 3: Questions for Game Engagement, organised by phases of engagement (Adapted from [16])

<table>
<thead>
<tr>
<th>IMMERSION</th>
<th>CEQ 18</th>
<th>I really get into the game</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEQ 01</td>
<td>I lose track of time</td>
<td></td>
</tr>
<tr>
<td>CEQ 17</td>
<td>I play longer than I meant to</td>
<td></td>
</tr>
<tr>
<td>CEQ 13</td>
<td>My thoughts go fast</td>
<td></td>
</tr>
<tr>
<td>CEQ 02</td>
<td>Things seem to happen automatically</td>
<td></td>
</tr>
<tr>
<td>PRESENCE</td>
<td>CEQ 16</td>
<td>Playing makes me feel calm</td>
</tr>
<tr>
<td>CEQ 15</td>
<td>I play without thinking about how to play</td>
<td></td>
</tr>
<tr>
<td>CEQ 12</td>
<td>Playing seems automatic</td>
<td></td>
</tr>
<tr>
<td>CEQ 07</td>
<td>I get wound up</td>
<td></td>
</tr>
<tr>
<td>FLOW</td>
<td>CEQ 05</td>
<td>The game feels real</td>
</tr>
<tr>
<td>CEQ 19</td>
<td>I feel like I just can’t stop playing</td>
<td></td>
</tr>
<tr>
<td>CEQ 06</td>
<td>If someone talks to me, I don’t hear them</td>
<td></td>
</tr>
<tr>
<td>CEQ 11</td>
<td>I can’t tell that I’m getting tired</td>
<td></td>
</tr>
<tr>
<td>CEQ 10</td>
<td>I don’t answer when someone talks to me</td>
<td></td>
</tr>
<tr>
<td>ABSORPTION</td>
<td>CEQ 09</td>
<td>I feel spaced out</td>
</tr>
<tr>
<td>CEQ 08</td>
<td>Time seems to kind of stand still or stop</td>
<td></td>
</tr>
<tr>
<td>CEQ 03</td>
<td>I feel different</td>
<td></td>
</tr>
<tr>
<td>CEQ 14</td>
<td>I lose track of where I am</td>
<td></td>
</tr>
<tr>
<td>CEQ 04</td>
<td>I feel scared</td>
<td></td>
</tr>
</tbody>
</table>

6. RESULTS
6.1 System Usability Scale
The twenty SUS scores for the 10 participants (in two conditions) were calculated (See Table 4). We expected a usability score of above 50 and ideally around 70 that would indicate that the interface is of acceptable quality for its intended use (Bangor et al, 2008). The results for Shadow (M=58.5, SD=19.8) and SHADE (M=57.3, SD=20.9) indicate a reasonable level of acceptance, although there was considerable variability in the 10 user ratings. Five participants rated the usability of Shadow higher than SHADE and 5 participants rated SHADE higher. Three users rated SHADE at least 15 points higher (17.5, 17.5, 27.5) on the SUS scale than Shadow. In contrast 2 users gave even larger differences in favour of Shadow over SHADE (32.5, 45.5).

We also note that the SUS measure is only a single reference score and not intended to be valid on a question-by-question basis and so we only analysed results for the high level SUS score only as it is intended, ignoring variations in individual questions [7].

TABLE 4: System Usability Scale Results(0-100)

<table>
<thead>
<tr>
<th></th>
<th>Shadow</th>
<th>Shadow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.2 Perceived Usefulness and Ease of Use

We calculated an average Perceived Ease of Use score for each participant by using selected questions from the SUS questionnaire (See Table 1, Table 5). With a range for 1-5 (1-low perceived ease of use and 5 for high perceived ease of use) there was no significant difference between Perceived Ease of Use for Shadow (M=2.4, SD=0.88) and SHADE (M=2.4, SD=0.94) and SHADE (M=3.7, SD=0.52).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Shadow</th>
<th>SHADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>37.5</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>77.5</td>
</tr>
<tr>
<td>3</td>
<td>42.5</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>82.5</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>47.5</td>
</tr>
<tr>
<td>6</td>
<td>42.5</td>
<td>70</td>
</tr>
<tr>
<td>7</td>
<td>90</td>
<td>85</td>
</tr>
<tr>
<td>8</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>9</td>
<td>67.5</td>
<td>35</td>
</tr>
<tr>
<td>10</td>
<td>77.5</td>
<td>32.5</td>
</tr>
</tbody>
</table>

TABLE 5: Mean Perceived Ease of Use for each Participant (1 Low - 5 High)

We also calculated a mean Perceived Usefulness score for each participant by adding together their Likert responses for each of the six questions in the Perceived Usefulness questionnaire (See Table 2, Table 6). To make the scales consistent with the Perceived Ease of Use we reversed the Usefulness measure by subtracting the Likert value from 8 and then scaling it to between 1 and 5. So with a range for 1-5 (1-low perceived usefulness and 5 for high perceived usefulness) there was no significant difference between the mean Perceived Usefulness score for Shadow (M=4.0, SD=0.50) and SHADE (M=3.7, SD=0.52).

6.3 Engagement Questionnaire

We calculated an average Engagement score for each participant by using the Likert responses from the Engagement questionnaire (See Table 3, Table 7). With a range for 1-5 (1-low engagement and 5 for high engagement) there was no significant difference between engagement for Shadow (M=2.6, SD=0.62) and SHADE (M=2.4, SD=0.82).

Finally we calculated the number of responses for each of the stages of engagement covered in the questionnaire for both Shadow (See Table 8) and SHADE (See Table 9). This allows for some level of comparison between the level of immersion, presence, flow and absorption reported by users in Shadow and SHADE. Due to the relatively small sample sizes used in this pilot study we did not check these results further for significance.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Immersion n=10</th>
<th>Presence n=40</th>
<th>Flow n=90</th>
<th>Absorption n=50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>1</td>
<td>9</td>
<td>50</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>11</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>4,5</td>
<td>4</td>
<td>20</td>
<td>21</td>
<td>8</td>
</tr>
</tbody>
</table>

TABLE 7: Mean Engagement score for each Participant (1 Low - 5 High)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Immersion n=10</th>
<th>Presence n=40</th>
<th>Flow n=90</th>
<th>Absorption n=50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>1</td>
<td>19</td>
<td>53</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>4,5</td>
<td>1</td>
<td>13</td>
<td>16</td>
<td>10</td>
</tr>
</tbody>
</table>

TABLE 8: Frequency of responses scores for engagement categories with Shadow (1,2–Disagree, 3–Not sure, 4,5–Agree).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Immersion n=10</th>
<th>Presence n=40</th>
<th>Flow n=90</th>
<th>Absorption n=50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>1</td>
<td>19</td>
<td>53</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>4,5</td>
<td>1</td>
<td>13</td>
<td>16</td>
<td>10</td>
</tr>
</tbody>
</table>

TABLE 9: Frequency of responses scores for engagement categories with SHADE (1,2–Disagree, 3–Not sure, 4,5–Agree).

7. DISCUSSION

In this pilot study we have used a number of questionnaires to try and measure both usability and engagement for two alternative approaches to providing assistance with psychological counselling. There was considerable variation measured with the group of 10 users. None of the measures indicated any significant difference in usability or engagement between the serious game Shadow and the online program SHADE. During the study we also collected a considerable amount of think-aloud data that will be analysed that may provide further insight into the role of usability and engagement. Although we did not necessarily expect to note any differences in usability between the two systems the variability in participant responses highlights the need for larger
sample sizes if the intention is to find significant differences using the questionnaires utilised in this study.

8. REFERENCES


[28] Desurvire, H., and Wiberg, C. 2009. Game usability heuristics (PLAY) for evaluating and designing better games:
The next iteration Online Communities and Social Computing (pp. 557-566): Springer.


Rovio Mobile. 2009. Angry Birds. Espoo, Finland


Wall, P., and Panksepp, J. 2013. Neuroimaging helps to clarify brain affective processing without necessarily clarifying emotions. INTECH Open Access Publisher.

