GAMIFYING PSYCHOLOGICAL TESTING:
INSIGHTS FROM GAMIFYING THE TAT

Thesis Presented

by

Borna Fatehi

to

The Department of Game Design

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Game Science and Design

Northeastern University

Boston, Massachusetts

December, 2017
Nowadays, gamification, using game elements in a non-game context, is becoming progressively popular in the research due to their affordances. Technology has the promise of increasing the accessibility and gamification has the capability of bringing engagement. This combination can increase participation in non-engaging contexts such as a test. In this study, we gamified a psychological test, Thematic Apperception Test (TAT), and ran an empirical study to test the effectiveness of such gamification. TAT consists a set of cards showing human figures in ambiguous situations and subjects write up stories about each card. The written stories are used to reveal aspects of the personality of the writers. We used a platform called Mad Science to implement our games and manipulated the test with achievement, exploration and social elements. In the empirical study, all participants did both the paper-based version of the TAT and the gamified version of it and rated their experience in each setting through surveys. We compared the results obtained from the traditional setting and the gamified version in terms of personality themes and detail of the stories to understand the consistency of results in the two settings and if stories in one setting could better reveal aspects of personality. We also looked at the survey results to understand the participants’ preference in each setting. Our results showed that the gamified version of the TAT provided a more motivating and enjoyable experience while at the same time could capture the essence of the TAT. The game could also reveal more personality themes in some of the cases.
TABLE OF CONTENTS

Abstract ........................................................................................................................................5

Introduction ................................................................................................................................5

Background ................................................................................................................................7

Affordances of game-like environments .....................................................................................7

Affordances of gamification of a psychological test .....................................................................7

The relation between in-game behavior and personality ...............................................................8

Games as assessment tools ...........................................................................................................9

Thematic Apperception Test ........................................................................................................9

Mad Science .................................................................................................................................10

Method ..........................................................................................................................................11

Participants ................................................................................................................................11

Materials .....................................................................................................................................11

Thematic Apperception Cards .......................................................................................................11

Game ..........................................................................................................................................12

Surveys .........................................................................................................................................14

Design .........................................................................................................................................15

Procedure .....................................................................................................................................16

Data Analysis .................................................................................................................................16

Result ..........................................................................................................................................19

Survey Result .................................................................................................................................19

Descriptive Statistics ....................................................................................................................19

Inferential Statistics .......................................................................................................................20
Gamifying Psychological Testing: Insights from Gamifying the TAT

Borna Fatehi
Advisor: Casper Harteveld

Abstract

Nowadays, gamification, using game elements in a non-game context, is becoming progressively popular in the research due to their affordances. Technology has the promise of increasing the accessibility and gamification has the capability of bringing engagement. This combination can increase participation in non-engaging contexts such as a test. In this study, we gamified a psychological test, Thematic Apperception Test (TAT), and ran an empirical study to test the effectiveness of such gamification. TAT consists a set of cards showing human figures in ambiguous situations and subjects write up stories about each card. The written stories are used to reveal aspects of the personality of the writers. We used a platform called Mad Science to implement our games and manipulated the test with achievement, exploration and social elements. In the empirical study, all participants did both the paper-based version of the TAT and the gamified version of it and rated their experience in each setting through surveys. We compared the results obtained from the traditional setting and the gamified version in terms of personality themes and detail of the stories to understand the consistency of results in the two settings and if stories in one setting could better reveal aspects of personality. We also looked at the survey results to understand the participants’ preference in each setting. Our results showed that the gamified version of the TAT provided a more motivating and enjoyable experience while at the same time could capture the essence of the TAT. The game could also reveal more personality themes in some of the cases.

Introduction

Game-like environments are becoming increasingly popular in research and have an enormous capacity for advancing social and behavioral sciences [1]. Games can supply an engaging environment for research purposes, education, health, to name a few, and their potential has been the subject of extensive research. However, application of games as a way to assess the psychological characteristics and individual differences has been studied only recently [2]. Players have different kind of interactions and a variety of options to choose in a game environment. The taken actions and decisions diverge among the players due to their individual differences and the investigation of them can reveal insights into the individual’s personality.

The idea of using games as an assessment tool first appeared at the end of the 19th century [3] when the potential of games for performance assessment was introduced [4]. Since then, the capability of games to show other forms of evaluation such as cognitive-performance [5] has been studied and games grabbed the attention of cognitive researchers to study games to understand the cognitive process by focusing on how
the players’ skill advancement happen in a game[6]. To understand to what extent the results of these assessments in games can be a true reflection of a player in the real world, researchers showed how the behavior of a player in a game is similar to his real life behavior [7] and how personality can describe the reason behind a chosen action in a game environment [8], [9]. Considering this potential of games, researchers started to think about making specific games for the purpose of assessment. As an example, Holmgard [2] made a game to assess intelligence.

The incentive behind embedding an assessment in a game setting was due to the affordances that such environment can provide. Gamification, or the use of game design elements in a non-game context, not only can have positive effects [10] in providing the motivation for a task, and therefore increase the engagement, but also may be able to address some of the challenges that a current assessment tool has; especially in the case of psychological tests.

Motivated participants, reliability, and validity are typical challenges for psychological tests [11]. Our hypothesis is that doing a psychological test in a game-like environment can address aforementioned challenges. Psychological tests are usually limited to be available only in clinics, narrowing its accessibility to a few people. Application of a test in a game environment can make that test available for more people. Moreover, participation in a psychological test might be motivating for only a small group of people. Games have the potential to provide an incentive for a group of people who may not be intrinsically motivated to participate in a test. Games are engaging in nature; hence, participants may be more willing to respond to a test in a game environment rather than in its traditional form, especially for time-consuming tests. In addition, a test in a game environment most likely is considered as a game rather than a test; consequently, it can ensure the performance of players’ typical behavior in that test. Since players can have access to games in a relaxed environment rather than a clinic or a lab, they might answer more honestly and accurately which can lead to a more reliable test result. Thus, games can increase accessibility, attract a larger population, ensure the typical behavior of subjects, and therefore, guarantee the reliability and validity of a psychological test.

For this study, we provided a framework to test our hypothesis. In this framework, we made games embedding one of the validated, widely used personality tests: the Thematic Apperception Test (TAT), and we made a comparison between the results obtained from the game and its traditional paper-based method. To test out the framework, we ran an empirical study on 18 participants. Participants do the TAT once in the gamified and once in its traditional non-gamified environment. Then, a comparison between the results is made for each individual to understand the consistency of the results in the two settings. In addition, participants self-report their experience in the two conditions and make comparisons between them through the surveys.

To make the evaluation of our framework possible we chose to use Thematic Apperception Test (TAT), because a) the test is widely used in clinical environments and b) the test’s validity and reliability has been shown in numerous studies [4], [12]–[22]. Using an acceptable test in this process is necessary because if we find inconsistencies in our study, we need to ascertain that the problem is with the framework rather than the test. The TAT consists a set of cards depicting human characters in ambiguous situations and the respondent should come up with stories about each card. Respondents usually project themselves in their stories, therefore, the stories are meant to reveal personality aspects of the writers.
We implemented two games in the Mad Science environment, which is a platform to perform experimental research. The characterization of the main game character in these games, which happens before a player is asked to tell his stories, is expected to make the player motivated and prepare him for the task that is about to come. In addition, we provided elements of achievement, exploration, and social elements in these games: achievement is addressed by manipulation of reward as a form of point; the exploration is given to the player by the available options during the conversations; the social element is addressed by the player’s interaction with the Non-Playable Characters (NPCs).

The result obtained from our empirical study supported the hypothesis showing that the gamification of a psychology test indeed has a positive effect both in providing an incentive for the participants for doing that test and, in providing more accurate, enriched test results. The self-report questionnaire showed a significant amount of preference, enjoyment, and motivation in the game setting over the paper-version setting for most of the participants.

**Background**

In this section, first, we discuss the affordances of game-like environments in general and the positive effects that “gamification” might bring. Then we describe the affordances of gamification of psychological tests and argue how such gamification has potential to address the current problems with psychological tests. After that, we investigate to what extent the result obtained from a gamified psychological test can be reliable and if games can reflect the players’ personality. Then we look into the literature that used games as a form of assessment and some examples of gamified psychological tests. Finally, we review the literature about the platform and the test that we used in our framework: the TAT and Mad Science.

**Affordances of game-like environments**

The first appearance of the term “gamification” goes back to 2008. Deterding et al. [23] studied the literature about gamification terms and came up with their definition as “the use of game design elements in non-game contexts”. Hamari et al. [10] provided a framework consisting motivational affordances, psychological, and behavioral outcomes to study the effectiveness of gamified environments. They concluded that gamification indeed provides positive effects; however, the effects are dependent both on the context and on the users. Cechanowicz et al. [24] gamified a survey, applied it to a real-world market research domain and showed it can significantly increase the participation.

**Affordances of gamification of a psychological test**

Usually, the scoring of psychological tests/methods requires adequate norms. These norms are made by collecting the answers of a broad range of demographics to that test/method. In fact, some of the critiques on some of the psychological test is that they are only applicable for a certain demographic since they are normalized based on them (e.g., [25], [26]). Recruiting a huge range of participants covering all the demographics is challenging and only a small group of people might be intrinsically motivated to participate in a research. Thus, for those who do not find participation in a research intrinsically motivating, there would be a need for an additional or a segregated element to keep them engaged [1]. Embedding a psychological test in a gamified environment might provide incentive while at the same time reduce the cost of deployment by making it available for a wider majority of subjects [2].
Besides, implementing a psychological test in a game environment will increase its accessibility. People might tend to avoid going to psychological clinics due to expenses or to preserve the social image. In most cases, people are not aware of the problems they may have. Nevertheless, after all an assessment game is still considered as a game rather than a test, and at the same time reveals information from the unconscious mind. This information will help the player to understand the reason behind some of his unconscious actions and decisions, which as a result facilitate personality improvements.

In addition, being nervous or being under the influence of stereotype threat [27] might cause the risk that subjects do not perform at their best in a psychological test. Furthermore, when participants are enforced in an artificial environment with paper-and-pencil tasks that do not provide immediate feedback, they may not respond authentically. Engagement in a game environment not only collects the subject’s information in a natural setting but also can encourage players to be involved in a task that requires a significant amount of time. Because games can be played anywhere and anytime, it extends the possibility of getting information from a broad range of demographics [1]. Thus, game-like environments can involve participants in authentic situations, gain behavioral data without grabbing participants’ conscious attention and attract a large range demographics in a sustained period of time [1].

The gamification of a psychological test not only can benefit that test but also can open a new advanced era for the future generation games. Nowadays, more games allow customization of game characters or settings for their players. Games like Elder Scroll Skyrim, Fallout, and Grand Theft Auto support a wide range of player types and experiences. The ability to embed a personality assessment tool in a game will allow the automation of setting or character customization and will expand the customizations’ capacity in games, and, therefore, offer more experiences that are meaningful to players.

**The relation between in-game behavior and personality**

To address the possibility of assessment through a game we need to know if the behavior of the players is related to their personality. In other words, we should know if the actions that players take within a game have some reflection of their personality.

Blascovich indicated that the behavior of people in virtual and real environments are similar [7]. Canossa showed that the interaction with NPCs and actions player take in game Fallout 3 can provide information similar to answering a personality test [28]. Van Lankveld investigated the player behavior in the game *Neverwinter Nights* and showed personality profile of the players are correlated with their in-game behaviors [8], [9] and Yee found similar results in World of Warcraft [29]. Canossa found that the behavior of a player in a game is not only related to personality but also the situation in which he is placed [28].

Thus, the relation between behavior in game and personality of the players are not only correlated but also different personalities follow different motivations and goals in a game. The history of this kind of studies goes back to Bartle’s anatomy of player who categorized players as socializers, killers, achievers and explorers [30]. In a follow-up study, Yee [31] confirmed explorers, achievers and, socializers in his empirical model as achievement, social and immersion components of MMOs. Jeng found that each personality has a specific motive and follows a specific goal in playing online video games [32]. Canossa et al. [33] studied the profile of Minecraft players and found that all players share the motives of curiosity, saving, independence, honor, and idealism while not being interested at all in acceptance or status.
Games as assessment tools

Thus far, we covered the affordances of game-like environments and showed that there might be an apparent relationship between personality and in-game behaviors. Now we will discuss studies that made use of games as an assessment tool and show the validity of assessment through a game.

Psychological assessment through tests and measurements started from the end of the 19th century [3] and the idea of games as assessment tools has a similar age. Jones et al. [4] and Jones [34] described the potential of video games for performance assessment.

The advancement of technology made it possible to test advanced contexts in recent years. Boot [5] reviewed the literature in the use of video games as tools to understand cognitive processes. Comparably, Gray [6] argued that studying game data could provide insight in how humans progress from novice to expert in games and theories of human performance. He believes that utilization of game data can benefit the advancement of cognitive, behavioral, or social science theories. Thomson and Tekofsky assessed performance speed and reaction time in games. Tekofsky [35] studied the relation between age and the video game’s behavior and found a correlation between age and initial play-style with a peak around 20, and decrease in performance and speed of play with the increase in age in Battlefield 3. Thomson [36] did a similar research with Starcraft 2 and found the reaction time in the game begins to slow down at age 24. Collmus discussed how games in social media reflect individual’s talents and how this information can be used for companies’ recruitment [37].

In terms of developing games specifically for assessment, Holmgard developed a game to assess the computational intelligence and cognitive performance [2] and Shute [38], [39] developed a game to teach some economic concepts and embedded the assessment of individuals’ advancement on those topics. She showed that providing such assessment results as a form of feedback to the player would enhance the learning procedure.

As we have seen, assessment games have been the subject of various research. But does the gamification affect the test performance? Attali and Ariel-Attali [40] ran two studies that gamified a mathematic assessment in a form that points were awarded based on accuracy and the speed of the answers. Although he had a small sample size, he concluded that the point manipulation had no effect on the main performance outcome and the response accuracy but increased the speed of response.

Thematic Apperception Test

Thematic Apperception Test is developed by Henry Murray and Christiana Morgan. The test consists of a set of cards depicting human characters in ambiguous situations (i.e., a young woman looking from an upper viewpoint to another young woman who is running). The respondent should write a story about (a) what led up to the events depicted on the card, (b) events are occurring on the card, (c) events will occur in the future and (d) the thinking and feelings of characters on the card [41]. TAT is an open-ended test with millions of possible stories to describe the cards.

Thematic Apperception Test is one of the widely used projective techniques in the clinical practice [15]–[17], [42]. It also provides the most comprehensive appraisal since it is vastly being researched among the projective techniques [43]. Watkins showed by surveying 412 assessment-active clinical psychologists that 90% were on an agreement that clinical psychology students should be competent of Thematic
Apperception Test and as another element of their survey they demonstrated TAT is one of the top 5 front runners with 82% of the clinical psychologists using it at least occasionally [17]. Being a highly used test, it suggests it should have good properties of reliability and validity.

Reliability. Based on Fineman study, the interscorer reliability of TAT indexes for achievement motivation has typically been between .8 to .9 range [12]. However, the test-retest reliability has a different story. Winter and Stewart found that the test-retest reliability was higher when the subjects were told to either a) write the same story or b) not worry if their story is similar or different. Nevertheless, it was not better than chance when they were asked to write a different story. To tackle that issue they suggested having some instructions in each test to “pull for consistency (reliability)” such as “give the first natural answer that comes to your mind”. [13]. Based on Mayer’s study “TAT have reasonable evidence supporting its reliability and validity and is not noticeably deficient in its psychometric properties relative to other assessment procedures commonly used in psychology, psychiatry, and medicine” [22].

Validity. The TAT in achievement motivation and task incentives has been validated in a number of studies [12], [14], [18], [44]. McClelland realized managers in non-technical fields with leadership motive pattern in TAT showed significantly higher levels of management advancement 8 years after testing than those without it [19]. Mussen showed a strong correlation between overt aggression and expressed aggression on TAT test. [20] Karon and O Granny found a high predictive validity for the ratings of emotional health from the TAT in schizophrenics patients [21].

The advantages of a projective test over a questionnaire are their capability to a) bypass the conscious defenses of the respondents, b) allow the access to the unconscious mind and c) prevent participants from hiding aspect of their personality to maintain social image [45], [46].

**Mad Science**

Mad Science is an easy to use platform made to help researchers make experimental research. It provides the ability to make any kind of behavioral or social scenarios in a 2D environment. Mad Science consists of four elements: a) character creator: to build or customize characters, b) scene creator: to make scenes/backgrounds to place objects and characters on it, c) scriptor which is mostly the logic and sequence of the game and d) manipulator to set dependent and independent variables [47].

Each project in Mad Science consists of several scenes. A background can be assigned to each scene and there are several built-in characters and objects, which can be placed on them. Each scene has its own script in which the logic of the scene can be defined: Dialogues and the choice of them, movement of characters, visibility of the objects, the setting of the variables and the transition between the scenes are defined in the script section. All the information is provided to the player through dialogue. The user interaction in a scene is either by the choosing dialogue options or by writing open responses in a textbox.

The evaluation of Mad Science’s potential has been done through replicating experiments [48], [49].
Method

Participants

A total number of 18 participants, who are students at Northeastern University, have been recruited. Participants are mixed gender (7 female, 11 male), mixed race (6 Asian, 12 White), at least have some college degrees (2 some college degree, 9 four year degree, 6 professional degree and, 1 doctorate) and are aged between 20 and 33 ($M = 26.61$, $SD = 3.55$).

Materials

Thematic Apperception Cards

We chose six cards from the TAT cards in total to be used in the settings. Three of the cards are used in the game setting and the other three are used in the control setting: The cards in the two versions of the games are the same but different from the cards in the control setting. Since the participants will play only one of the games, we chose the same cards to maintain consistency. On the other hand, since all the participants will do both the control setting and one of the game settings we chose different cards to avoid fatigue effect. Although the cards in the game setting and the control setting are different, they share the same themes. We asked two experts in the psychology field to confirm the type of cards chosen and the similarity of their theme. In Table 1 the cards used in the control setting and the game setting are shown. The cards in each row share the same themes with each other. In the first row, both characters’ face is covered and their gender is not specified. In addition, their gesture is somehow similar. In the second row pictures, there are two same gender characters, which one is young and the other is old. Finally, in the third row, the characters are the same gender in each picture and both share the same theme of the dominance of one character over the other.

Table 1 - The TAT cards in the control setting and the game setting - Cards in each row share similar themes

<table>
<thead>
<tr>
<th>Control Setting</th>
<th>Game Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Control Setting Card" /></td>
<td><img src="image2.png" alt="Game Setting Card" /></td>
</tr>
</tbody>
</table>
We have implemented The Thematic Apperception Test in two different game settings and one control setting, which is aimed to function as the paper-pencil version of the test, on Mad Science platform. The settings are as follows:

1) Game 1: The game character is a talented but desperate freelance writer who gets happy only when he writes. He gets a phone call from one of the companies that he works with and gets informed that his writings have been received good feedback from readers. The manager of that company decides to allocate a story section in the next edition of the magazine and nominate it for national magazine awards. The story section would be TAT images and the main character’s writings about them. The main character is asked to go to the company to write stories about the cards he sees with addressing the four questions according to the TAT procedure. The four
questions are: (a) what led up to the events depicted on the card, (b) events are occurring on the card, (c) events will occur in the future and (d) the thinking and feelings of characters on the card.

2) Game 2: The game character is a famous critic and is invited to one of his friend’s gallery to provide her a report about her work. The work in the gallery are TAT cards and the NPC friend asks from the player the four main questions in form of a conversation while the player is seeing the images.

3) Control setting: There is no gamified element. There is only text giving instructions to the participant to write stories about the cards as the cards appear. We implemented the paper-version in the Mad Science environment to keep the input method consistent with the game versions.

In the first scenes of the games, we tried to make characterization for the writer and the critic. We showed some ordinary life of the main character which is all about writing. There is a mood meter, which indicates the happiness and the sadness of the main character. The mood meter goes high as the player starts to write or receive a compliment about his writing and goes down as he chooses an inappropriate choice of dialogue in conversation: e.g. reject to write for a magazine.

The initial characterization in the game is expected to make the players believe that they are a famous writer and motivate him in the task of writing stories about the cards. In addition, we tried to provide motives for the different player types of Bartle’s anatomy of player, which were also confirmed in the empirical research of Yee [31]. The mood meter can facilitate the gameplay for achievers: They might be incentive to keep the main character in a good mood. The availability of various options in the dialogues is aimed to motivate explorers: They might enjoy choosing various options in the dialogues and see the outcome of their choices. Finally, the presence of an NPC character in the galley in the second game and the conversation between player and her is meant to address the incentive for socializers.
Surveys

The survey aims to measure three aspects of demographics, gamer profile, and experience with 5-point Likert items for each question. There are four surveys in total: Post paper-version questionnaire, Post game-version questionnaire, Comparison questionnaire and the demographics questionnaire.

In the post paper-version questionnaire, the participants rate their experience about the control setting. Participants should rate their enjoyment and encouragement to write in detail (from a great deal to not at all), the length of the experiment (from extremely reasonable to extremely unreasonable), their comfort to write honest (from extremely comfortable to extremely uncomfortable). In the final question, we asked them if they think their experience could be better if they would write the story in a word document or on a hard paper (from far above average to far below average), to understand to what extent the implementation of the test in the Mad Science environment could have any effect on their answer. In the end, they are asked to address any problem they had in doing the test or any additional comments.

The game questionnaire consists of the same materials along with believability of the characters and game environment (from extremely believable to extremely unbelievable). They should also rate to what extent: a) they could relate to the characters, b) they found the game aesthetically pleasing c) they immersed in the game world (from a great deal to none at all). In addition, there are some questions about the game itself and the playstyle of the players in it, which they should rate from strongly agree to strongly disagree. They should rate to what extent they a) tried to keep the mood meter as high as possible, b) tried to describe the...
images in detail, c) were careful to address everything that was asked about the images, d) tried to make no errors, e) wanted to enjoy the game, f) wanted to try the options they had, g) took the responsibility to accomplish the tasks well; while playing the game. Then they rate the game’s logical flow, the funniness of the game story and, the easiness of a) using game interface, b) learning how to play the game and, c) following the story of the game with the same Likert scale items. Since Mad Science does not have advanced graphics, we asked them to rate if they found the games’ graphic distracting.

In the comparison questionnaire, participants should compare all the aforementioned aspects of the game and the control setting. They answer to what extent a) do they think the game can capture the essence of the test in the paper version (from a great deal to none at all) b) they agree with “I preferred the game version over the paper version” (from strongly agree to strongly disagree). Then they compare in which experiment they felt most comfortable to write a) honestly and b) in more detail, which experiment they c) enjoyed the most, d) found too lengthy. The available options for these questions are paper-version, game-version, equal for both with the exception of “found it too lengthy” question that has a fourth option of “I did not find any of them lengthy”.

Finally, in the demographic questionnaire, general demographic of the players including age, gender, education, and ethnicity of the participants are collected along with their gaming profile. In the gaming profile, they should address the amount of gameplay in a week, rate the game genre they like most and rate the aspects of the games they care most by sorting the available options. The game aspects are: Spatial Involvement (Having the sense of being connected to game space with exploration), Affect Involvement (The game feels real and ability to evoke emotions), Narrative Involvement (The game’s story and narrative), Ludic Involvement (The goals in a game and the strategy or tactics required to achieve that goal), Kinesthetic Involvement (The feelings of interaction with controls and mastery of them), and Shared Involvement (Competition or collaboration with other players or other characters in a game).

**Design**

The experiment is designed within subjects in the control and the game setting: The result of the first phase is compared to the result of the second phase for each participant. The independent variables are the game setting and the control setting, and the gaming profile. The dependent variables are details of stories, associated themes to them, and, participants’ self-report of enjoyment and motivation.

We looked into the a) stories in the game and the control version, b) post-questionnaires, and c) comparison questionnaire. We compared the stories of the participants to understand: a) if the game could capture the same themes as the paper version, b) if the stories in one setting were more detailed, and c) if one setting could reveal more personality traits or more themes. We looked at the survey data to understand if participants had more motivation, enjoyment, or preference in any of the settings.

In summary, in order to compare the stories in the game environment and the paper version, we are looking into four factors (dependent variables):

a) If stories in one setting has more describing details than the other

b) If stories in the game setting could reveal the same themes found in the control setting

c) If stories in one setting could reveal more themes of personality than the other
d) The post questionnaires and the comparison, which is aimed to ask participants’ ideas about their overall experience in the two different settings

**Procedure**

The study is designed into two phases with a one-week gap in between. In the first phase, participants do either the control version or one of the game versions. Then they answer the related questionnaire based on what they have done. In the second phase, they do the other setting: the ones that do the game setting in the first phase do the control setting in the second and vice versa. Then they answer the related questionnaire, comparison questionnaire, and the demographic questionnaire. To avoid introduction of confounding variables we used counterbalancing: The order of doing the experiments is randomly assigned to each individual. In the second phase, they were told to not worry about their answer be similar or different to their previous experience. The participants could choose to do the test in a lab or do it at their convenient place. A link to the experiment, surveys, and their assigned ID emailed to the latter group.

The participants were randomly assigned to one of the two game settings: eight played the first game and nine played the second one. In terms of ordering of the settings, five participants did the paper version first and the rest 12 did the game version first. All the participants completed the surveys, control setting, and game setting. The one-week gap is designed to reduce the memorability effect. Table 2 shows the summary of this procedure.

**Table 2 - Summary of the procedure**

<table>
<thead>
<tr>
<th>The experiment</th>
<th>Game-1 and Post Game Survey</th>
<th>Game-2 and Post Game Survey</th>
<th>Control Setting – Post paper-version Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Participants in the</td>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>First Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Participants in the</td>
<td>3</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Second Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>11</td>
<td>18</td>
</tr>
</tbody>
</table>

**Data Analysis**

The TAT stories are meant to reveal some aspect of personalities of storywriters. In other words, participants present some aspect of themselves in their stories. As people write more detailed, enriched stories describing the feelings they get from the pictures, they give more clues about their personality.

The current problem with the TAT is that usually people do not describe adequate and thus their stories alone cannot provide enough information about them. With the implementation of the test in a game environment, we are trying to provoke them for detailed stories that can provide enough information for analyses of their personality. If the stories in the game environment could reveal more themes or more aspect of personalities, the game could better bypass the conscious mind and defense mechanisms, therefore, it could be more successful with the aim of TAT.
The reports for the questionnaire are for all the 18 participants but we only selected nine of them for detailed analysis.

Analyzing the themes in the stories

The stories have been examined qualitatively by having the researchers reading all the stories. In the procedure, we looked for the themes associated with the stories. We tagged the stories with themes. In the process of tagging, the information of the participants was removed from the stories and, the comparison and re-attaching the participant’s info happened after the coding has been done.

We used the Personality Research Form (PRF) scales developed by Jackson [50] to understand the story themes. Based on PRF there are 22 main scales associated with personality traits. The scales are relevant to the functioning of individuals in a wide variety of situations and are used to understand the areas of normal functioning rather than psychopathology. We used PRF since a) It is based on the Variables of Personality set defined by Henry Murray [51], the same person who made the TAT, therefore a good match with the TAT b) it is better than the Variables of personality because it is modified with much research evidence [50]. Some examples of the scales are Achievement: maintains a high standard and is willing to work toward distant goals, Affiliation: Enjoys being with friends and people in general; accepts people readily and, Play: does many things “just for fun”.

We read all the stories in the two phases and tagged them with the PRF themes. As we stated earlier, we removed all the info associated with stories while reading them to avoid any bias. Then we compared the themes that were found in the settings for each participant. We only used a few set of cards to understand the affordance of our game environment; therefore, we expected that not all the themes might appear in the writings. Based on the nature of the cards we used, some of the themes such as Affiliation and Nurturance appeared in high frequency among the subjects, some of the themes such as Play and Harm avoidance appeared only for a few subjects, while some of the others such as Infrequency or Desirability never appeared in any of the subjects. To give the readers some example of how we used the themes, we tagged “The woman killed the man” with aggression, “The man in the top left is trying to persuade the man in the bottom right” with dominance and, “Younger woman seems to be certain, decisive, bold and strong” with autonomy. The full coding of the themes associated with the stories of each participant in each phase is available in the appendix.

Analyzing the details in the stories

As we discussed, the describing details might lead to revealing more information about a person. A number between zero and three is assigned to each story independently scoring the describing details.

Three is assigned to the writings that have fully shape of a story with described feelings and the thinking of the characters. Two is assigned to the writings that are less descriptive but has satisfied all the four aspects of TAT and has semi-form of a story. One is assigned to the writings that have only addressed the four criteria and reveals some information about how the writer has interpreted the picture. Finally, zero is assigned to the writings that are not in a form of story, has not addressed all the four criteria, or does not give any information about the writer. Examples of each score for the first picture of the game version along with explanation are given below to give the readers a general idea about the scoring procedure.

An example of writing with the score of 3 on the detail:
‘He had asked her to come to the room and thats when he told her, her mother has passed away.
Devastated by the news and crying she opened the door to let some light in, so maybe she can think clear
as if the sorrow and dark have blocked her mind. She opens the door, looks at his face through all the
tears and asks for more details.”

This writing, although short, has a form of story and a reader can understand the framing of the writer. It
has described the feeling of the character very well and also has addressed the rest three areas of what
happened before, what happens after, what has led up to the event.

An example of writing with the score 2 on the detail:

‘‘A woman is holding her head in her hand. She is experiencing some negative emotion. Her other hand is
holding open a door. The work is in black and white. She is feeling sad and disappointed. Her boyfriend
was cheating on her. She gets over her trauma and moves on’’

This writing has a frame that addressed all the four areas. But it has a semi-form of a story and the sentences
are not very connected to each other.

An example of writing with the score 1 on the detail:

What is happening: Somebody knocked the door, she was sleeping, just woke up, trying to open the door
for the person.

The feeling of the characters: Sleepy, impatient

What happened before: Somebody knocked the door

What happens after: she is going to open the door for the person behind the door

In this scenario, there is only an explanation of an event happening and the actions taken by the character.
It is not a story but the writer has more or less tried to address the four areas.

Example of a story with score 0 on the detail

‘‘She looks upset

Frustrated

Lack of sleep

She's going to have a nap”

This writing does not have a form of a story at all and the addressing of the areas is very short and concise.
In addition, the feelings are not really connected with each other: Frustration and upset due to lack of sleep.
This writing does not give much information about the writer.
Result
Survey Result
Descriptive Statistics

Post Paper Questionnaire
Despite that most of the participants (83.33%) found the length of the experiment reasonable or extremely reasonable and were encouraged at least a moderate amount to describe in detail, only three of them enjoyed the experiment a lot or a great deal and the rest (83.33%) reported an enjoyment of moderate amount or less. Nevertheless, almost all of them (17) were extremely or somewhat comfortable to write honestly. The paper-version was not actually on a hard paper or writing on a word document and seven participants (38.89%) reported they could have a better result, above average or far above average, if they could write their stories on a hard paper or on a word document.

Post-Game Questionnaire
Regardless of which game setting participants played, 94.44% enjoyed the game experience at least a moderate amount and only one of them did not enjoy at all. At the same time, 83.33% found the length of the experiment somewhat or extremely reasonable and 72.23% found the overall game experience motivating and aesthetically pleasing at least a moderate amount. About the same number of participants self-reported that the game was extremely or somewhat believable and the game definitely or probably encouraged them to explain the stories in detail. In addition, 61.11% of the participants reported they immersed in the game world for a moderate amount or more but fewer of them (55.55%) mentioned that they could relate themselves to the game character.

About 94.44% of the participants were careful to address everything that was asked about the images and aimed to accomplish their tasks well. Sixteen (88.89%) tried to make no errors and the same number reported they wanted to enjoy the game. 77.77% tried their best to describe the images in detail and 72.22% wanted to try the options they had in the dialogues. Only ten of them (55.55%) tried to make the mood meter as high as possible.

More than 83.33% of the participants reported that the game had a logical flow, the interface was easy to use, it was easy to learn to play the game, and they could follow the story easily. Although the games did not have very advanced graphics due to the limitation of Mad Science, 61.11% of the participants did not find the games’ graphic distracting. Only 55.55% found the story of the game fun.

Comparison Questionnaire
Based on participants self-report, 83.33% believed that the game could capture the essence of the test in the paper version at least a moderate amount and 61.11% strongly or somehow preferred the game version to the paper version and 22.22% had neither preference. The rest of the results are shown in Table 3.

<table>
<thead>
<tr>
<th>Comparison of the experiments in terms of:</th>
<th>Game Version</th>
<th>Paper Version</th>
<th>Equal for both versions</th>
<th>None of them</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing more honestly</td>
<td>38.89% (7)</td>
<td>27.78% (5)</td>
<td>33.33% (6)</td>
<td></td>
</tr>
<tr>
<td>Writing in more Detail</td>
<td>44.44% (8)</td>
<td>44.44% (8)</td>
<td>11.11% (2)</td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>66.67% (12)</td>
<td>16.67% (3)</td>
<td>16.67% (3)</td>
<td></td>
</tr>
<tr>
<td>Finding it lengthy</td>
<td>38.89% (7)</td>
<td>22.22% (4)</td>
<td>1.11% (2)</td>
<td>27.78% (5)</td>
</tr>
</tbody>
</table>
Inferential Statistics

As we discussed earlier, in both the post-paper questionnaire and the post-game questionnaire, participants ranked to what extent they enjoyed each experiment from a great deal (1) to none at all (5). We found no significant result between the means of enjoyment in the two settings by including all the participants. However, by looking at the gamer profile of the participants we found interesting results.

Among the gamers, participants who reported playing games more than zero hours weekly, the report on the level of enjoyment in the game version (N= 14, M = 2.36, SD = 1.08) was numerically higher than the paper version (N=14, M = 3.07, SD = .83). To test the hypothesis that the gamers statistically significantly enjoyed the game version more than the paper version, a Wilcoxon paired samples rank test was performed. Prior to conducting the analysis, the assumption of homogeneity of variances was examined through the Pitman-Morgan test. The assumption was considered satisfied, as the spearman’s correlation between the sum of the respective paired scores and the difference between those was estimated at \( r = -.246, p=.396 \). It will also be noted that the correlation between the two conditions themselves was estimated at \( r = .621, p < .02 \) suggesting that the Wilcoxon paired sample test is appropriate in this case. The null hypothesis of equal rank means was rejected, \( W(10) = 4, z = -2.33, p = .020 \) with a medium effect size (\( r = .54 \)). Thus, the enjoyment in the game version was statistically significantly higher than the paper version. The median score on the enjoyment decreased from paper version (\( Mdn = 3: \) A moderate amount) to game version (\( Mdn = 2: \) a lot).

We performed the same test for the rest of the paired variables in questionnaires but did not find any statistically significant difference between the ranked means. The result for the encouragement to write in more detail was \( z = -.55, p = .672 \) and the reasonability of the length of the experiments was \( z = -.30, p = 1.00 \).

In order to understand if the ordering of the tasks in the phases had any statistically significant effect on the obtained results, a between-group ANOVA was performed. The two groups based on the ordering of the tasks are a) Phase 1: paper version, phase 2: game version and b) Phase 1: game version, phase 2: paper version. The assumption of homogeneity of variances and the normality was evaluated and determined to be satisfied before performing the test. The between-group ANOVA result for the score in detail in the paper version is \( F = 2.09, p = .192 \) and \( F = 3.61, p = .099 \) in the game version. The result for the number of themes in the paper version is \( F = .41, p = .542 \) and \( F = 1.08, p = .333 \) in the game version. The result between the two groups for the total number of themes, including the repeated themes, in the paper version and in the game version are \( F = .22, p = .652 \) and \( F = .66, p = .444 \), respectively. The result for the word count in the paper version is \( F = 4.44, p = .073 \) and \( F = .37, p = .564 \) in the game version. Finally, the percentage of paper explained is \( F = .33, p = .585 \) and the percentage of similarity of themes is \( F = .16, p = .705 \). Thus, there was no statistically significant difference between the means of the related items for the two groups and the ordering of the phases had no effect on the obtained results.

Personality themes result

We read all the stories for nine of the participants; due to a technical problem while recording the data, we lost the story data for the rest of them. We tagged the stories with the themes from the Personality Research Form. The results are shown in Table 4 in the Appendix. About 50% or more of the themes appeared in the paper-version stories also appeared in the game-version stories for six out of nine participants and in three of them, the consistency was 75% or above.
The number of unique themes was not excelling in any of the settings in general. However, the total number of themes, which includes the repeated themes, increased in the game version for five of the cases (55.55%), remained the same for one of the cases (11.11%) and declined in the rest (33.33%). The same pattern was found for the total score in details but this time paper version overtakes the game version: the stories in the paper-version was more detailed in five of the cases (55.55%), remained the same for one of the cases (11.11%), and only declined in three of cases (33.33%). Finally, the percentage of the similarity of themes in each version was 50% or below for all the cases.

**Discussion**

In the empirical study, we have not had any form of compensation or extrinsic motivation for participation in the study. Thus, we may expect that all the participants were intrinsically motivated to participate at least to some extent. In fact, their self-report on the questionnaire in regards to high rate of agreeability on taking the responsibility to accomplish the tasks well, being careful to address everything that was asked about the images, trying to make no errors and trying to describe the images in detail and honest in both settings can support their initial intrinsic motivation. Therefore, any sign of increased motivation or enjoyment among the participants in game setting can be referred as a strong indicator of success in the gamification’s goal. Based on the Bartle’s anatomy of player we tried to address the achievers and explorers in the first game setting, included the socializers in the second game setting, and tried to understand how the gamification could enhance the test experience. The results obtained from the survey showed a significant amount of enjoyment, motivation, pleasure, and preference of the game setting over the paper setting for almost all of the participants. Although the game experiment was much lengthier than the paper-version, more than half of the participants either did not find any of them lengthy or found the paper version lengthier. This supports the engagement of the participants in the game settings. These results are well explained by one of the participants as:

“The game version was more fun to play and made it more clear what I am asked to do. In the paper version I was not sure, should I write a story or explain the picture. Game version was more successful since I had to complete a scenario. Game version was a bit long though, as it took a while to get to the point where things started to happen. Having different options for dialogue was really cool.”

The high rate of agreeability on logical flow of the game, and easiness of the interface, following the story, and learning to play the game support the games’ prosperity and almost all the participants were on an agreement that the game could capture the essence of the paper version. This suggests that the games were able to operate the same test while facilitate it by providing a more engaging environment.

The enjoyment in the game version was higher for most of the participants as reported in the comparison questionnaire but we only found a statistically significant increase of enjoyment among the gamers. This might suggest that the gamification was only successful among the gamers but we only had five non-gamer participants, therefore, there might be a need of more non-gamers’ data to get any general conclusion about them.

Based on these results, we can conclude that the games were successful in its goal: providing a more engaging environment for the test; however, the games were not without drawbacks. The Mad Science platform is still in beta version and there are some problems with it. For example, the sound system does
not work properly. In the game settings, there were two different sounds to feedback the player when the mood meter increased or decreased. Due to some technical problems, the sound system did not work for some of the participants. These participants reported confusion about the mood meter’s function. This might explain the reason that only 55.55% tried to keep the mood meter high. Another problem with Mad Science was that the characters movement does not have animation: The characters slide rather than actually move. In fact, about 40% found the game’s graphics distracting and this might explain the lower rate of immersion and relate to the game characters. As one of the participants said: “The interface was messy and the inconsistent characters threw me out of it”. Even though the options in the dialogues were satisfying for most of the participants and they reported that they wanted to try the options they had, some did not find them compelling. As one of the participants said, the options in dialogues were meaningless: There was not any tradeoff in the options; rather you knew which option you needed to choose in order to proceed. This might explain the reason why only half of the participants found the story of the game to be fun. Designing a more fun and interactive game in a better environment might have led to an even better result.

As we mentioned earlier, to maintain the maximum consistency between the game settings and the control setting we also used the Mad Science platform for the control setting so that the input remains the same in all the three versions. Seven of the participants (38.89%) reported they could have a better result either above or far above average if they could write their stories in a hard paper or a word document. We did not make separate questions about the hard paper and the word document but as mentioned by some of these participants in the additional comments, this preference was due to their need for spell check. The spell checking option is also not available on the hard paper that the actual test is currently running. This might have affected their overall experience and their self-report in the surveys.

We coded the stories for some of the participants and compared the coding of the stories in the two settings for further analyzing. The coding in the game setting had a consistency of above average with the paper setting for most of the participant suggesting that the game could indeed capture the essence of the real test. Interestingly, the total number of themes appeared in the game setting, including the repeated themes, was more than the total number of themes appeared in the paper setting for more than half of the participants and was equal for one of them. The participants, who showed more themes in the game version in comparison to the paper version, either, were more motivated to write detailed stories or could better bypass the defense mechanism in the game setting. The high rate of self-report regarding “I wanted to enjoy the game” shows that most participants were consistent with this finding. This can lead to a better accurate result for the purpose of TAT. For the other group who had more themes in the paper, they either did not find the game motivating, did not understand that the objective of the game or did not take the doing the task in the game serious. For example, one of the participants said he was more interested playing the game and when he faced the story section, he was eager to see what happens next so he wrote short stories and was interested in coming back soon to the main game story. It was only at the end of the game that he realized the whole game was about those stories. Furthermore, in the second game setting, the participants were asked to address the four questions of the TAT individually rather than as a whole. This caused that most of the participants to answer shortly and almost none of them came up with a fully structured story in this game setting. This might explain the reason why the score of detail in the paper version was higher than the game-version for most of the participants. However, for the participants who played the first game scenario, the stories in the game version was more detailed and almost all of the participants came up with fully structured stories in this setting.
The low percentage of similarity between the themes in the game setting and the paper setting might be due to various reasons. First, as we discussed earlier the cards in the two settings were not the same, therefore, the cards in one setting might capture some aspect of the personality that may not be the same with a different set of cards. Especially the participants who wrote short sentences or short stories, only one or two themes could be associated with their stories; therefore, each card could only reflect a few themes for them. There might be a need for more cards for these participants in order to understand all of their personality themes. However, the themes in the two settings were more consistent for those who wrote long stories and in most cases, these participants also had more associated themes and higher detail score in the game setting than the paper setting. In the actual test in a clinical setting, psychologists use at least 10-12 cards. If we could use this number of cards in each setting, we might have had a better consistency for the two settings. Due to the limitation of using three cards to shorten the total length of experiments, the similarity of at least 40 percent, which has been appeared for all of the participants except one, might be an acceptable result.

While most of the literature in gamification claim an increase of motivation in a gamified environment, some argued the opposite. For example, Hanus and Fox [54] utilized a gamified environment for an elective course and they concluded that the motivation, satisfaction, and empowerment were less among the students who received the gamified curriculum than those who did not. They argued that “giving rewards for already interesting tasks (i.e., ones where an individual is already intrinsically motivated to do) decreases intrinsic motivation, however, incentives given for boring tasks might actually increase intrinsic motivation”. Since the course was an elective meaning that students were already intrinsically motivated about the course, adding rewards to their motivation resulted in an opposite effect. This might be also the case in our study for some of the participants and explain some of the results since all the participants were intrinsically motivated for participation; as one of the participants said, “The game version was really funny to me, but the paper version is probably the better version.”

Games are strong media which can influence its players. Klimmt et al. [52] concluded based on Goldstein’s vicarious self-perception theory that when games have a narrative in which a role is assigned to the players and they control a specific character, they will self-identify themselves with that character. In their research, by using Implicit Association Test (IAT), they showed identification occurs with video game characters. Based on their research when subjects are engaged in a racing game (Need for Speed: Carbon) they imagine themselves as if they are a driver and have the attitudes of such character, while in a shooter game (Call of Duty 2) they do so as a soldier [45]. Similarly, Domínguez et al. [53] showed when a player is given a role in interactive narrative role-playing games, implicit or explicit, they remain consistent with their roles, and their chosen actions in games are under the influence of their roles. They explicitly gave roles to one group and did not do so for their control group. Interestingly, participants in the control group were consistent with an implicit role but there were no agreements upon them. They called this phenomenon the “Mimesis Effect”. In our study, we tried to give the role of a writer and a critic to encourage players to write detailed stories about the TAT cards so that more aspects of personality can be drawn from their writings. The games were simple and not very engaging for all the participants but we found detailed stories and therefore more associated themes for those who could relate themselves to the game characters. Making games that are more engaging and can appeal a broader range of audiences, might have led to results that are more substantial. Our study is an initial step toward the era of assessment through games instead of its traditional paper-based form. We hope advancement in this field not only approach assessment through games possible but also opens access to treatment or advancement in psychological phenomena.
Limitation of study

The number of participants that we had for the empirical study was not adequate for an extensive conclusion. In addition, all the participants in this study shared similar demographics. There is a need for a wider range of demographics and a greater number of participants to ascertain the conclusion made in this study.

Concluding based on the self-report questionnaire might not lead to an accurate result especially in our case that the number of participants is few. We could not run significance testing for all of our variables and there were some inconsistencies in the self-report of the questionnaires. For example, 83.33% of the participants reported the length of the experiment was extremely or somehow reasonable both in the game version and in the paper version. However, in the comparison questionnaire, only five or 27.85% reported that they did not find any of them lengthy.

The comparisons made on this study is based on accepting the validity and reliability of the TAT, however, there are still questions about the TAT’s reliability and validity. Lilienfeld [43] has covered literature related to problems with TAT reliability and validity. As an example for one of the problems with TAT, Wildman [55] asked a group of clinical psychologists to determine, on basis of various combination of test results, whether respondents’ were psychiatric patients or non-patients (nurses) who had been matched to the patients for age and education. Although adding MMPI to the TAT resulted in an increase in accuracy from 57% to 80%, adding TAT to the MMPI resulted in a decrease in accuracy from 88% to 80%. Another critic on this test, which can also be applied to lots of other tests, is about its sample in which has been validated. Usually, tests have been validated with known subjects in equal size. Samples consist of two groups: one with X phenomena or disorder and the other group without it. In the validation studies, researchers try to find out how well a test can identify the first group compared to the other. However, there are some arguments that the test is not as useful and valid for diagnosis: to see if a person has a disorder or not [56].

Future Work

In this study, we tried to understand the effect of gamification of the TAT with a small sample group. In the future, we will expand our research with a broader range of participants and use a better version of the games to understand the degree of the consistency of our initial findings. Having more participants can provide results that are more considerable and facilitate more comparisons. For example, we could not run a between subject study to understand the difference between the two games but with more participants, this can also be done.

Summary

In this paper, we discussed affordances of game-like environments and the way such environments can be used in an assessment. We implemented the Thematic Apperception Test in an environment called Mad Science with two different game settings and compared the results with the paper-based version of that test for the same participants. Results showed a significant amount of increased enjoyment, motivation, and preference in the game setting. At the same time, the game could not only capture the essence of the traditional test but also could reveal more aspect of personality for some of the participants.
Acknowledgment

I would like to thank Professor Casper Harteveld for his guidance through all of this research. This work could never be done without his help and contribution. I would like to thank Professor Randy Colvin and William Sharp from the department of psychology at Northeastern University who helped me with the psychological materials. I would like to thank Andy Hall who customized the Mad Science platform for this research and provided all the required materials in the Mad Science platform. Finally, I would like to thank all the participants who participated in this study.

References


1998.


Appendix
Demographic and gaming preference questionnaire: [https://neu.co1.qualtrics.com/jfe/form/SV_bxObcwiOrQznuT3](https://neu.co1.qualtrics.com/jfe/form/SV_bxObcwiOrQznuT3)
Post-Game Questionnaire: [https://neu.co1.qualtrics.com/jfe/form/SV_e3uG4maVeGL0TgF](https://neu.co1.qualtrics.com/jfe/form/SV_e3uG4maVeGL0TgF)
Post-Paper Questionnaire: [https://neu.co1.qualtrics.com/jfe/form/SV_cD7uk7XiF4KsR49](https://neu.co1.qualtrics.com/jfe/form/SV_cD7uk7XiF4KsR49)
Comparison Questionnaire: [https://neu.co1.qualtrics.com/jfe/form/SV_bx9CwuSTJvk8oT3](https://neu.co1.qualtrics.com/jfe/form/SV_bx9CwuSTJvk8oT3)
Game 1 Link: [https://web.northeastern.edu/madscience/borna1.html](https://web.northeastern.edu/madscience/borna1.html)
Game 2 Link: [https://web.northeastern.edu/madscience/borna2.html](https://web.northeastern.edu/madscience/borna2.html)
Control Setting Link: [https://web.northeastern.edu/madscience/borna3.html](https://web.northeastern.edu/madscience/borna3.html)

Coding of the stories
In Table – 3, ID is an associated number to each individual, order of the tasks shows the order of the procedure for each participant, and the game setting shows the associated game to that participant.
Table – 4 shows the summary of the coding of the stories. The first column is the participant ID. The second column is the total codes that are assigned for all the three pictures in the paper version. The codes with an exclamation mark mean that the opposite of that theme was assigned to a story. For example, “! Change” means that the participant showed the opposite spectrum of change. It is notable that the codes that appeared repeatedly on different stories are only shown once. The third column shows the total number of unrepeated codes that are found in the paper version while in the fourth column, the number of repeated codes is also included in the calculation of total themes. Column fifth is the sum of scores in detail among all the three stories in the paper version. The next four columns, column sixth, seventh, eighth and, ninth, shows the same variables but in the game version. Column tenth, the percentage of paper explained, shows how much of the tags that are used in the paper version are also depicted in the game version and is calculated by the number of same tags that are in the paper version and the game version divided by the total number of tags in the paper version. This percentage can indicate to what extent the manipulation of gamification could affect the result of the actual non-gamified test. The percentage of similarity in the last column shows the similarity of the themes in both settings. It is calculated by the number of similar tags in both settings divided by the larger number of tags.

Table 4 - The summary of the procedure for each participant

<table>
<thead>
<tr>
<th>ID</th>
<th>Order of the tasks</th>
<th>Game Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>P03</td>
<td>Game – Paper</td>
<td>Game-2</td>
</tr>
<tr>
<td>P04</td>
<td>Game – Paper</td>
<td>Game-2</td>
</tr>
<tr>
<td>P07</td>
<td>Game – Paper</td>
<td>Game-2</td>
</tr>
<tr>
<td>P09</td>
<td>Paper – Game</td>
<td>Game-2</td>
</tr>
<tr>
<td>P13</td>
<td>Paper – Game</td>
<td>Game-1</td>
</tr>
<tr>
<td>P20</td>
<td>Game – Paper</td>
<td>Game-2</td>
</tr>
<tr>
<td>P22</td>
<td>Paper – Game</td>
<td>Game-1</td>
</tr>
<tr>
<td>P29</td>
<td>Game – Paper</td>
<td>Game-2</td>
</tr>
<tr>
<td>P32</td>
<td>Game – Paper</td>
<td>Game-2</td>
</tr>
</tbody>
</table>
Table 5 – Summary of the coding of the stories

<table>
<thead>
<tr>
<th>ID</th>
<th>Code In the Paper Version</th>
<th># of unique themes</th>
<th>Total # of themes</th>
<th>Detail</th>
<th>Code In The Game Version</th>
<th># of unique themes</th>
<th>Total # of themes</th>
<th>Detail</th>
<th>% of paper explained</th>
<th>% of similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>P03</td>
<td>Affiliation - Nurturance Play - Understanding</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>Achievement - Affiliation Endurance - Nurturance</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>P04</td>
<td>Aggression - Autonomy - Cognitive Structure - Dominance - Endurance – Harm avoidance Nurturance - Sentience - Understanding</td>
<td>10</td>
<td>11</td>
<td>4</td>
<td>Achievement - Affiliation Cognitive Structure - Nurturance - Sentience - Understanding</td>
<td>6</td>
<td>11</td>
<td>4</td>
<td>44.45</td>
<td>44.45</td>
</tr>
<tr>
<td>P07</td>
<td>Abasement - Achievement Play - Understanding</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>Achievement - Affiliation Endurance - Impulsivity Nurturance - Understanding</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>50</td>
<td>33.3</td>
</tr>
<tr>
<td>P09</td>
<td>Achievement - Affiliation Impulsivity - Nurturance</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>Achievement - Affiliation - Aggression - Dominance Impulsivity</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td>P13</td>
<td>Affiliation - Autonomy - Play - Social Recognition - Understanding</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>Affiliation - Achievement - Aggression - Autonomy - Dominance - Endurance – Harm avoidance - Play - Social Recognition</td>
<td>10</td>
<td>13</td>
<td>8</td>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td>P20</td>
<td>Achievement - Cognitive Structure Endurance</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>Achievement - Affiliation - Change - Cognitive Structure - Dominance - Endurance</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Achievement - Affiliation - Autonomy - Nurturance Social Recognition</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>Affiliation - Nurturance - Understanding</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>---</td>
<td>---------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>P22</td>
<td>Achievement - Affiliation - Aggression - Change - Cognitive Structure - Endurance - Nurturance</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>Achievement - Affiliation - Endurance</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>42.9</td>
<td>42.9</td>
</tr>
<tr>
<td>P29</td>
<td>Achievement – Abasement - Affiliation - Aggression Dominance - Impulsivity</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>Achievement - Aggression Impulsivity</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>