CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

Evaluating the Impact of Procedurally Generated Content on Game Immersion

Abstract. This paper describes a study that examines the impact that procedurally generated content has on the quality of gaming experience. To that end, an experimental study has been undertaken where gamers play two versions of an otherwise identical game, the only difference being that in one version the game levels are designed by a human designer and in the second version they are procedurally generated. A game immersion questionnaire is used to capture the quality of the gameplay experience and the results across the two groups compared. Whilst there are observable differences in perceived total immersion, statistical analysis using one way ANOVA testing suggests that the difference is not statistically significant. Detailed analysis of the questionnaire responses identifies where variation between the two groups is statistically significant.

Keywords. Gameplay Experience, Genetic Algorithms, Immersion, Procedural Content Generation

1 Introduction

Game content is an important factor in keeping players engaged in gaming worlds, yet games are becoming increasingly complex which has a corresponding impact on game development (Blow, 2004). As well as dealing with this complexity in terms of the underlying code, there is an increasing demand for new game content. In general, game content and game asset creation is one of the most significant costs of a game (Köhler et al., 2012). Manual content production is expensive and potentially not scalable (Iosup, 2011). In contrast to manual content production, Procedural Content Generation (PCG) is the application of computer software to generate game content, specifically the algorithmic generation of game content with limited or no human contribution (Togelius et al., 2013). However, PCG is considered difficult as it not only incurs extensive computational overhead, but also requires the ability to compute the technical and cultural values of the generated instances (Hendrikx et al., 2013).

Whilst PCG has been the subject of considerable interest, much of the research in this area is focused on the technical feasibility of the various approaches. In contrast to this, the research in this paper focuses instead in attempting to quantify how procedurally generated content is received by the game players. To that end, this study involves measuring to what extent players immerse themselves in a relatively simple game. A group of test subjects were provided a game where the level map was procedurally generated using a genetic algorithm and the results compared to a control group who played an otherwise identical game where the game levels were designed by a human game designer. The outcomes of this study are presented in the following sections, with section 2 providing a short overview of related work and section 3 describing the game implementation. The experimental design is covered in section 4 and the results of the study presented in section 5. These results are analysed in section 6, and section 7 concludes the paper.

2 Background and Related Work

The creation of game content, such as models, levels, textures, and other items within the game world, is time-consuming and costly (Edwards, 2006). In addition, it has been noted that the manual creation of game content has a range of other drawbacks that include a lack of flexibility and suggestions that manual approaches are inherently unscalable (Roden and Parberry, 2004). The automatic creation of game content is not new, with examples dating back to the 1980s (Togelius et al., 2013), however there are continuing challenges in finding ways to reduce unwanted artefacts that can be encountered using simple random generation of content. To that end, a number of approaches have been considered with a growing emphasis on the use of computational intelligence techniques such as evolutionary computation. The following sections provide an over of both Procedural Content Generation (PCG) in general, as well as specific examples of evolutionary approaches to PCG.

2.1 Procedural Content Generation

This section provides a brief introduction to existing work in the area of Procedural Content Generation. It is not an exhaustive review, as such surveys are already published in the literature (Hendrikx et al., 2013). Instead, this section introduces PCG for later purpose of identifying evolutionary approaches to PCG.

Togelius, Kastbjerg, Schedl, and Yannakakis (Togelius et al., 2011a) discuss PCG in an attempt to classify PCG through a process of defining what it is not. Implicit in this work is the distinction between online PCG and offline PCG, the former being where content is generated at runtime during gameplay and the latter being where content is generated during the design of the game or prior to gameplay (Khaled et al., 2013). Online PCG is both challenging and fascinating with the potential to impact factors such as the replayability of games as well as promoting the emergence of new game dynamics (Smith et al., 2010), however this work specifically addresses offline PCG.

Offline PCG facilitates the game development process and typically involves systems that assist game developers in their design process (Smith et al., 2010, Treanor et al., 2012) or through the creation of game assets (Whitehead, 2010) or levels (Mark and Berechet, 2014, Smelik et al., 2010). Traditional approaches to PCG utilise a number of techniques or theoretical frameworks, such as L-systems (Dormans, 2010) or other space based approaches (Bourke and Shier, 2013), statecharts (Dragert et al., 2011) and petri nets (Lee and Cho, 2011), along with an emergence of declarative approaches (Smelik et al., 2010) or those using techniques such as answer set programming (Smith and Mateas, 2011) and those that involve user direction (Kruse et al., 2016). However, one of the dominant trends in recent years is the growing interest in search based or evolutionary based approaches to PCG.

2.2 Evolutionary Procedural Content Generation

Evolutionary Procedural Content Generation (EvoPCG) is a specific case of search based PCG where the search algorithm utilises an evolutionary base. In general, search based PCG (Togelius et al., 2010b, Togelius et al., 2011b) provides the opportunity to explore massive search spaces and find unique solutions that may not be generatable using more traditional approaches. The identification of "unexpected" solutions using search algorithms has been seen in a variety of domains, such as software engineering (Harman, 2007) and engineering design (Bentley and Wakefield, 1998, Connor et al., 2000) to name but two. The application of evolutionary algorithms offer the same potential in terms of the possibility to identify surprising and novel solutions in the creation of game content.

Whilst the term search based PCG was only recently utilised (Togelius et al., 2010b), the use of evolutionary algorithms has been relatively extensive prior to this. As with PCG in general, surveys of existing search based approaches to PCG exist in the literature (Barreto et al., 2014, Togelius et al., 2011b) and this paper does not attempt to replicate this work. Instead, it focuses on a number of implementations with the specific aim of identifying current work and investigating how the outcomes are validated. In this regard, most of the work in this area undertakes a very limited validation of the work of which few, if any, consider whether the use of evolutionary algorithms is addressing the fundamental premise that PCG is valuable because manual content creation is expensive and time consuming.

For example, Cardamone et al. (2011) discuss a system that uses an interactive evolutionary algorithm for the design of racetracks for a high end racing game. In this work, they evaluate the resulting tracks using a combination of quantitative fitness function data and user perception of the generated tracks. However, it does not compare whether the results with any other approach, even randomly generated tracks. Nor does the work look at the time or computational cost, and compare this with manual creation of content. In these regard, the results are essentially verified against their goal, but there is a lack of validation as to whether the goal meets a desired need. This should not detract from the quality of the research nor from the quality of outcome the system has produced. Instead, it highlights the need for greater degrees of empirical testing and validation of PCG methods.

A review of recent applications show this to be relatively prevelant, with many authors stating the goals and purpose of PCG which is to save time and money in the production of engaging content, but not fully evaluating whether there is such a saving or whether the content is engaging. For example, Togelius et al. (2010a) consider a multiobjective approach for the generation of strategy maps, and whilst the use of pareto-optimal sets in exploring trade-offs in design objectives is unquestionably valuable, there is no indication of the computational cost of implementing this approach. Other studies often undertake some form of empirical evaluation, for example Ashlock et al. (2011) consider different representations for the same problem of creating maze-like levels and draw conclusions about the relative merits, but again they do not undertake a comparison with other approaches that may be procedural, random or manual. In general, most of the recent research in this area falls in to one of the two categories presented above. The first being where the output of the process is evaluated in terms of whether it achieved the desired design goal (Merrick et al., 2013, Shaker et al., 2013) and the second being where some comparison is made, often between different representations or different tuning of the algorithms (Liapis et al., 2015).

There are exceptions, for example Raffe et al. (2015) compare the playability of optimised levels against that of random levels. However, there is still an unanswered question as to whether the computational cost of deploying evolutionary algorithms has any benefits in terms of meeting the goals of PCG in terms of the efficient creation of engaging content.

2.3 Immersion in Games

The notion of immersion itself is not well defined; despite the wide spread use of the term in the game industry and game media. While it has been identified that it does indeed occur when playing games (Brown and Cairns, 2004, Haywood and Cairns, 2006), the specific attributes that contribute to such a state is still in the process of the being studied. In Jennett et al.'s work (Jennett et al., 2008), the term immersion has a specific definition, relating to other notions; flow, cognitive absorption and presence.

These ideas all overlap in areas, but are all related to the process of being engaged playing video games in some way. Flow describes the process by which "people become so absorbed in their activities that irrelevant thoughts and perceptions are screened out" (Jennett et al., 2008). Cognitive absorption is concerned with how people become deeply involved with software. Cognitive absorption is applicable to video games as games are a form of software. Presence describes the physiological state of being in a virtual environment. Since games typically involve simulated virtual worlds, presence may be applicable to video games. These three notions are all applicable to the measurement of video games, however Jennett et al. (Jennett et al., 2008) identifies that these notions do not cover all aspects of games and may not be an accurate measure of the success of games. Pulling from the parts that overlap and are universally applicable between these ideas, Jennett et al. go on to establish immersion as a potential measure of the quality of a video game.

3 Implementation

The purpose of this research is to evaluate the impact of using PCG in terms of producing engaging content by using the concept of immersion as a measure of the game quality. In order to achieve this objective, an accompanying game and PCG method have to both be instantiated to be used in the experiment. As this paper is focused on the player evaluation of the content, this section primarily focuses on the game environment as perceived by the player. Complete details of the PCG method and evaluation can be found in previously

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

published work (Reference blinded for review).

3.1 Game Environment

The game developed for this research is a simple 2D top down shooter, where the player must navigate a labyrinth and find the end of the level. Along the way, there are a variety of Non-Player Characters (NPCs) that engage the player in a hostile manner. The player must successfully evade and/or overcome these hostile NPCs and find the end of the level in order to proceed to the next level.

The game was designed to be simple and straight forward. Whilst this is the case, the game can be considered in many ways to be representative of more complex first person perspective games. The game, named Vektor, was not given any fiction, or narrative that informs the gameplay. For the purpose of the research, only immersion derived from the mechanics of the gameplay is relevant to the study. Any possible effect of immersion from the design of a narrative or the visual aesthetic is not desired, as it could potentially interfere with the immersion of the game and the variants. Because of this, the game can be described as an abstract game, where the use of basic shapes and symbols are used to communicate the purpose of gameplay elements, without invoking any emotional relationships between entities in the game, relative to peoples past experiences. A screenshot of the game is shown in Figure 1.

Fig. 1. Vektor screenshot.

While it may not sound like the game's design may have a tangible impact of the success of the research, the relationship between the design of a game and PCG systems is in truth important. The design of the game mechanics have to enable the design of content produced by a PCG system to influence immersion. In the interaction between the player and the game, it is conceivable that there are three things that can influence the perceived immersion. The first is the player. The immersion due to the players' previous experiences and expectations is hard to minimize. The second and third is the mechanics of the game and the content of the game, both of which have the potential to impact the extent to which a player experiences immersion. Game content includes aspects such as narrative as well as visual and audio content. Such elements have been shown to have an impact on immersion (Smith, 2016). Game mechanics refers to the rules that governs interactions within the game environment (Call et al., 2012). Gameplay, on the other hand, represents the distinct combination of mechanics that comprise a single game as experienced by the player (Kuo et al., 2017). The individual mechanics of a game may not be compelling on their own, but the sum total of a game's mechanics creates a unique gaming experience for the player when combined during play (Arsenault and Perron).

For this research, game levels have been generated through both through the use of a genetic algorithm and a human designer. Because the PCG system deals with levels, whether or not it is successful is based on the spatial configuration of the level that is produced. Hence, we need a game design where the contiguous space has a tangible effect on the immersion rating of the game. For this reason, a top down shooter game was implemented, as the mechanic of firing and avoiding projectiles necessitates the use of cover and having enough room to evade.

4 Experimental Design

This research uses a relatively simple experimental design based around the play testing of two game variants. The only difference between the two games is the method by which the level maps were generated. One variant uses maps produced by human game designers, whilst the other uses procedurally generated levels.

For this research, player immersion was chosen to be evaluated to measure the success of the game variants. In particular, an existing player immersion questionnaire (Jennett et al., 2008) was utilized. However, it was decided to not utilise the final question of the questionnaire that explicitly referenced the concept of immersion using a 1-10 scale, which differed from the other questions that are based on five point Likert scales. Therefore, measures of immersion from this work are not comparable with other studies. Given the only difference between the two games is the level design, it can be assumed that any variation in perceived immersion arises directly from this difference. Accordingly, the null hypothesis is as follows:

H0: The immersion ratings between the two game variants will not be significantly different

The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of the two groups. This approach is common in game immersion studies (Denisova and Cairns, 2015) and the one-way ANOVA is considered a robust test against the normality assumption. The possibility of bias as a result of the small sample size is noted.

5 Results

A total of 20 participants were selected for the study and randomly assigned a game variant to play. Each participant only played one version of the game. Immediately prior to play testing, demographic and game playing experience was collected using a pre-participation question. During the playtesting sessions it was noted that a number of the less experienced gamers were experiencing difficulties with the controller and navigating through the game. It was therefore decided to do a post-hoc pruning of the data to remove participant data from the analysis for those participants not playing video games at least an hour per week. The following section compares the demographics of the playtesters.

5.1 Participants

Following the post-hoc pruning, data was available for 7 players who played the human designed game levels and 9 players who had played the PCG levels. A further demographic analysis of the groups was conducted to show that the compositions of the groups were roughly comparable. Figure 2 presents data that compares the ages of the participants in each group.

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

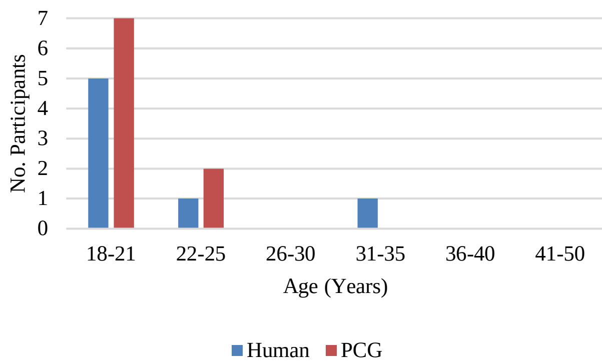


Fig. 2. Comparison of ages in groups.

The two groups of playtesters had a mix of ages, however the majority in both group were in the 18-21 age bracket. Figure 3 presents the data related to the gender of the two groups.

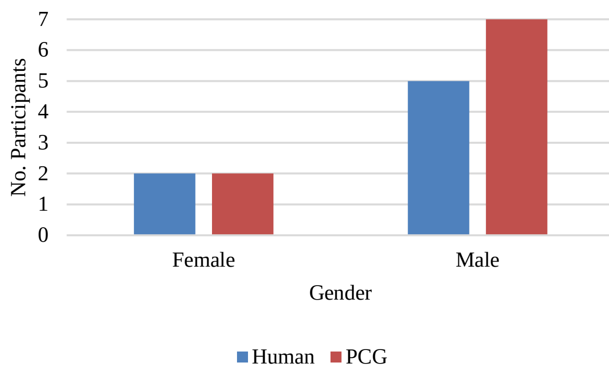


Fig. 3. Comparison of gender distribution in groups.

Again, only minor variances in the composition of the groups are noticeable as a result of the difference in sizes following the post-hoc pruning. Figure 4 presents the data related to the gaming experience of the groups.

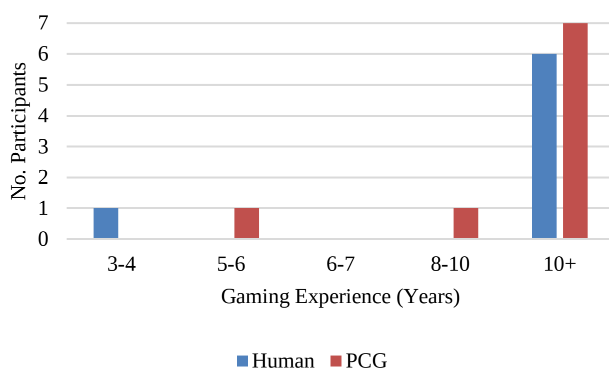


Fig. 4. Comparison of gaming experience in groups.

There is perhaps greater variability between the two groups in terms of gaming experience, however the majority in both groups have more than ten years gaming experience. Figure 5 and Figure 6 show the frequency of play and duration of play sessions respectively.

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

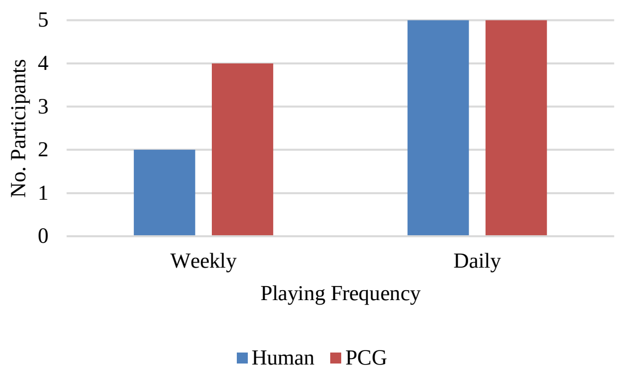


Fig. 5. Comparison of playing frequency in groups.

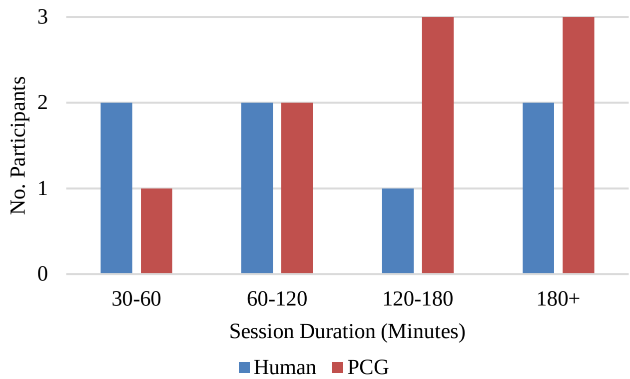


Fig. 6. Comparison of session duration in groups.

Both groups would be considered active gamers, with the majority in each group playing games on a daily basis for over an hour. By a small margin, the playtesters exposed to the procedurally generated levels might be considered more active than the other group. Whilst there are some minor variations between the two groups, they are sufficiently close to be considered a homogenous mix of active gamers, and as a result the likelihood of bias being present in the data as a result of differences between the two groups is low.

5.2 Total Immersion Analysis

Total immersion is a concept previously used by Denisova & Cairns (Denisova and Cairns, 2015) where the responses of participants in the Immersive Experience Questionnaire (IEQ) are simply summed to produce a total score for the participant. To facilitate this, the data captured using the IEQ was adjusted so that all questions resulted in a score of 5 being the best score. The results for each participant for questions 5, 7, 8, 9, 17 & 19 in the IEQ were flipped to the corresponding value if the scale was reversed (e.g. a 2 becomes a 4, a 5 becomes a 1). Figure 7 shows an analysis of total immersion for the data.

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

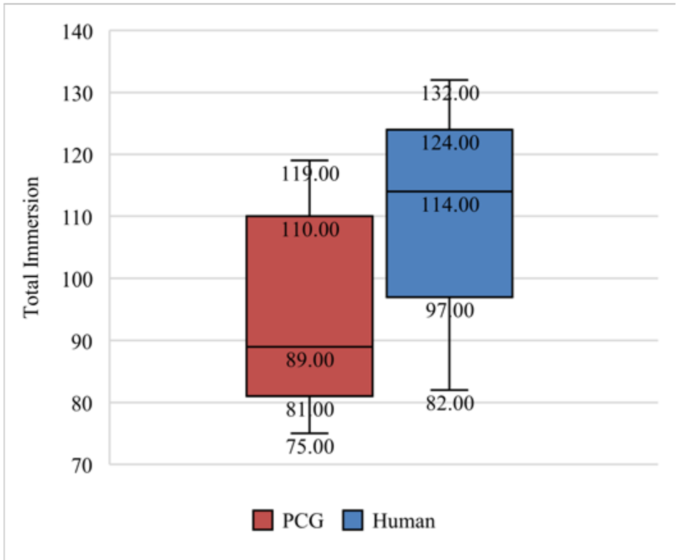


Fig. 7. Comparison of total immersion between groups.

This figure shows in general that the total immersion of players of the PCG variant is lower than that of players of the human designed variant. This is confirmed by the data relating to the plots shown in Table 1.

Table 1. Total immersion descriptive statistics

	Human Design	Procedurally Generated
Mean	112.14	94.56
Median	114.00	89.00
Range	82.00 – 132.00	75.00 – 119.00
Interquartile Range	97.00 – 124.00	81.00 – 110.00

Both the mean and median scores of the PCG group are lower than that of the group who played human designed levels. The difference in mean scores between the two groups is 18, and the difference in median scores is 25. This can be placed into context by considering the number of questions in the questionnaire, and as such a typical difference between individuals of each group might be 1 point on the 5 point Likert scale for just over half the questions. Whilst this difference is small, there is still a need to formally address the null hypothesis through the use of statistical significance testing. This can be achieved using a one sided analysis of variance (ANOVA), as proposed by Denisova & Cairns (2015). Table 2 shows the summary data generated for the ANOVA testing.

Table 2. ANOVA summary data

	Count	Sum	Average	Variance
Procedurally Generated	9	851	94.55556	249.778
Human Designed	7	785	112.1429	296.1429

Table 3 shows the ANOVA outcomes generated by using the Excel single factor ANOVA analysis.

Table 3. ANOVA summary data

Source of Variation	SS	df	MS	F	P-value

Between Groups	1217.9206	1	1217.9206	4.5167	0.0518
Within Groups	3775.0794	14	269.6485		
Total	4993.0000	15			

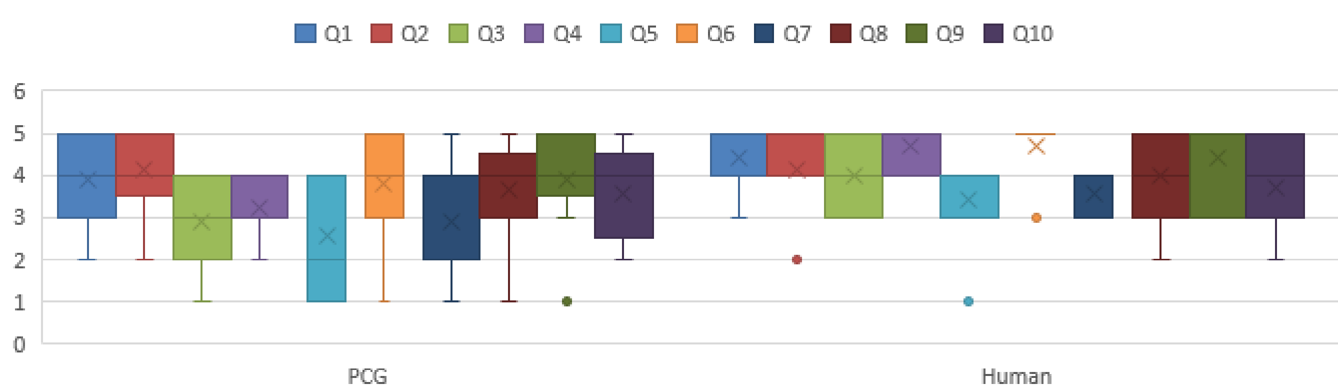
Based on an α-value of 0.05, this shows that there is no statistically significant differences between group means as determined by one-way ANOVA ($F(1,14) = 4.5167$, $p = 0.0518$). This is somewhat contrary to the data presented in Table 1 and Fig. 6 where a clear difference is visible, suggesting firstly that the actual distribution of values is not sufficiently distinct for the differences to be clear. It also suggests that the sample size may not be sufficiently large for there to be high confidence in the outcomes.

Given that the p-value is only marginally greater than the α-value, and that the F-value is smaller than F_{crit}, the outcomes of the ANOVA are definitely borderline, as might be expected from the visual interpretation of Figure 7. Whilst the one-way ANOVA is considered a robust test against the normality assumption, which implies that it tolerates violations to its normality assumption rather well, this robust only applies to skewed or kurtotic distributions. In particular, with small sample sizes, platykurtosis can have a profound effect.

To confirm the null hypothesis, a non-parametric Kruskal-Wallis test is conducted. The Kruskal-Wallis test statistic is approximately a chi-square distribution. Based on the significance level of $\alpha = 0.05$, and the number of degrees of freedom is $df = 2 - 1 = 1$, the rejection region for this Chi-Square test is R = {X2: X2 > 3.841. The H statistic ($\approx X^2$) is calculated to be $H = 4.045 > R$ (= 3.841). It is then concluded that the null hypothesis is rejected. Similarly, the p-value is $p = 0.0443$, and since $p = 0.0443 < 0.05$, it is concluded that the null hypothesis is rejected. The outcomes of the Kruskal-Wallis test contradict that of the ANOVA, however the P-value is still close to the α-value which suggests again a borderline outcome in terms of statistical significance. The slim margin may be related to the aggregation of the questionnaire results into a single measure of total immersion. Therefore the following section analyses the questionnaire responses at a higher level of granularity to identify any hidden detail.

5.3 Detailed Questionnaire Analysis

The Immersive Experience Questionnaire consists of 30 questions and the box plots in Figure 8 show an analysis of individual question responses.



CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

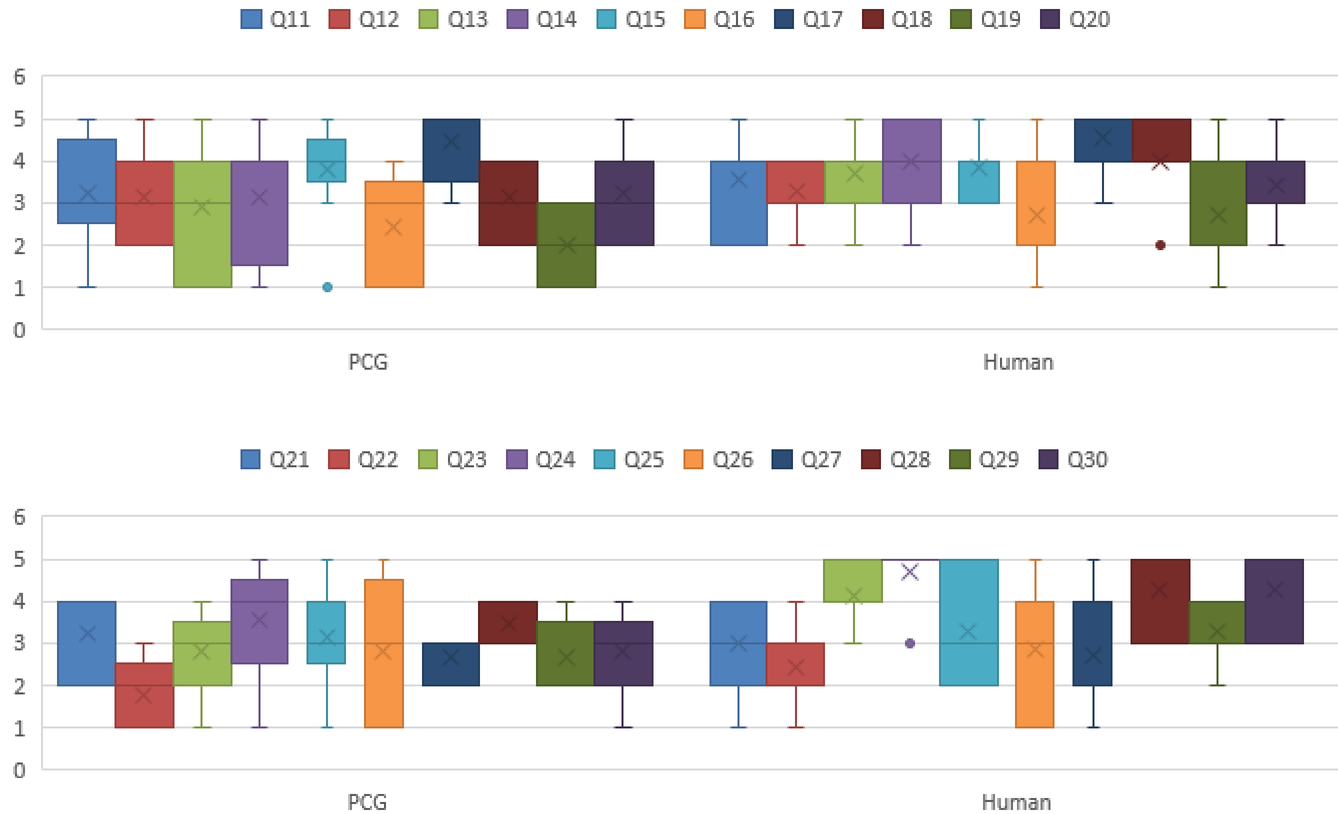


Fig. 8. Comparison of individual question responses.

The descriptive statistics of the responses to each question are presented in Table 4. In terms of the mean responses to each question, the data related to participants playing the human designed game variant is consistently higher than that for participants playing the procedurally generated levels. The one exception to this is for Q21 ("How well do you think you performed in the game?"). In terms of the median responses, a number of questions have produced data that is comparable. These include Q2, Q8, Q10, Q12, Q14, Q15, Q17, Q20, Q22, Q25 and Q26.

Of these questions, of particular interest are Q2 ("To what extent did you feel you were focused on the game?"), Q22 ("To what extent did you feel emotionally attached to the game?") and Q25 ("Were you in suspense about whether or not you would win or lose the game?") where the standard deviation of the data related to the procedurally generated level is equal to or smaller than that of the data related to the human designed variant. This suggests that there is more coherence and agreement in the responses to these questions.

Table 4. Summary of questionnaire results by question

Question	Human Designed			Procedurally Generated		
	Mean	*Median*	*SD*	*Mean*	*Median*	*SD*
1	3.89	4.00	1.10	4.43	5.00	0.73
2	4.11	4.00	0.99	4.14	4.00	0.99
3	2.89	3.00	0.99	4.00	4.00	0.76
4	3.22	3.00	0.63	4.71	5.00	0.45

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

5	2.56	2.00	1.34	3.43	4.00	1.05
6	3.78	4.00	1.23	4.71	5.00	0.70
7	2.89	2.00	1.29	3.57	4.00	0.49
8	3.67	4.00	1.15	4.00	4.00	1.07
9	3.89	4.00	1.20	4.43	5.00	0.90
10	3.56	4.00	1.07	3.71	4.00	1.03
11	3.22	3.00	1.23	3.57	4.00	1.05
12	3.11	3.00	1.10	3.43	3.00	0.49
13	2.89	3.00	1.45	3.71	4.00	0.88
14	3.11	4.00	1.37	4.00	4.00	1.07
15	3.78	4.00	1.13	3.86	4.00	0.64
16	2.44	3.00	1.17	2.71	2.00	1.28
17	4.44	5.00	0.83	4.57	5.00	0.73
18	3.11	3.00	0.87	4.00	4.00	0.93
19	2.00	2.00	0.82	2.71	2.00	1.28
20	3.22	3.00	1.03	3.43	3.00	0.90
21	3.22	4.00	0.92	3.00	3.00	1.07
22	1.78	2.00	0.79	2.43	2.00	0.90
23	2.78	3.00	0.92	4.14	4.00	0.64
24	3.56	4.00	1.26	4.71	5.00	0.70
25	3.11	3.00	1.10	3.29	3.00	1.28
26	2.78	3.00	1.55	2.86	3.00	1.36
27	2.67	3.00	0.47	2.71	2.00	1.28
28	3.44	3.00	0.50	4.29	5.00	0.88
29	2.67	2.00	0.82	3.29	3.00	0.70
30	2.78	3.00	0.92	4.29	5.00	0.88

To be able to further confirm where there are similarities between the responses of the two groups, a one-way ANOVA has been conducted on the responses for each individual question.

Table 5. ANOVA P-values for responses to individual questions

Question	ANOVA P-value	Question	ANOVA P-value
1	0.3122	16	0.6866
2	0.9535	18	0.7693
3	***0.0375***	18	0.0873
4	***0.0002***	19	0.2247
5	0.2064	20	0.7009
6	0.1137	21	0.6821
7	0.2340	22	0.1720
8	0.5887	***23***	***0.0072***
9	0.3686	24	0.0598
10	0.7832	25	0.7878
11	0.5829	26	0.9214
12	0.7370	27	0.9245
13	0.2344	***28***	***0.0404***
14	0.2071	29	0.1566

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

15	0.8789	***30***	***0.0077***

Based on an α-value of 0.05, analysis of these p-values suggests that there is a statistically significant difference between the responses from the two groups for Q3 ("How much effort did you put into playing the game?"), Q4 ("To what extent did you lose track of time?"), Q23 ("To what extent were you interested in seeing how the game's events would progress?"), Q28 ("How much would you say you enjoyed playing the game?") and Q30 ("Would you like to play the game again?").

Cross-referencing these questions to the descriptive statistics in Table 4 would suggest that the human designed game levels have achieved a higher level of immersion in the dimensions related to these questions when compared to the procedurally generated game. Whilst the median and mean for other questions are higher for the human designed game players, the spread of responses to these questions is such that the difference is not statistically significant.

6 Discussion

The results presented in the previous section allow the impact of the use of procedurally generated content on the game player experience to be evaluated. This is made possible by the two game variants being identical in terms of the game mechanics, with the only variation being whether the game levels were designed by a human designer or procedurally generated.

As a general analysis, it would seem that there is a difference in the extent to which players became immersed in the procedurally generated game. Whilst statistical analysis suggests that this not a significant difference, the relatively small sample size of playtesters is a limitation of the study. Whilst the playtesters playing the procedurally generated content were marginally more active in their gaming, it is unlikely that the observed differences in immersion were a result of players being more active. It is not a conclusive outcome, but it is probable that the lower engagement with the game was a result of the nature of the game levels. However, the cause of this lower engagement is not clear. For example, the PCG method could be inappropriate, or implemented in a fashion that limits the ability to generate game levels that create a sense of immersion. Alternatively, the final human interaction with the generated level maps could have produced unforeseen aspects of gameplay. Whilst the final interaction just involved the placement of keys and enemies, this has potential to massively enhance or detract from the gameplay experience. Further work is required to ensure that the entire game level content is generated procedurally to remove this potential bias, and similarly a sensitivity analysis of the fitness function used in the method is producing the most playable levels that it can.

This work does have a number of limitations, firstly the sample size of playtesters is relatively small which raises doubts about the generalizability of the statistical analysis. It is also a source of potential bias in the outcomes, for example whilst the participants were all active gamers they may not actively play top down shooter games. A larger population of participants would reduce this potential bias. The main contribution of this work is therefore conceptual in suggesting that there is value in considering whether procedurally generated content is engaging. The analysis framework can be redeployed in larger studies that have more value in terms of the statistical analysis.

The work is also limited in that considers only one approach to procedural content generation, and also only in the context of a relatively simple game environment. Further work will consider more sophisticated approaches to procedural generation and more complex game environments.

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

7 Conclusion

This paper has outlined a study that compares the perceived immersion resulting from playing two versions of a game, one where game levels were designed by human designers and the other where games levels were procedurally generated. Whilst the results are statistically inconclusive, there is sufficient evidence that suggests that there are particularly areas where the procedurally generated content does not succeed in engaging the players as well as human designed levels. Further work will investigate approaches to both close this gap in perceived immersion by providing more sophisticated approaches to generating content. In addition, a larger scale study will be undertaken to ensure that the statistical analysis results in more conclusive outcomes.

8 Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

ARSENAULT, D. & PERRON, B. In the frame of the magic cycle: the circle(s) of gameplay. *In:* PERRON, B. & WOLF, M. J. P. (eds.) *The Video Game Theory Reader 2.* New York: Routledge.

ASHLOCK, D., LEE, C. & MCGUINNESS, C. 2011. Search-based procedural generation of maze-like levels. *Computational Intelligence and AI in Games, IEEE Transactions on,* 3**,** 260-273.

BARRETO, N., CARDOSO, A. & ROQUE, L. Computational Creativity in Procedural Content Generation: A State of the Art Survey. Proceedings of the 2014 Conference of Science and Art of Video Games, 2014.

BENTLEY, P. J. & WAKEFIELD, J. P. 1998. Finding acceptable solutions in the pareto-optimal range using multiobjective genetic algorithms. *Soft computing in engineering design and manufacturing.* Springer.

BLOW, J. 2004. Game development: Harder than you think. *Queue,* 1**,** 28-37.

BOURKE, P. & SHIER, J. Space Filling: A new algorithm for procedural creation of game assets. Proceedings of the 5th Annual International Conference on Computer Games Multimedia & Allied Technology, 2013.

BROWN, E. & CAIRNS, P. A grounded investigation of game immersion. CHI'04 extended abstracts on Human factors in computing systems, 2004. ACM, 1297-1300.

CALL, J., WHITLOCK, K. & VOORHEES, G. 2012. From dungeons to digital denizens. *Dungeons, dragons, and digital denizens: The digital role-playing game.* New York: Continuum.

CARDAMONE, L., LOIACONO, D. & LANZI, P. L. 2011. Interactive evolution for the procedural generation of tracks in a high-end racing game. *Proceedings of the 13th annual conference on Genetic and evolutionary computation.* Dublin, Ireland: ACM.

CONNOR, A. M., CLARKSON, P. J., SHAPAR, S. & LEONARD, P. Engineering design optimization using Tabu search. Design for Excellence: Engineering Design Conference 2000, 2000. John Wiley & Sons, 371-378.

DENISOVA, A. & CAIRNS, P. The placebo effect in digital games: phantom perception of adaptive artificial intelligence. Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, 2015. ACM, 23-33.

DORMANS, J. Adventures in level design: generating missions and spaces for action adventure games. Proceedings of the 2010 workshop on procedural content generation in games, 2010. ACM, 1.

DRAGERT, C., KIENZLE, J., VANGHELUWE, H. & VERBRUGGE, C. 2011. Generating extras: Procedural AI with statecharts. Technical Report SOCS-TR-2011.1.

EDWARDS, R. 2006. The economics of game publishing. *IGN Entertainment Inc.*

<Blinded for review. Masters thesis>

HARMAN, M. The current state and future of search based software engineering. 2007 Future of Software Engineering, 2007. IEEE Computer Society, 342-357.

HAYWOOD, N. & CAIRNS, P. 2006. Engagement with an Interactive Museum Exhibit. *In:* MCEWAN, T., GULLIKSEN, J. & BENYON, D. (eds.) *People and Computers XIX — The Bigger Picture: Proceedings of HCI 2005.* London: Springer London.

HENDRIKX, M., MEIJER, S., VAN DER VELDEN, J. & IOSUP, A. 2013. Procedural content generation for games: A survey. *ACM Transactions on Multimedia Computing, Communications, and Applications,* 9**,** Article 1.

IOSUP, A. 2011. POGGI: generating puzzle instances for online games on grid infrastructures. *Concurrency and Computation: Practice and Experience,* 23**,** 158-171.

JENNETT, C., COX, A. L., CAIRNS, P., DHOPAREE, S., EPPS, A., TIJS, T. & WALTON, A. 2008. Measuring and defining the experience of immersion in games. *International journal of human-computer studies,* 66**,** 641-661.

KHALED, R., NELSON, M. J. & BARR, P. 2013. Design metaphors for procedural content generation in games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Paris, France: ACM.

KÖHLER, B., HALADJIAN, J., SIMEONOVA, B. & ISMAILOVIĆ, D. Feedback in low vs. high fidelity visuals for game prototypes. Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques, 2-9 June 2012 Zurich, Switzerland IEEE, 42-47.

KRUSE, J., SOSA, R. & CONNOR, A. 2016. Procedural urban environments for FPS games. *Proceedings of the Australasian Computer Science Week Multiconference.* Canberra, Australia: ACM.

CITATION: Connor, A.M., Greig, T.J. & Kruse, A.M. (2017) Evaluating the Impact of Procedurally Generated Content on Game Immersion. The Computer Games Journal, 6(4), 209-225

The final publication is available at Springer via http://dx.doi.org/10.1007/s40869-017-0043-6

KUO, A., HILER, J. L. & LUTZ, R. J. 2017. From Super Mario to Skyrim: A framework for the evolution of video game consumption. *Journal of Consumer Behaviour,* 16**,** 101-120.

LEE, Y. S. & CHO, S. B. 2011. Context-Aware Petri Net for Dynamic Procedural Content Generation in Role-Playing Game. *IEEE Computational Intelligence Magazine,* 6**,** 16-25.

LIAPIS, A., YANNAKAKIS, G. N. & TOGELIUS, J. 2015. Constrained novelty search: A study on game content generation. *Evolutionary computation,* 23**,** 101-129.

MARK, B. & BERECHET, T. 2014. *Procedural 3D Cave Generation.* MA thesis. IT University of Copenhagen, 2014.: http://benjaminmark.dk/Procedural_3D_Cave_Generation. pdf.

MERRICK, K. E., ISAACS, A., BARLOW, M. & GU, N. 2013. A shape grammar approach to computational creativity and procedural content generation in massively multiplayer online role playing games. *Entertainment Computing,* 4**,** 115-130.

RAFFE, W. L., ZAMBETTA, F., LI, X. & STANLEY, K. O. 2015. Integrated approach to personalized procedural map generation using evolutionary algorithms. *Computational Intelligence and AI in Games, IEEE Transactions on,* 7**,** 139-155.

RODEN, T. & PARBERRY, I. 2004. From artistry to automation: A structured methodology for procedural content creation. *Entertainment Computing–ICEC 2004.* Springer.

SHAKER, M., SARHAN, M. H., AL NAAMEH, O., SHAKER, N. & TOGELIUS, J. Automatic generation and analysis of physics-based puzzle games. Computational Intelligence in Games (CIG), 2013 IEEE Conference on, 2013. IEEE, 1-8.

SMELIK, R., TUTENEL, T., DE KRAKER, K. J. & BIDARRA, R. Integrating procedural generation and manual editing of virtual worlds. Proceedings of the 2010 Workshop on Procedural Content Generation in Games, 2010. ACM, 2.

SMITH, A. M. & MATEAS, M. 2011. Answer Set Programming for Procedural Content Generation: A Design Space Approach. *IEEE Transactions on Computational Intelligence and AI in Games,* 3**,** 187-200.

SMITH, G., WHITEHEAD, J. & MATEAS, M. Tanagra: A mixed-initiative level design tool. Proceedings of the Fifth International Conference on the Foundations of Digital Games, 2010. ACM, 209-216.

SMITH, J. 2016. *The musical parameters of immersion and flow: involving the player, emotionally and physically, in a video-game.* Masters Thesis, University of Huddersfield.

TOGELIUS, J., CHAMPANDARD, A. J., LANZI, P. L., MATEAS, M., PAIVA, A., PREUSS, M. & STANLEY, K. O. 2013. Procedural content generation: Goals, challenges and actionable steps. *In:* LUCAS, S. M., MATEAS, M., PREUSS, M., SPRONCK, P. & TOGELIUS, J. (eds.) *Artificial and Computational Intelligence in Games* Dagstuhl: Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik.

TOGELIUS, J., KASTBJERG, E., SCHEDL, D. & YANNAKAKIS, G. N. 2011a. What is procedural content generation?: Mario on the borderline. *Proceedings of the 2nd International Workshop on Procedural Content Generation in Games.* Bordeaux, France: ACM.

TOGELIUS, J., PREUSS, M. & YANNAKAKIS, G. N. Towards multiobjective procedural map generation. Proceedings of the 2010 workshop on procedural content generation in games, 2010a. ACM, 3.

TOGELIUS, J., YANNAKAKIS, G. N., STANLEY, K. O. & BROWNE, C. 2010b. Search-based procedural content generation. *Applications of Evolutionary Computation.* Springer.

TOGELIUS, J., YANNAKAKIS, G. N., STANLEY, K. O. & BROWNE, C. 2011b. Search-Based Procedural Content Generation: A Taxonomy and Survey. *IEEE Transactions on Computational Intelligence and AI in Games,* 3**,** 172-186.

TREANOR, M., BLACKFORD, B., MATEAS, M. & BOGOST, I. Game-o-matic: Generating videogames that represent ideas. Procedural Content Generation Workshop at the Foundations of Digital Games Conference, 2012.

WHITEHEAD, J. Toward proccedural decorative ornamentation in games. Proceedings of the 2010 Workshop on Procedural Content Generation in Games, 2010. ACM, 9.