Queensland University of Technology
Brisbane Australia

Klarkowski, Madison, Johnson, Daniel, Wyeth, Peta, Smith, Simon, & Phillips, Cody
(2015)
Operationalising and measuring flow in video games. In
Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction on - OzCHI '15, ACM, Melbourne, Vic, pp. 114-118.

https://doi.org/10.1145/2838739.2838826

Operationalising and Measuring Flow in Video Games

Madison Klarkowski, Daniel Johnson, Peta Wyeth, Simon Smith, Cody Phillips
Queensland University of Technology (QUT), Brisbane, Australia
{m15.clark, dm.johnson, peta.wyeth, simon.smith, c1.phillips}@qut.edu.au

ABSTRACT

This paper explores the obstacles associated with designing video game levels for the purpose of objectively measuring flow. We sought to create three video game levels capable of inducing a flow state, an overload state (low-flow), and a boredom state (low-flow). A pilot study, in which participants self-reported levels of flow after playing all three game levels, was undertaken. Unexpected results point to the challenges of operationalising flow in video game research, obstacles in experimental design for invoking flow and low-flow, concerns about flow as a construct for measuring video game enjoyment, the applicability of self-report flow scales, and the experience of flow in video game play despite substantial challenge-skill differences.

Author Keywords

Player experience; flow; self-report.

ACM Classification Keywords

K. 8. 1. Personal Computing: General- games.

INTRODUCTION

While the concepts of 'enjoyment' or 'fun' are difficult to define and quantify, the concept of 'flow' provides a useful lens through which to approach the player's experience of pleasure during gameplay (Csikszentmihalyi, 1990). Marked by characteristics including total focus, engagement, and an altered sense of time, flow is a well-established concept that lends itself intuitively to the nature of video game play (Cowley, Charles, Black, & Hickey, 2008). As such, measuring the flow experience during video game play allows for a keener understanding of both player experience and the construct of flow.

Flow is described as a gratifying experience of a task undertaken for its own sake where the challenge or demands of the task are adequately met by the skill of the person completing the task (Sweetser & Wyeth, 2005; Csikszentmihalyi, Abuhamdeh, & Nakamura, 2005). Should the person find the task either overwhelming (where challenges outstrip skill) or too simple (skill outstrips challenges), Csikszentmihalyi et al. (2005) posits that flow is unlikely to be experienced - thus, challenge-skill balance is considered a central prerequisite for the experience of flow.

OzCHI '15 , December 07 - 10 2015, Melbourne, VIC, Australia

http://dx.doi.org/10.1145/2838739.2838826.

This paper reports the level design process and pilot testing undertaken in preparation for a study of the psychophysiological characteristics of flow during video game play. To this end, it was necessary to create video game levels likely to induce discrete flow and low-flow states. To allow further understanding, we aimed to distinguish flow from two types of low-flow states; boredom and overload. Thus, we sought to create levels designed to invoke three kinds of player experience: a balance condition (in which flow was expected through challenge-skill balance), boredom (in which skills were greater than challenge), and overload (in which challenge is greater than skill). This paper is designed to articulate the challenges (and some solutions) involved in designing levels likely to create (or prevent) flow, as well as subjectively measuring flow during video game play.

FLOW MANIPULATION

Manipulated challenge-skill levels have been employed previously in video games research. Nacke and Lindley (2008) used three game conditions to induce immersion, flow, and boredom in a first-person shooter video game. Of relevance to this research are the modifications made to the flow and boredom condition. Flow was designed for through the inclusion of challenges focused on "interesting combat mechanics", gradual increase in difficulty, and "cooldown spots" (areas that allow players to rest and restock with sparse ammo and health supplies). The boredom condition focused on reduced challenge (weaker enemies and high amounts of health and ammo supplies) and a 'boring' play environment (repeating textures, linearity, and a limited choice of weapons). Nacke and Lindley's boredom condition was rated by participants to be low on challenge, immersion, and flow. Conversely, the flow condition scored highest on flow, challenge, and tension. The flow condition also elicited greater physiological arousal and greater positive valence than did the boredom condition.

Keller and Bless (2008) achieved flow and low-flow conditions through direct manipulation of challenge-skill compatibility in a Tetris-type video game. Flow was designed for in a condition referred to as 'adaptive', in which task demands automatically adapted to player skill through dynamic response to performance metrics. Low flow states were achieved via boredom and overload conditions. In the boredom condition, blocks fell at a very slow speed regardless of player ability. In the overload condition blocks fell at a fast base speed, which would continually increase during play. Keller and Bless' manipulation was successful in achieving greater experience of flow than the low-flow conditions, as indicated by participant reports of several core flow characteristics. Participants playing the adaptive condition experienced an altered perception of time, perceiving the adaptive condition to be shorter than the boredom and overload conditions. Participants also reported greater enjoyment and involvement in the adaptive condition than in the boredom and overload condition. This study identified challenge-skill balance as a significant predictor of intrinsic motivation.

ARTEFACT DESIGN PROCESS

The requirements for the video game selected for use in the current study were as follows: (i) representative of standard commercial video games and likely to induce flow, (ii) 'moddable' – alterable to the extent of being able to develop both flow and low-flow video game levels, (iii) allows for uninterrupted eight minute play sessions (to ensure adequate time for a tonal psychophysiological experience to emerge when using the game in future studies), (iv) allows players to 'jump in' – no prior investment in the game required, and (v) intuitive enough to be enjoyed by participants of all skill and experience levels after exposure to a short tutorial.

On the basis of these requirements, the video game chosen was Valve Corporation's *Left 4 Dead 2*. The game was also chosen due to its native inclusion of Dynamic Difficulty Adjustment (DDA) in the form of an entity known as the 'AI Director'. DDA provides real- time adjustment of the game's difficulty in response to player status to ensure challenge-skill balance and has proven successful in invoking flow in prior studies (Keller & Bless, 2008).

First Artefact Design

Initial artefact design followed the approach undertaken by Nacke and Lindley (2008); that is, the boredom condition incorporated a linear level with weak opponents and repeating textures, no real winning condition, a limited choice of weapons, a high amount of health and ammo supplies, and a lack of surprises. To achieve repeating textures and models within a linear level, the map took the form of a simple corridor without environmental clutter. Character conversations (friendly AI engaging in chatter with one another that reveal story) were removed. Other changes to facilitate boredom focus on manipulation of game challenge, and include (i) enemy AI altered to react slowly to player presence, (ii) notably diminished enemy health, (iii) player health unable to drop below 90%, (iv) high amount of ammo and health pack supplies, (v) only one available gun, and (vi) no 'winning condition' – the corridor does not end. The boredom condition also removed 'special infected': a prominent challenging enemy type in *Left 4 Dead 2*.

The balance (flow) condition required very little modification to the original game, but the game was altered to remove what early pre-testing revealed to be luck-based game changers. These were two 'boss' type zombies, as their chance of spawn and spawn location were randomised and could easily invoke an early loss condition for novice players. As such, the majority of level design manipulation occurred in the boredom condition to prevent flow through challenge-skill imbalance. The map level 'The Parish', a winding maze- like cemetery, was chosen due to its high average completion time, under the assumption that no players would be able to complete the level before the allotted play session time expired.

First Artefact Design Flaws

Both conditions were rejected after initial pilot testing suggested flow was unnecessarily confounded with aesthetic quality, and that the desired playtimes were not being achieved. While challenge-skill imbalance was achieved through direct manipulation of combat difficulty, aesthetic experience was compromised by the removal of environmental assets in the boredom condition. As it would be difficult to separate the impact of reduced flow from the impact of reduced aesthetic quality, this condition was discarded in favour of developing a boredom condition derived only from challenge-skill imbalance (low-flow), with all environmental assets and character chatter intact. The balance condition revealed errors in map selection, as expert players were able to complete the map in less than the required playtime and novice players easily got lost.

Second Artefact Design

The second attempt at artefact design followed the framework established by Nacke and Lindley (2008) only in regards to the manipulation of challenge-skill balance through combat difficulty; Keller and Bless' (2008) singular focus on challenge-skill compatibility largely provided the framework for this design phase. Assuming challenge-skill balance to be a central antecedent of flow, all conditions directly manipulated this balance. To avoid confounds associated with aesthetic differences, or potential differences in experiences if exposed to different environments, all play conditions now took place within the same map. The map selected was the 'The Port', in which players are required to fetch sixteen objects scattered throughout the map. Pilot testing revealed that players were unable to complete the condition in less than eight minutes for any condition, thus ensuring that no participant encountered a 'win' condition.

Both the balance and boredom conditions remained identical to the first artefact design, with the exception of the re-introduction of aesthetic elements to boredom in the form of a new map ('The Port') for both conditions - meaning the boredom condition now featured the same environmental assets, textures, and character conversations as the balance condition. The boredom condition continued to use all challenge manipulations detailed within 'First Artefact Design'. The simple level design of 'The Port' made navigation considerably easier. As with the first artefact design, the majority of level manipulation was undertaken to prevent flow in the low-flow conditions; due to the presence of DDA, it is expected that the game is capable of inducing flow without modification (in the balance condition).

In the second artefact design process, the overload condition was also added. The inclusion of this 'overload' condition enables a complete picture: the measurement of skill > challenge (boredom), matched challenge-skill (balance), and challenge > skill (overload/anxiety). The overload condition was achieved through hyper-reactive enemy AI, extra enemy health, increased enemy count, extra damage from enemy hits and friendly fire, friendly AI vision range reduced (less reactive to enemies) and larger enemy 'mobs'.

METHOD

Twenty university students (nineteen male), aged 17 - 31 (mean age of 20.2, *SD* = 3.24), volunteered for the study. Participants rated their experience with video games on a custom 7-point Likert scale, with '7' representing 'extremely experienced'. Participants self-rated as an average of 6 (*SD* = 1.10), indicating that participants were generally experienced video game players. Compensation was a game software key. Each participant played the three conditions (boredom-balance-overload) in individual experiment sessions not exceeding sixty minutes. All participants first played a four-minute tutorial to familiarise themselves with the game and control scheme to minimise potential learning effects. Ten participants then played the boredom condition, and ten participants played the balance condition. In all sessions, the overload condition was played last. This was to ensure any carry over effects (i.e. mood, frustration) evoked by the overload condition did not interfere with the experiences of boredom and flow. Each condition self-terminated after 10 minutes of play.

After each condition, participants spent five to ten minutes answering self-report surveys addressing flow and mood. The survey results investigated in this paper are from the Long Flow State Scale (FSS-2) (Jackson, Eklund & Martin, 1995). The Long Flow State Scale is a 36 item survey measured on a 7-point Likert scale, with '1' representing 'strongly disagree' and '7' representing 'strongly agree'. The scale consists of subscales (4 items each) devoted to each of the eight components of flow – skill, concentration, clear goals, unambiguous feedback, action-awareness, sense of control, loss of self-consciousness and transformation of time – as well as a subscale devoted to measuring the autotelic experience associated with flow. Scores are calculated for each of the 9 subscales as well as total flow (the 9 subscales combined). The FSS-2 is a validated measure for evaluating the experience of flow in various settings, and has previously been successfully applied to video games (Nacke & Lindley, 2008).

RESULTS AND DISCUSSION

A within-subjects MANOVA was conducted using gameplay condition (boredom, balance, overload) as the independent variable and all outcome measures (the 9 subscales as well as total flow) as dependent variables. All statistical assumptions of MANOVA were met with the exception of univariate outliers identified on the challenge-skill balance and transformation of time subscales from a single participant (Field, 2013). No substantive differences in results were found with outliers removed and in the interest of statistical power the results reported here include all cases. Using Wilk's Lambda, there was a statistically significant effect of condition on the combined dependent variables (Λ=0.259, $F(18,60) = 3.221$, $p < 0.005$; partial $\eta2$=.491). Univariate follow-ups revealed differences in terms of challenge-skill balance ($F(2,38) = 13.744$, $p < .005$), merging of action and awareness ($F(2,38) = 13.744$, $p = .007$), and sense of control ($F(2,38) = 4.552$, $p = .017$). Total flow was also found to differ significantly between conditions ($F(2,38) = 4.867$, $p = .013$).

Pairwise follow up tests using Bonferroni corrections were conducted on these three subscales and total flow. Challenge-Skill balance was significantly higher for the boredom condition than for the overload condition (p=.008); it was also significantly higher for the balance condition than for the overload condition (p<.005). The boredom condition also scored significantly higher than the overload condition in the merging of action and awareness (p=.005) and sense of control subscales (p=.046). Finally, with respect to total flow, no difference between the boredom condition and the overload condition was identified (p=.097), but the balance condition was significantly higher than the overload condition (p=.017). For all results see Table 1 and Figure 1. Overall, the results show that greater total flow was experienced in the boredom condition than in the overload condition. This difference seems to be a function of participants reporting relatively high levels of challenge-skill balance, merging of action-awareness, and sense of control in the boredom condition. Additionally, no significant differences were found between the boredom condition and the balance condition.

Subscale	Boredom		Flow		Overload	
	Mean	SD	Mean	SD	Mean	SD
Challenge-Skill Bal.	3.58	0.91	3.83	0.66	2.55	0.82
Merging Act-Aware.	4.15	0.58	3.99	0.48	3.63	0.69
Clear Goals	4.47	0.45	4.35	0.49	4.22	0.66
Unambig. Feedback	4.38	0.57	4.03	0.50	3.97	0.84
Conc. On Task.	4.21	0.81	4.41	0.52	4.32	0.60
Sense of Control	4.31	0.62	3.93	0.77	3.58	0.94
Loss Self-Consc.	4.05	0.63	4.07	0.91	4.23	0.74
Trans. Of Time	3.75	0.93	3.80	0.86	3.67	1.01
Autotelic Exp.	3.63	0.92	3.72	0.72	3.49	0.91
Total Flow	36.56	4.11	36.17	2.68	33.70	4.08

Table 1. Means and Standard Deviations for FSS-2.

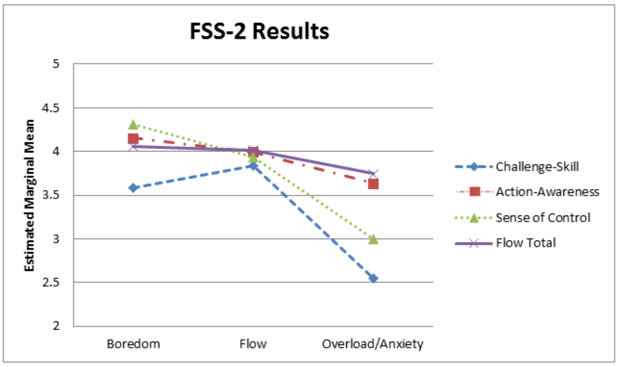

Figure 1. Self-Report Flow Results.

Difficulties in Reducing Flow in Immersive Games

A proposed explanation for the absence of differences in flow experiences between the boredom and balance condition is the potential for flow experience regardless of challenge-skill imbalance in immersive video games. This suggests that a challenge-skill imbalance does not necessarily negate flow in immersive video games. As *Left 4 Dead 2* features a highly detailed environment, it is possible that some participants may have derived aspects of flow (such as altered perception of time, focused concentration, and loss of self- consciousness) from world exploration. In other words, regardless of very low levels of challenge offered by the enemies in the game, participants were able to achieve flow through exploring and/or enjoying the aesthetic qualities of the game world.

It is possible that the challenge > skill imbalance achieved by the overload condition was successful in preventing flow because it did not allow players the opportunity to 'immerse' themselves within the world. This may be due to repeated player deaths, limited mobility (constrained by enemies), and reduced chance for exploration. As many commercial titles feature detailed environments, this may point to issues with the applicability of the flow construct to video games**.**

Unsuccessful Condition Design

The absence of any discernible difference in the flow experience between boredom and balance conditions may also point to errors in condition design. It may be that participants were still sufficiently challenged by the demands of the boredom condition. However, this seems unlikely given the range of constraints introduced to ensure very low levels of challenge. Specifically, player health never fell below 90%, enemies would die from a single shot, no 'special infected' would spawn, and a low enemy spawn rate was used (less than one third of that seen in the flow condition). Regardless, it may be that participants found unexpected ways to seek challenge in the boredom condition. For example, they may have felt challenged as a function of not being sure when enemies would appear (the anticipation of enemies in some ways balancing the relative lack of enemies), or perhaps challenged by obtaining as many fuel canisters as possible before the conclusion of the play time in the condition (confirmed anecdotally by two participants to the experimenter).

Challenge-Skill as an Antecedent

Csikszentmihalyi et al. (2005) identify challenge-skill balance as an antecedent of flow. However, Fong, Zaleski and Leach (2014) note that it may be that this balance between challenge and skill is commonly associated with, but not always necessary for, flow to occur. As such, it may be that the experience of flow is possible despite challenge-skill imbalance; in this case, sole focus on challenge-skill balance may not have been adequate for successful flow manipulation. Rheinberg, Vollmeyer and Engeser (2003) propose that flow is divisible into two factors: 'absorption by activity' and 'fluency of performance'. In this approach, flow through absorption is often associated with balanced or slightly challenging activities, whereas flow through fluency is stronger under low challenge activities. It is possible that participants experienced flow through absorption in the balance condition, and flow through fluency in the boredom condition. Further investigation into dissociation between types of flow is recommended.

Scale Applicability

In contrasting a flow experience with a boredom experience, the study has raised the question of the applicability of the FSS-2 scale to some video game experiences. The FSS-2 is a commercial scale, and specific scale items cannot be published; as such, the subscales as a whole will be discussed. The 'merging of action and awareness' subscale contains items focused on performing actions automatically and without much thought. The 'sense of control' subscale contains items focused on feeling control over what one is doing. While these experiences are true of the flow experience, they are also arguably true of a boring or unchallenging experience. As the boredom condition was not mentally taxing, it follows that participants did not need to put 'too much thought' into their actions; similarly, as the game is not mechanically challenging and was selected for its intuitiveness, it is likely that participants generally felt particularly confident in their control of the boredom condition. This aligns with Keller and Bless's (2008) findings, in which the highest levels of perceived control were reported for the boredom condition. It may be that video games that are otherwise not likely to induce flow still offer participants the opportunity for high levels of sense of control and merging of action awareness. In this way, the FSS-2 may indicate high levels of flow in video games with these features.

It is particularly notable that no significant difference was found between the balance and boredom conditions for the 'challenge-skill balance subscale. A possible explanation stems from the subscales including items that could be interpreted as asking if the respondent has sufficient skills to meet the presented challenge. Participants could have sensed that their skills were enough (or more than enough) to meet the challenges of the boredom level leading to high scores on this subscale. This would corroborate Rheinberg et al.'s (2003) description of flow through fluency. The results from the current study raise the possibility that the FSS-2, when applied to boring or exceptionally easy video game scenarios, may result in high ratings of flow when flow may not actually be occurring; however, future research (that rules out some of the other explanations for the pattern of results in the current study) is needed before any firm conclusions about the applicability of the FSS-2 to video game research can be made.

CONCLUSION

The development and evaluation of flow and low-flow experimental video game levels has uncovered difficulties around operationalising and measuring flow in games research and raises questions regarding the role of challenge-skill balance in flow. The current study highlights the difficulties inherent in manipulating flow. As aesthetically 'immersive' experiences are common attributes of commercial games, it is recommended that future research investigating challenge-skill balance be aware of this as a potential confound with flow. With respect to the boredom condition used in the current study, more research is needed in order to determine whether people experienced flow regardless of the low level of challenge experienced in the level or whether higher levels of challenge were created by participants in unexpected ways. Finally, further investigation is recommended to determine the applicability of Jackson et al.'s FSS-2 scales to video games.

ACKNOWLEDGEMENTS

We thank Valve Corporation for supporting this research by kindly providing game keys for participant compensation.

REFERENCES

Cowley, B., Charles, D., Black, M. and Hickey, R. Toward an understanding of flow in video games. Computers in Entertainment 6, 2 (2008), 1-27.

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

Csikszentmihalyi, M., Abuhamdeh, S. and Nakamura, J. Flow. In Handbook of Competence and Motivation, Andrew Elliot and Carol Dweck (eds.). New York: The Guilford Press, (2005).

Field, A. Discovering Statistics Using IBM SPSS Statistics (4th. ed.). SAGE Publications Ltd, (2013).

Fong, C., Zaleski, D. and Leach, J. The challenge-skill balance and antecedents of flow: a meta-analytic investigation. Journal of Positive Psychology 10, 5 (2014), 425-446.

Jackson, S. A., Eklund, R. C. and Martin, A. J. The Flow Scales Instrument and Scoring Guide. Mind Garden, Inc (1995).

Keller, J. and Bless, H. Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation. Journal of Personality and Social Psychology Bulletin 34, 2 (2008), 196- 209.

Nacke, L. and Lindley, C. A. Flow and immersion in first-person shooters: Measuring the player’s gameplay experience. In Proc. Futureplay ’08, ACM (2008), 81-88.

Rheinberg, F., Vollmeyer, R. and Engeser, S. Die Erfassung des Flow-Erlebens [The assessment of flow experience]. In Diagnostik von Selbstkonzept, Lernmotivation und Selbstregulation [Diagnosis of Motivation and Self-Concept], Stiensmeier-Pelster, J. and Rheinberg, F. (eds.). Hogrefe-Verlag, Germany (2003), 261-279.

Sweetser, P. and Wyeth, P. GameFlow: a model for evaluating player enjoyment in games. Computers in Entertainment 3, 3 (2005), 1-24.