USING FLOW THEORY TO DESIGN VIDEO GAMES AS EXPERIMENTAL STIMULI

David Sharek, Psychology Department
Eric Wiebe, Department of Math, Science, and Technology Education
North Carolina State University, Raleigh, NC 27695-7650 USA

The goal of this study was to evaluate the use of Flow Theory (Csikszentmihalyi, 1990) in the development of a video game for specific use as an experimental stimuli. Flow Theory was used to manipulate the level of challenge and indirectly the level of perceived skill to create three design conditions: Boredom, Flow, and Frustration. Results showed that Flow Theory provide a strong theoretical framework for manipulating skill and challenge. The intrinsic characteristics of the game mechanics provided robust, real-time performance measures that were used in a manipulation check to ensure that the conditions that were intended to be designed were indeed designed. These performance measures also provide useful data that can be combined with self-report data to produce high measurement diagnosticity and sensitivity. Validated conditions of Boredom, Flow, and Frustration can be used in studies of training and decision-making.

INTRODUCTION

The goal of this study was to evaluate the use of Flow Theory (Csikszentmihalyi, 1990) in the development of a video game for specific use as a stimulus for several experiments focused on measuring user engagement. Though the gaming industry has leveraged the high face-validity of Flow Theory to design games for many years, the use of empirical research to assess the internal validity of Flow Theory in game design has only just begun to gain momentum (Pavlas, Heyne, Bedwell, Lazzara, & Salas, 2010).

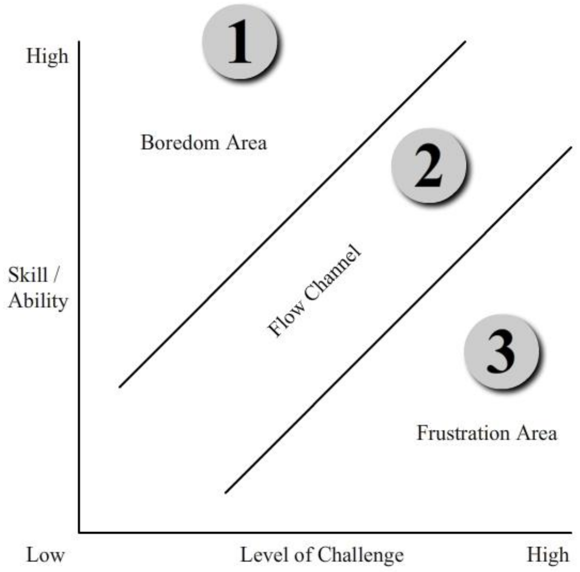
Flow Theory was used to create three design conditions (Boredom, Flow, and Frustration) by manipulating the potential for skill development and the level of challenge required to play the game. The requirements for the video game were that it must include parameters that could be manipulated to create each of the design conditions. Additionally, the video game must provide data that can be used to determine if the three conditions were created correctly.

Flow Theory's assessment of optimal experiences (Csikszentmihalyi, 1990) provides insight into understanding the interactions between user skill and task challenge. By designing video games that help users maintain a state of flow through a balance of user skill and task challenge, a state of game play may be achieved that is neither frustrating nor boring and can lead to higher levels of engagement (O'Brien & Toms, 2008; Rabin, 2005).

Flow is described as an optimal experience that includes feelings of exhilaration and deep enjoyment (Csikszentmihalyi, 1990) where being in an optimal experience is similar to being fully engaged. Csikszentmihalyi proposed that one of the most powerful experiences in flow occurs when a person is faced with difficult obstacles that they deem to be worthwhile to overcome. For example, including a scoring system in video games can help a player stay in their flow by providing feedback on uncertain outcomes such as whether or not they will receive the highest score as they progress though the game. Malone (1980) refers to this as a *metagoal* and explains that though the main goal for the game may be to finish a level, the inclusion of a scoring system can also motivate players to score as high as possible thus creating more of an interest in the game.

Figure 1 shows the hypothesized Flow Channel (2). A person can fall into the Frustration Area (3) if their skill level is not matched up with a comparable difficulty level while playing a game or participating in a task. People usually begin a new task with a low set of skills and their task should match their skill level with an appropriately low set of challenges. According to Flow Theory, as a person progresses through a task, their flow will more likely be maintained if their task difficulty manageably increases to match their growing skill. Boredom (1) can result if task difficulty is never increased to match the user's growing skill through experience with the game.

Figure 1. Three Conditions based on Flow Theory - Adopted from Csikszentmihalyi (1990)



Copyright 2011 by Human Factors and Ergonomics Society, Inc. All rights reserved DOI 10.1177/1071181311551316

In a similar vein to Flow Theory, Loewenstein (1994) provides a model of curiosity which is based on the idea that people have the desire to fill the manageable gaps that they identify in their existing declarative or procedural knowledge structure. Gaps that are too great can stimulate learned helplessness (Frustration Area) and gaps that are too little can cause apathy (Boredom Area) (Loewenstein, 1994). This has implications for video game design since it highlights that it is not only the balance between challenge and acquired skill that determines a person's placement in the flow channel; the intrinsic desire to simply gain knowledge can also motivate people to continue participating in a task.

Using Flow Theory was not only useful for designing the three distinct experimental conditions; it also introduced a practical approach for developing a manipulation check to ensure that what was intended to be designed, was actually designed.

The usefulness of developing a game such as the one herein described, is that it allows the experimenter to collect performance data that can be compared to post-hoc questionnaire and self-report data and produce a broader and more accurate picture of a participant's experience. For example, a person experiencing boredom may report different amounts of cognitive load compared to someone in the Flow condition. Additionally, using a diagnostic self-report measurement scale such as the NASA-TLX, a participant's cognitive load may load differently on each of the subscales compared to someone in the Flow condition with a similar overall cognitive load score.

METHOD

Participants

A total of 169 people were recruited through Amazon's Mechanical Turk (Amazon.com, 2009). Thirteen participants indicated that they had trouble loading some of the game levels (most likely due to poor Internet connections) during the experiment and so were removed from the analysis. Of the remaining 156 participants, 58% were female and 42% were male (mean age = 30.79, *SD* = 10.22). Seventy-eight percent were from the United States. Condition 1 (Boredom) had 48 participants, Condition 2 (Flow) had 53 participants, and Condition 3 (Frustration) had 55 participants. Thirty-three percent indicated that they had previous experience with a similar type of game.

Apparatus

The experiment was hosted on an experimental psychology Web site maintained by the author. The Web site used a MySQL database and a PHP server-side scripting language to provide the storage of participant responses. The experiment was accessible from any computer with Internet access. The computer requirements included a minimum computer monitor resolution of 1024x768 and an installed version of the Adobe Flash player v.10 or greater.

Participants played an online Flash-based strategy game called Block Walk (Sharek, 2009a). The game user experience is based on Bloxorz, a game developed by Damien Clarke (Clarke, 2007).

Figure 2. Screenshot of game play in the video game, Block Walk (Sharek, 2009a)

As **Figure 2** shows, the goal of the isometric tile-based game is to move a rectangular block, made up of two differently colored cubes, towards a goal point so that it is standing up on top of the goal. In more difficult levels, the goal will only accept the end of the block that is of the same color as the goal. The movement of the block depends on the starting position of the block on each tile. If the block is standing up on a tile, it will fall down and occupy two tiles in the direction that the user moves it; thus successive movements in this direction will increment the block by two tiles. If the block is on its side, it can either be tilted upward for a two-tile move or it can be rolled sideways for a one-tile move. This makes it difficult to predict the block's future position a few moves from its starting point. There are certain sequences of moves that can be learned to position the block more accurately; learning these sequences are usually only possible through experimentation and practice over time. The number of moves required to position the block over the goal and the complexity of the moves were manipulated to create varying levels of difficulty.

Participants in the *constant low difficulty* (Boredom) condition were introduced to the game by playing easy game levels where only a few simple moves were required to position the block over the goal. As players progressed through levels, the game difficulty did not increase.

Participants in the *skill matched with difficulty* (Flow) condition were introduced to the game through easy levels similar to those in the Boredom condition. As each level was solved, the subsequent levels increased in difficulty due to an increase in complexity and number of required moves. As players progress through the easier levels it was theorized based on Flow Theory, that they have, to some degree, mastered the game mechanics and are ready for more difficult levels where strategic thinking becomes increasingly critical.

Participants in the *constant high difficulty* (Frustration) condition were presented with game levels where many combinations of the two types of moves were required to correctly position the block over the goal. The number of moves could easily reach into the hundreds. The difficulty level was also enhanced by requiring a specifically colored end of the block to connect with the goal. In many cases the level may seem impossible to solve. The participants in this condition were not given an opportunity to learn the idiosyncrasies of the block's movements by scaffolding through simpler levels.

Design

Three flow-state conditions (see **Figure 1**) were derived using Flow Theory (Csikszentmihalyi, 1990):

- **Design Condition 1 (Boredom: Constant Low Difficulty)** – Video game will begin and stay at a low difficulty level. According to Flow Theory, the user will quickly become bored and apathetic while playing because their skill will quickly accommodate and exceed the game's difficulty.
- **Design Condition 2 (Flow: Skill Matched with Difficulty)** – Video game will begin at a low difficulty level and incrementally become more difficult as the user progresses through the game. According to Flow Theory, the user will be able to gain the skills necessary to accommodate the game's difficulty level. By creating an experience where skill and difficulty are managed and appropriately matched, the user should remain in a flow channel.
- **Design Condition 3 (Frustration: Constant High Difficulty)** – Video game will begin and stay at a high difficulty level. According to Flow Theory, the user will quickly become frustrated while playing because their skill will not be able to reach the game's level of difficulty due to the lack of incremental scaffolding.

Dependent Variables. The number of times a participant changed the block's direction for each level (**Directions***)*, the number of times the block was moved off of the game board and into the water (**Errors**), the number of times the block was moved (**Moves**), and the total length of time participants spent playing each level (**Time per Level**) was analyzed as individual performance measures. It was expected that a high-performing user would minimize all of these variables while playing the game. **Cognitive Load** was measured using the NASA Task Load indeX (NASA-TLX) (Hart & Staveland, 1988).

Procedure

The experiment was conducted over an Internet connection. Once participants provided consent for the terms of the task, they were randomly assigned to one of three conditions. They then proceeded to the general game instructions that identified the game's goal and key game mechanics. The experiment began after the instructions were read.

After completing the game, participants were asked to complete an online version of the NASA-TLX (Sharek, 2009b). When the questionnaire was completed, participants were taken to a final screen where they were debriefed and thanked. On this page, participants were given an experimental completion code which they used to paste into their Mechanical Turk user page to indicate that they have completed the experiment and required payment in the amount of 80 cents.

RESULTS

Performance and self-report cognitive workload measures for the NASA-TLX are presented in this section. An alpha level of .05 was used for all analyses described below.

A manipulation check was conducted to determine if the three design conditions (Boredom, Flow, and Frustration) accurately reflected their corresponding flow-states. Descriptive statistics on behavioral and performance data for all three design conditions can be found in **Table 1**.

Table 1.Individual Game Level M and SD by Condition

Condition		*M*	*SD*
Boredom	Directions	5.14	1.42
	Errors	.31	.21
	Moves	13.07	3.14
	Time per Level	18.62	10.28
Flow	Directions	19.98	7.50
	Errors	1.70	1.55
	Moves	39.70	13.70
	Time per Level	47.93	17.93
Frustration	Directions	40.94	24.46
	Errors	4.50	4.08
	Moves	95.54	57.68
	Time per Level	112.26	51.48

Note: (n=155)

Results from five one-way analysis of variances (ANOVA) indicated that there were significant main effects between all three design conditions and:

- *Directions* - the number of times a participant changed the block's direction for each level, F(2,155) = 72.51, p < .001,
- *Errors* - the number of times the block was moved off of the game board and into the water, F(2,155) = 35.49, p < .001,
- *Moves* - the number of times the block was moved, F(2,155) = 74.50, p < .001,
- *Time per Level* - the total length of time participants spent playing each level, F(2,155) = 111.30, p < .001.

Table 2. Design Condition Post-hoc Mean Differences

IV	Groups	Boredom	Flow
Directions	Boredom	--	--
	Flow	14.84***	--
	Frustration	35.80***	20.96***
Errors	Boredom	--	--
	Flow	1.39***	--
	Frustration	4.20***	2.81***
Moves	Boredom	--	--
	Flow	26.63***	--
	Frustration	82.46***	55.83***
Time per Level	Boredom	--	--
	Flow	29.31***	--
	Frustration	93.64***	64.33***

Note: *n=155, ***p*<.001

Table 2 shows the results from a Games-Howell post-hoc test where significant interactions were found for all conditions and IVs. Participants in the boredom condition (*M*=5.14) changed the block's direction the least number of time per level followed by those in the flow condition (*M*=19.98), and those in the frustration condition (*M*=40.94) changed the block's direction the greatest number of times per level. Participants in the boredom condition (*M*=.31) made the least number of errors followed by those in the flow condition (*M*=1.70), and then those in the frustration condition (*M*=4.50). Participants in the boredom condition (*M*=13.07) moved the block the least number of times per level followed by those in the flow condition (*M*=39.70), and then those in the frustration condition (*M*=95.54). Finally, participants in the boredom condition (*M*=18.62) spent the least amount of time per level (measured in seconds) followed by those in the flow condition (*M*=47.93), and then those in the frustration condition (*M*=112.26).

A one-way ANOVA was conducted to investigate cognitive load differences between the three design conditions based on participant ratings using the NASA-TLX. A significant main effect was found, *F*(2,153) = 16.65, *p* < .001. A Bonferroni post-hoc test was conducted to determine which design conditions were significantly different for the cognitive load dependent variable. Post-hoc results indicated that those in the Boredom condition experienced significantly lower levels of cognitive load (*M*=36.82) compared to those in the Flow condition (*M*=47.92) and those in the Frustration condition (*M*=50.76). There were no significant differences between the Flow and Frustration conditions.

Leveraging the inherent diagnosticity of the NASA-TLX, each of the six subscales that measure self-report workload demand were analyzed using one-way ANOVAs. Descriptive statistics for all six subscales can be found in **Table 3**.

Table 3.TLX Workload Breakdown Descriptives

TLX Scales	Conditions	*N*	*M*	*SD*
Mental	Boredom	48	36.15	25.563
	Flow	53	65.00	16.899
	Frustration	55	71.09	20.292
Physical	Boredom	48	16.15	20.323
	Flow	53	16.70	17.622
	Frustration	55	22.36	22.066
Temporal	Boredom	48	20.31	18.720
	Flow	53	26.04	20.484
	Frustration	55	30.36	22.399
Performance	Boredom	48	82.40	18.878
	Flow	53	68.96	18.328
	Frustration	55	43.91	29.118
Effort	Boredom	48	39.58	29.658
	Flow	53	61.42	17.469
	Frustration	55	66.64	20.482
Frustration	Boredom	48	26.35	23.218
	Flow	53	49.43	27.151
	Frustration	55	70.18	26.246

Results from the ANOVAs revealed significant main effects for Mental demand (*F*(2,153) = 39.42, *p* < .001), Temporal demand (*F*(2,153) = 3.04, *p* = .05), Performance (*F*(2,153) = 38.00, *p* < .001), Effort (*F*(2,153) = 19.80, *p* < .001), and Frustration (*F*(2,153) = 37.34, *p* < .001). Physical demand was not found to be significant, *F*(2,153) = 1.56, *p* = .21.

Table 4. NASA-TLX Subscale Post-hoc Mean Differences

Subscale	Groups	Boredom	Flow
Mental Demand	Boredom	--	--
	Flow	28.85***	--
	Frustration	34.95***	6.09
Temporal Demand	Boredom	--	--
	Flow	5.73	--
	Frustration	10.05*	4.33
Performance	Boredom	--	--
	Flow	-13.43**	--
	Frustration	-38.49***	-25.05***
Effort	Boredom	--	--
	Flow	21.83***	--
	Frustration	27.05***	5.22
Frustration	Boredom	--	--
	Flow	23.08***	--
	Frustration	43.83***	20.75***

Note: *n*=155, **p*<.05, ***p*<.01, ****p*<.001

Table 4 shows the results from the post-hoc test. Participants in the Boredom condition (*M*=36.15) reported significantly lower *mental demands* than those in the Flow (*M*=65) and Frustration (*M*=71.10) conditions. Participants in the Boredom condition (*M*=20.31) reported significantly lower *temporal demands* than those in the Frustration condition (*M*=30.36). Participants in the Boredom condition (*M*=82.40) reported significantly higher perceived *performance* than those in the Flow (*M*=68.96) and Frustration (*M*=43.91) conditions. Additionally, participants in the Flow condition reported significantly lower perceived *performance* than those in the Frustration condition. Participants in the Boredom condition (*M*=39.58) reported significantly lower perceived *effort* than those in the Flow (*M*=61.42) and Frustration (*M*=66.64) conditions. Participants in the Boredom condition (*M*=26.35) reported significantly lower *frustration* than those in the Flow (*M*=49.43) and Frustration (*M*=70.18) conditions. Additionally, participants in the Flow condition reported significantly lower *frustration* than those in the Frustration condition.

DISCUSSION

When developing a complex stimulus, such as the game described above, it is imperative to first ensure that the conditions were successfully designed to reflect the desired condition states. By analyzing the individual game-level, user-behavior data based on a Flow Theory framework, a strong case can be made to support and provide confidence that what was intended to be measured was indeed measured.

The six subscales measured using the self-report NASA-TLX shed additional insight into the perceived differences between the three conditions. Analyzing each subscale individually reveals a trend that, although the user-behavior data for the Flow and Frustration conditions were significantly different, they may have not been different enough to create environments where cognitive load and engagement differences could be found. The *performance* and the *frustration* subscales were the only two measurements that produced significant differences between the Flow and Frustration conditions. As would be expected, the Frustration condition was perceived to be more frustrating than the Flow condition, but whether the high levels in frustration contributed to the lower levels of perceived performance in the Frustration condition is unclear. Additional research into the direction of influence between these two factors could provide insight into how feelings of frustration affect performance.

The usefulness of this game as an experimental tool is that different game levels can be easily programmed to cause participants to feel varying levels of frustration and boredom. By extension, manipulations of this affective dimension and its impact on engagement can be used in studies of human-machine interactions in the areas of training and decision-making. The game described only shows one example of how manipulating game challenge can affect skill development and optimal feelings of Flow. Other implementations of these manipulations could be used to create more than three conditions, or only one condition that internally moves between boredom Flow and frustration over time. Analysis of performance data coupled with appropriate self-report data could be used to draw conclusions with confidence.

REFERENCES

Amazon.com. (2009). Mechanical Turk. Retrieved May 2, 2009, from https://www.mturk.com/mturk/welcome

Clarke, D. (2007). Bloxorz [Computer Game]: DX Interactive.

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

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75-98.

Malone, T. W. (1980). *What makes things fun to learn? heuristics for designing instructional computer games*. Paper presented at the Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems.

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology, 59*(6), 938-955.

Pavlas, D., Heyne, K., Bedwell, W., Lazzara, E., & Salas, E. (2010). *Game-based Learning: The Impact of Flow State and Videogame Selfefficacy*. Paper presented at the Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA.

Rabin, S. (2005). *Introduction to Game Development*. Boston: Charles River Media.

Sharek, D. (2009a). Block Walk (Version 1.0) [Computer Game]. Raleigh, NC.

Sharek, D. (2009b). Online NASA-TLX (Version 0.1) [Computer Software]. Raleigh, NC.