This article is a part of a symposium titled: *Engagement and Simulation/Gaming*

Simulation & Gaming
2014, Vol. 45(4-5) 569–592

DOI: 10.1177/1046878114554176
sag.sagepub.com

Measuring Video Game Engagement Through the Cognitive and Affective Dimensions

David Sharek[1] and Eric Wiebe[1]

Abstract

Aim. This study investigates a novel **technique** for **measuring** video game **engagement** by capturing **behavioral data** with little task interference.

Background. **Flow Theory** and **Cognitive Load Theory** provide insight into understanding engagement by analyzing the **interactions** between **skill** and **task challenge**. The development of this real-time measurement of engagement provides a more precise **diagnostic** method for designing challenging, yet cognitively engaging, tasks.

Method. Flow Theory guided the design of three conditions (**Boredom**, **Flow**, and **Frustration**) for a video game played by 156 participants. The authors tested a potential measure of engagement based on the number of times a participant clicked a game-clock during gameplay and during intermissions, along with performance and workload data. We differentiated the three conditions by synthesizing **game-clock clicks** during gameplay, during intermission, and overall cognitive load.

Results. **Boredom** showed lower cognitive load than **Flow** and **Frustration**. Frustration had significantly lower game-clock clicks during gameplay and significantly higher clicks during intermission than Flow or Boredom.

Conclusion. This study's measurement approach could potentially be used to measure **cognitive** and **affective** elements of **engagement**, helping to

[1]North Carolina State University, USA

Corresponding Author:
David Sharek, Department of Psychology, North Carolina State University, 640 Poe Hall, Campus Box 7650, Raleigh, NC 27695-7650, USA.
Email: dsharek@playgraph.com

identify where in a task a person may reach a point of **disengagement**, and where they may choose to **reengage**.

Keywords
behavioral data, Boredom, Cognitive Load Theory, disengagement, education, engagement, Flow, Flow Theory, Frustration, game-clock clicks, games, interactions, mental workload, reengagement, skill, task challenge, training

Recent research on computer-based gaming as a paradigm for training and educational software design has often hinged on the contention that gaming environments are inherently engaging (Admiraal, Huizenga, Akkerman, & ten Dam, 2011; Burgos, Tattersall, & Koper, 2007; Ciavarro, Dobson, & Goodman, 2008; Whitton, 2011). In fact, national competitions (e.g., Joan Ganz Cooney Center [JGCC], 2013) have been initiated to encourage developers to create game-based learning environments under the assumption that these environments will enhance student engagement in hard to teach subject areas. The assumed model is that higher levels of engagement leads to increased learning around skills-based or higher order content. However, in a review of engagement literature, it quickly becomes apparent that almost as many definitions of engagement exist as articles on the subject (e.g., O'Brien & Toms, 2008; Paradise, 2008; Przybylski, Rigby, & Ryan, 2010; Webster & Ho, 1997; Whitton, 2011).

Many researchers and policy makers have called for increased attention to the affective dimension of learning, including behaviors surrounding the construct of engagement (Rodrigo, 2011; Walker, Greene, & Mansell, 2006). Fredricks, Blumenfeld, and Paris (2004) frame their literature review broadly, and include three broad ways in which educational engagement can be framed: (a) behavioral engagement (e.g., participation in school events and activities), (b) emotional engagement (e.g., positive and negative reactions to teachers and classmates), and (c) cognitive engagement (e.g., thoughtful and purposeful investment in learning). Our interest here is in an outcome of higher cognitive engagement, but we acknowledge that understanding cognitive engagement will, out of necessity, have to account for the role of emotional (i.e., affective) engagement (Baker, D'Mello, Rodrigo, & Graesser, 2010). Furthermore, the impetus for studying general video game engagement here will be to inform future research on engagement in the more complex realm of serious games (Abt, 1970) where individuals participate in game-based computer environments for the purposes of learning. In such game-based environments, one of the key design strategies is to enhance intrinsic motivation of an individual to engage in the game environment (Cordova & Lepper, 1996; Malone & Lepper, 1987). Researchers have recognized the central role of challenge in creating this motivation to stay engaged in gaming environments (Malone & Lepper, 1987; Webster & Ho, 1997). Challenge can be linked to engagement in a number of ways, including the opportunity to demonstrate competency in order to enhance motivation (Przybylski et al., 2010). Game developers have also recognized that individuals cannot sustain high levels of engagement all of the time and that gamers, out of physiological and psychological necessity, will go through cycles of engagement and disengagement (Garris, Ahlers, & Driskell, 2002; O'Brien & Toms, 2008). Through structural devices such as game levels and cut scenes, most video games provide natural break points to allow for a natural cycle of engagement and disengagement.

Flow Theory (FT; Csikszentmihalyi, 1988, 1990) provides a particularly compelling high-level model that has been used by a number of researchers and game developers to look at engagement in the context of gaming environments. Flow is particularly well suited to describing the positive psychological state when individuals engage in experiences that are perceived to be challenging yet still obtainable (Rabin, 2005; Schell, 2008) often leading to high engagement. Central to this understanding of engagement through challenging contexts is a model of cognitive workload that can describe both the level of cognitive load that is inherent to the task design relative to the user's ability, but also account for the level of voluntary cognitive effort that users are willing to commit to the task. Cognitive Load Theory (CLT; Kalyuga, Chandler, & Sweller, 2004; Paas, Renkl, & Sweller, 2003; Sweller, 1988) is a particularly apt framework for understanding the different elements that contribute to the overall workload of an individual in a learning environment. The purpose of this article is to operationalize a measure of engagement through the combination of FT and CLT into a theoretical framework to help explain the engagement process. Together, CLT and FT provide a robust cognitive-affective framework for exploring engagement in learning.

Flow Theory

FT is one of the more thorough and practical frameworks from which to study engagement (O'Brien & Toms, 2008; Pavlas, Heyne, Bedwell, Lazzara, & Salas, 2010). Flow is described as an optimal experience that includes feelings of exhilaration and deep enjoyment (Csikszentmihalyi, 1990). People in the flow state are intrinsically motivated and commonly report the following: focused concentration, feelings of control, and a lack of awareness of time (Csikszentmihalyi, 1988). Csikszentmihalyi proposed that one of the most powerful experiences in flow occurs when a person is faced with difficult obstacles that they judge are worthwhile to overcome. These experiences are commonly found in video games (Weibel, Wissmath, Habegger, Steiner, & Groner, 2008), which typically provide a progression of challenging, but obtainable experiences (Rabin, 2005; Schell, 2008). This willingness to pursue obtainable challenges is deeply rooted in social cognitive theory of self-efficacy (Bandura, 1977).

Csikszentmihalyi (1990) indicates that one of the hallmarks of flow is a state of engaged concentration, an important component of deep learning. Baker and colleagues (2010) summarize engaged concentration as a state of engagement with a task such that concentration is intense, attention focused, and involvement complete. During such periods of intense concentration during flow, one of the possible

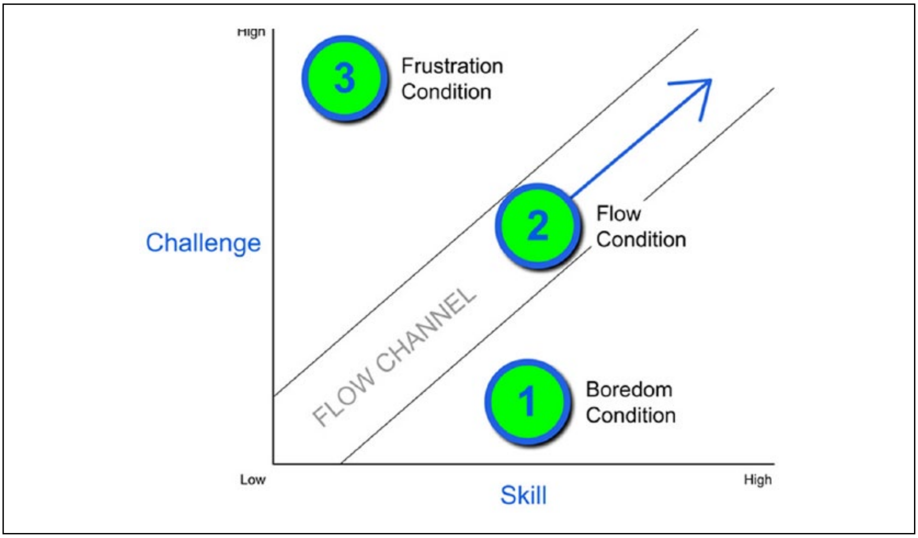

Figure 1. Three conditions based on Flow Theory.
Source. Adapted from Csikszentmihalyi (1975, 1990).

outcomes is a loss of a sense of time, or time distortion. As discussed in detail below, more classic models of cognitive resource allocation can also explain this—that is, if the engaged task encumbers all or most of one's cognitive resources, then that leaves few resources to monitor time on task.

To understand the dynamic nature of engagement, it is important first to take a high-level look at the relationship between engagement and flow. In order to engage, a person must be motivated to allocate their cognitive resources to a task (Burns & Gentry, 1998; Przybylski et al., 2010; Webster, Trevino, & Ryan, 1993). Once engaged, a person has the opportunity to enter a flow state. The flow state describes the active response a person has to the challenges of the task. At times during the task, a person can become disengaged (O'Brien & Toms, 2008). During this point of disengagement, it follows that the person is also released from their flow state. After a period of disengagement, a person may have the opportunity to reengage in the task, and subsequently, he or she may enter back into a state of flow.

Good game design techniques promote a state of flow by managing challenge (Garris et al., 2002), such that the given game task is dynamically balanced with individual skill levels (see Figure 1). When challenge is out of balance, either Boredom (high skill and low task difficulty) or Frustration (low skill and high task difficulty) may become the over-riding emotional state. Neither is considered part of an optimal flow state and either one may result in less than optimal learning through a decrease in motivation to continue with the learning experience (Baker et al., 2010; Kort, Reilly, & Picard, 2001; Patrick, Skinner, & Connell, 1993).

Cognitive Load Theory

CLT also provides a framework for understanding the relationship between challenge and skill. Similar to the predictions of FT, CLT also hypothesizes that an optimal cognitive state for learning will exist when the learner's ability (skill) and the complexity (challenge) of their learning task are appropriately balanced (Moreno & Park, 2010). Building on the assumptions that the learner has limited working memory, and virtually unlimited long-term memory, CLT can be used to aid the designer in selecting the appropriate elements to enhance training (Kalyuga et al., 2004).

Research involving CLT focuses on three types of load—extraneous, intrinsic, and germane—and assumes that all three loads are additive (Paas et al., 2003). Intrinsic load is of most interest for this study because it represents the load generated due to the spread between a person's current skill and the task challenges that the game presents to them. Intrinsic load is likely to be relatively uniform across users when facing a novel task, for example, when beginning a new video game without having previous experience with a similar type of game. As intrinsic load depends both on the learner's initial expertise and his or her developing schemata, it is important to remember that the relative challenge, and thus his or her cognitive load, will change as he or she develops expertise. Again looking at Figure 1, based on CLT, intrinsic load is likely to be highest when the ratio of challenge to skill is greatest (in the Frustration Area) and lowest where the ratio has reversed (the Boredom Area). We hypothesized that we would identify an optimal balance of challenge to skill in a learning or training situation that would represent a middle range of intrinsic load. By extension, we believe that as novices gain experience while playing a game, the game will need to raise the challenge level to keep individuals in this optimal challenge-skill ratio range. Outside this middle range, learning would not be optimized because either cognitive capacity has been overloaded (Frustration) or underutilized (Boredom).

Unlike intrinsic load, it is possible to reduce extraneous load, and indeed, cognitive load researchers have focused much of their time on finding ways to reduce it in instructional materials. Extraneous load is created by the formatting and presentation of information (Sweller, 1988). Although extraneous load is not clearly defined (Schnotz & Kuerschner, 2007), it can generally be thought of as the formatting and implementation of instructional design elements that do not directly aid schema acquisition or schema automation (Sweller, 2005). As the game interface is constant across conditions in this study, the authors assume that the level of extraneous load across users and conditions is uniform.

Germane cognitive load (Paas & van Merrienboer, 1994) can be thought of as the cognitive effort committed by the learner toward the learning goal. Another way of considering germane load is to think of it as a form of intrinsic motivation (Mayer & Moreno, 2010), and that being motivated toward task completion encourages a higher degree of cognitive processing. Using this line of reasoning, germane load provides a link to the intrinsic motivation and positive affect associated with active engagement and entering flow. This conceptualization of germane load is consistent with Flow Theory's predicted outcomes of learners in a state of flow. Careful management of intrinsic load creates a condition of difficult, but achievable challenge that induces a state of flow. Individuals experiencing this positive affect are likely to be motivated to put forth increased germane load. If an individual is in a state of frustration, neither the positive affect nor the cognitive capacity is available to devote to germane load. Conversely, in a state of boredom, neither the positive affect to motivate an individual nor the cognitive need exists for the individual to put forth resources toward germane load.

Currently, no well-established measure of germane load exists (Brunken, Paas, & Moreno, 2010); however, the NASA-TLX has been used extensively to measure overall cognitive load. The NASA-TLX is also a useful tool for diagnosing specific task-related components of load (Byers, Bittner, & Hill, 1989; Hart & Staveland, 1988; Wiebe, Roberts, & Behrend, 2010) based on its use of six subscales (Mental demand, Temporal demand, Performance, Effort, Frustration, and Physical demand). Although it is a retrospective, self-report measure, it has demonstrated itself as a reliable means of providing a snapshot of overall cognitive load (Wiebe et al., 2010). Furthermore, analyzing data at the subscale level has the potential of providing additional (albeit limited) diagnostic information on a person's affective and cognitive state.

Measuring Engagement

Probably the most common approach to measuring engagement is through the use of self-report questionnaires that are administered after a given task (e.g., O'Brien & Toms, 2010). Although useful, these measures lack a degree of diagnosticity due, in part, to the fact that they are retrospective, by nature, and therefore report on a summative assessment of experience. Therefore, it can be difficult to report retrospectively on different states of affect at different stages during a task. The solution of including self-report measures of engagement at more points during the task, in turn, risks disrupting the very focused engagement (and Flow) being sought. Unfortunately, the alternative of covert capture of physiological and/or behavioral data external to the software environment is fraught with its own challenges of cost, logistics, and scalability (cf. Baker et al., 2010; D'Mello, Taylor, & Graesser, 2007). Analysis of log file data of user-behavior within the game environment shows promise (e.g., Kapoor, Burleson, & Picard, 2007), but often comes up against considerable diagnostic complexities once one moves past the simplest low-level analysis of data. Measuring a person's awareness of the passage of time as a secondary task may provide another possible solution (McCarthy & Wright, 2004; O'Brien & Toms, 2008; Skadberg & Kimmel, 2004). O'Brien and Toms (2008) find evidence of this attribute of engagement and linked such spatiotemporal experiences to a person's self and external awareness. Similarly, Skadberg and Kimmel (2004) find that people's self-reported sense of a distortion of time was an accurate measure of flow.

FT predicts that when a person engages in a task, they may enter a flow state and experience positive affect and a distorted sense of time. However, FT does not provide an explicit cognitive or affective mechanism for how or why this happens. In contrast, we can think of a person who is very much aware of the passage of time as indicated by, for example, continual glances at his or her watch as not in flow, that is, bored by the task. McCarthy and Wright (2004) propose the idea that a person's sense of time can change depending on his or her level of emotional engagement. In addition, they suggest that a person's willingness to continue to engage in a task directly relates to his or her perception of time.

Although measuring engagement in a controlled laboratory setting has been largely under-studied (O'Brien & Toms, 2008), prospective memory research (Cook, Marsh, & Hicks, 2005; Einstein & McDaniel, 1990) introduces a potential clock mechanism for measuring one of the more observable characteristics of engagement; that is, the awareness of the passage of time (O'Brien & Toms, 2008). In addition, research on the effects of cognitive load on secondary task performance (Ogden, Levine, & Eisner, 1979) provides further support for investigating the awareness of time within the framework of FT and CLT. For example, research has found that when people participate in a task that requires the majority of their attentional resources, they will be less aware of the passage of time (Block, George, & Reed, 1980; Block, Hancock, & Zakay, 2010). These findings, then, predict that if a mechanism is made available for clicking on a button to check how much time is left for an experimentally set minimum amount of gameplay, those individuals who are in a flow state and engaged in gameplay are less likely to click on the clock button.

Current Study

As discussed above, a proposed integration of elements of both FT and CLT allows us to hypothesize this manipulation of the challenge-skill ratio to create conditions of frustration, boredom, and flow. What we report here is an attempt to shed light on how productive this line of thinking is and whether experimental manipulations and the instrumentation used here have potential for future explorations into the relationship of the cognitive and affective dimensions of engagement for learning. The study described in this article uses a video game designed to evoke frustration, boredom, or a flow state (see Figure 1) in conjunction with measurements of the awareness of the passage of time and cognitive load. This game design would specifically manipulate the ratio of challenge to skill in order to influence intrinsic cognitive load during gameplay. In addition, the game includes intermissions to provide natural points to transition between engagement and disengagement. The number of times a person metaphorically glances at his or her watch while playing or during intermission, the amount of time past the minimum time individuals are required to play, and a traditional subjective cognitive load measurement are collectively used to provide a more diagnostic picture of engagement (Wierwille & Eggemeier, 1993).

Based on CLT's framing of intrinsic load, this type of load will progressively increase as a person moves through the Boredom, Flow, and Frustration conditions. It is possible to predict what happens to germane load using just CLT. However, FT predicts higher motivation and engagement in the Flow condition based on the balance of challenge to skill. It could be that germane load will be higher in the Flow condition than in the Boredom and Frustration conditions. Assuming that extraneous load (the third component of cognitive load) is constant across the three conditions, it will be the differences in summation of intrinsic and germane load that will reveal differences in the self-report measure of overall load. With the increased germane load of the Flow condition, it is possible that the overall load of the Flow condition will be as much as the Frustration condition. However, the Boredom condition is likely to have lower overall load than either the Flow or Frustration conditions, because neither intrinsic nor germane load is likely to be high.

CLT and FT need to be brought together to understand the likely awareness of the passage of time. High overall cognitive load will potentially lead to a suppression of the awareness of the passage of time (fewer clicks on a game-clock button). Similarly, those in the Flow condition, for both cognitive and affective reasons, will be less inclined to click on the game-clock button. Those in the Boredom condition, however, will be in neither a cognitive nor an affective state to minimize game-clock button clicks. The checking of game-clock button clicks during intermission should act as a control; where momentary cognitive load is low for all three conditions, but the affective state of gameplay should still carry over. Gameplay beyond the minimum required time should be greatest for those in the intrinsically motivating Flow condition.

Hypothesis 1 (H1): Those in the Frustration condition will perform worse (based on a series of game metrics) than those in the Flow and Boredom conditions.
Hypothesis 2 (H2): During gameplay, those in the Boredom condition will demonstrate a greater awareness of the passage of time (higher game-clock clicks) than those in the Flow and Frustration conditions.
Hypothesis 3 (H3): During intermissions, those in the Frustration condition should see a relative rise in the number of game-clock intermission clicks, similar to those in the Boredom condition.
Hypothesis 4 (H4): Those in the Flow condition will play the game for a significantly longer period of time beyond the minimum required time compared with those in the Boredom and Frustration conditions.
Hypothesis 5 (H5): Those in the Boredom condition will rate their overall cognitive load significantly lower than those in the Flow and Frustration conditions.

Method

Participants

A total of 169 people were recruited and paid 80 cents through Amazon's Mechanical Turk (Amazon.com, 2012). The Mechanical Turk is an online crowdsourcing tool that connects people to online tasks. Research shows that the Mechanical Turk is a viable participant recruitment tool for online research (Behrend, Sharek, Meade, & Wiebe, 2011). After agreeing to complete the task, we randomly assigned individuals to one of three conditions. 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 we removed them from the analysis. Of the remaining 156 participants,

Figure 2. Screenshot of gameplay in the video game, BLOCK WALK.
Source. Sharek (2010).

78% were from the United States, 58% were female, and 42% were male (*M* age = 30.79, *SD* = 10.22). Thirty-three percent (evenly distributed between conditions) reported that they had previous experience with a similar type of puzzle game.

Stimuli and Apparatus

Participants played a strategy game called BLOCK WALK (Sharek, 2010; see Figure 2). The game mechanics are based on BLOXORZ (Clarke, 2007). The goal of the isometric tile-based game (see Figure 3) is to move a rectangular block, comprised of two differently colored cubes, toward a goal 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. Moving the block while it is standing up on a tile will cause it to fall down and occupy two tiles in the direction of movement; thus, successive movements in this direction will increment the block by two tiles. If the block is on its side, either a player can tilt it upward for a two-tile move or it can be rolled sideways for a one-tile move. This difference in block orientation and degree of movement makes it difficult to predict the block's future position a few moves from its starting point. Players can learn certain sequences of moves to position the block more accurately; learning these sequences are usually only possible through experimentation and practice. Game-level challenge is manipulated by (a) the placement of the tiles, (b) the complexity of directional moves needed to navigate to the goal, and (c) whether either end of the block can land on the target or whether a specific colored side needs to land on it.

FT guided the creation of three conditions (see Figure 1). Participants in Design *Condition 1 (Boredom: Skill greater than game challenge)* only played easy game

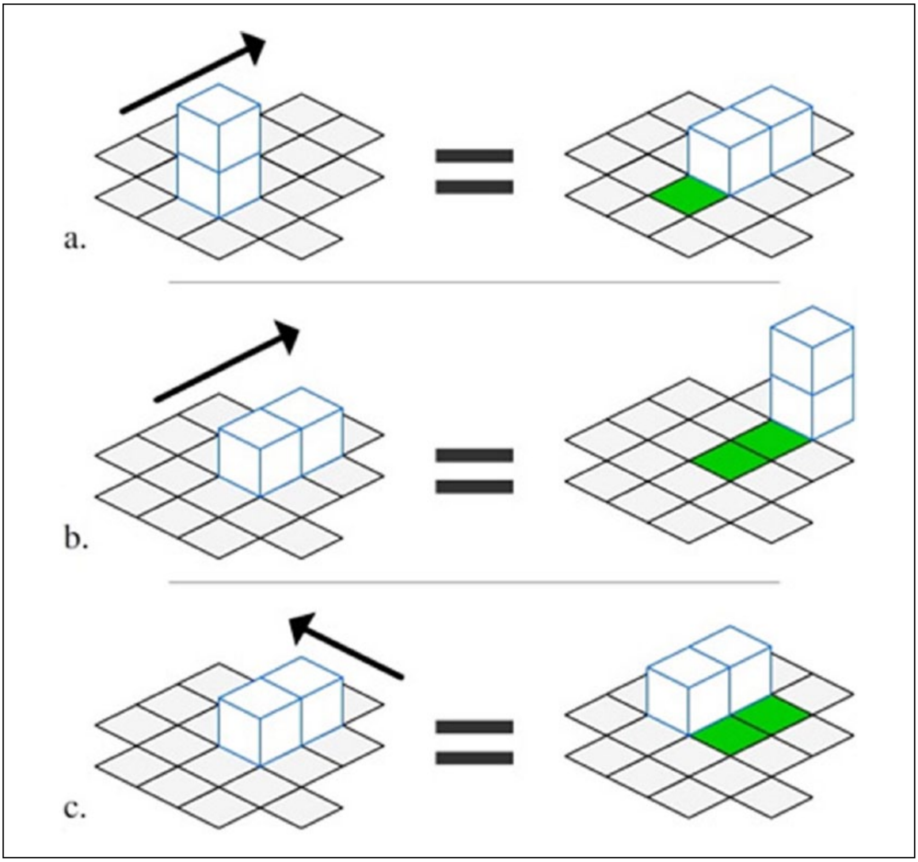

Figure 3. Moving the block around the game board in the BLOCK WALK game.
Note. a. When the block is standing up, any horizontal or vertical movement will cause it to fall down. Notice how the block no longer remains over the starting tile position (shown in green). b. When the block is laying down, it can be moved to a standing position if the movement is along the longer axis of the block. A combination of Moves a and b will move the block three squares along an axis. c. If the block is laying down and moved along block's shorter axis, it will continue to remain laying down and only move one square at a time. Combinations of Moves a and c can move and offset the block.

levels where a few simple moves were required to position the block over the goal. As players progress through levels, the game difficulty did not increase. Participants in *Design Condition 2 (Flow: Skill matched with game challenge)* were introduced to the game through easy levels similar to those in the *Boredom* condition. As they solved each level, the number of required moves and the complexity of the moves were increased. As the player progresses through the easier levels, it is assumed that they are mastering the game mechanics and strategies and are ready for more difficult levels where strategy becomes increasingly critical. Participants in *Design Condition 3 (Frustration: Skill less than game challenge)* were presented with game levels that could only be solved through complex sequences of moves with the number of moves commonly reaching into the hundreds. The game also enhanced the degree of challenge 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 game did not give the participants in this condition an opportunity to learn the idiosyncrasies of the block's movements and develop strategic knowledge progressively by scaffolding through simpler levels. A prior research study (Sharek & Wiebe, 2011) validated all three condition designs based on gameplay metrics and self-report data.

Procedure

The authors conducted the experiment over the Internet. We randomly assigned participants to one of the three conditions and then introduced to the game-clock, which functioned experimentally as a measure of the degree of engagement with the task, and indirectly the degree of intrinsic motivation. We asked participants to interact with the game-clock until they were familiar with its functions.

Although we told participants that the entire experimental procedure should take no longer than 30 minutes, we did not tell them that the minimum required gameplay time was 10 minutes. During gameplay, participants could click on a button labeled, "TIME," located at the bottom of the game's screen. Clicking on the button caused the game-clock to slide out and display a message indicating whether or not the minimum time was up without giving an indication of how much time had passed or was remaining. Clicking on the TIME button before reaching the minimum amount of time would cause a window to slide out with a red background and the words, "*The minimum time has not yet been reached*." After 3 seconds, the game-clock would automatically slide off-screen. The participants did not know that, when they reached the minimum time, a window would automatically slide out with a green background and the message, "*The minimum required time to play is up. You may continue playing if you want*." Two buttons would also be included with the message and labeled, "EXIT GAME" and "KEEP PLAYING." Clicking on the Keep Playing button would slide the window off-screen, while clicking on the exit button would immediately end the game. The game-clock could be pressed both during gameplay and during a short 5-second intermission between levels. After exiting the game, participants were asked to complete the NASA-TLX workload instrument (Hart & Staveland, 1988).

Results

Design Conditions Validation

The authors conducted a manipulation check to determine whether the three design conditions (Boredom, Flow, and Frustration) accurately reflected their corresponding flow states (i.e., challenge-to-skill ratio). Results from four one-way ANOVAs indicated significant main effects between all three design conditions and the four dependent variables. *Directions*—the number of times a participant changed the block's

Table 1. Gameplay Performance Behavior, Standard Deviations, and Post Hoc Mean Differences.

				M differences	
Condition	*n*	*M*	*SD*	Boredom	Flow
Directions					
Boredom	48	4.98	8.06	—	—
Flow	53	20.67	27.62	15.67***	—
Frustration	55	46.44	64.93	41.47***	25.77***
Errors					
Boredom	48	0.28	0.89	—	—
Flow	53	1.47	3.12	1.19***	—
Frustration	55	4.54	9.27	4.25***	3.07***
Moves					
Boredom	48	12.76	16.46	—	—
Flow	53	41.23	52.32	28.47***	—
Frustration	55	105.42	148.27	92.66***	64.19***
Time per level (s)					
Boredom	48	16.33	17.96	—	—
Flow	53	48.01	58.27	31.68***	—
Frustration	55	117.42	146.26	101.10***	69.41***

***$p < .001$.

direction for each level, $F(2, 3104) = 397.22$, $p < .001$; *Errors*—the number of times the block was moved off of the game board and into the water, $F(2, 3104) = 219.11$, $p < .001$; *Moves*—the number of times the block was moved, $F(2, 3104) = 399.93$, $p < .001$; *Time per level*—the total length of time participants spent playing each level, $F(2, 3104) = 450.25$, $p < .001$. Table 1 shows the means, standard deviations, and results from the post hoc tests.

Cognitive Load

We conducted a one-way ANOVA to investigate cognitive load differences between the three design conditions based on participant ratings using the unweighted composite NASA-TLX measure (Byers et al., 1989; Hart & Staveland, 1988; Wiebe et al., 2010). A significant main effect for cognitive load was found, $F(2, 153) = 16.65$, $p < .001$. Post hoc results indicated that those in the Boredom condition experienced significantly lower levels of cognitive load ($M = 36.82$) compared with those in the Flow condition ($M = 47.92$) and those in the Frustration condition ($M = 50.76$). We found no significant differences between the Flow and Frustration conditions.

Leveraging the inherent diagnosticity of the NASA-TLX, we analyzed each of the six subscales using one-way ANOVAs. Table 2 shows the means and standard deviations. Results from the ANOVAs revealed significant main effects for Mental demand,

Table 2. NASA-TLX Subscale Means, Standard Deviations, and Post Hoc Mean Differences.

				M differences	
Condition	*n*	*M*	*SD*	Boredom	Flow
Mental demand					
Boredom	48	36.15	25.56	—	—
Flow	53	65.00	16.90	28.85***	—
Frustration	55	71.09	20.29	34.95***	6.09
Temporal demand					
Boredom	48	20.31	18.72	—	—
Flow	53	26.04	20.48	5.73	—
Frustration	55	30.36	22.40	10.05*	4.33
Performance					
Boredom	48	82.40	18.88	—	—
Flow	53	68.96	18.33	−13.43**	—
Frustration	55	43.91	29.12	−38.49***	−25.05***
Effort					
Boredom	48	39.58	29.66	—	—
Flow	53	61.42	17.47	21.83***	—
Frustration	55	66.64	20.48	27.05***	5.22
Frustration					
Boredom	48	26.35	23.22	—	—
Flow	53	49.43	27.15	23.08***	—
Frustration	55	70.18	26.25	43.83***	20.75***

*$p < .05$. **$p < .01$. ***$p < .001$.

$F(2, 153) = 39.42, p < .001$; 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 2 shows the results from the post hoc tests.

Game-Clock Clicks and the Passage of Time

We also analyzed the number of game-clock clicks during gameplay, game-clock clicks during intermissions (an approximately 5-second transition period between levels), and the amount of time spent over and above the minimum time. To control for any zeros in the data, we conducted a log transform and then analyzed the variables across the three design conditions using a MANOVA. Results from this analysis revealed a significant main effect for the design conditions, $F(2, 156) = 7.44, p < .001, \eta^2 = .13$.

Follow-up ANOVAs revealed significant main effects between the three design conditions and game-clock clicks during gameplay, $F(2, 156) = 4.00, p = .001, \eta^2 = .01$, and intermission game-clock clicks, $F(2, 156) = 4.76, p < .001, \eta^2 = .16$. We found

Table 3. Game-Clock Clicks Means, Standard Deviations, and Post Hoc Mean Differences.

Condition	*n*	*M*	*SD*	*M* differences	
				Boredom	Flow
Game-clock clicks during gameplay					
Boredom	48	3.54	2.74	—	—
Flow	53	3.60	3.05	0.06	—
Frustration	55	1.96	2.33	−1.58*	−1.64**
Game-clock clicks during intermissions					
Boredom	48	0.69	2.62	—	—
Flow	53	0.79	1.03	0.1	—
Frustration	55	2.02	2.41	1.33**	1.23**

*$p < .05$. **$p < .01$.

no significant main effects for time spent over the minimum required time. Table 3 shows the means, standard deviations, and results from the post hoc tests.

Discussion

This study used FT and CLT to operationalize the design of three different conditions of skill/challenge matches. Although one can ostensibly call the material used in this study a "game," it is also a high face-value representation of the manipulation of skill and challenge found in most training scenarios. The results indicate that the three design conditions were operationalized as predicted, as revealed by the results of the manipulation check. With these cognitive-affective states established, the research demonstrated that integrating the game-clock measure during and between gameplay with an overall measure of cognitive load provided a productive method for conceptualizing and measuring engagement. The findings as a whole did not show a simple relationship between the level of reported cognitive load and affective state, as revealed by the game-clock. This insight arose in part from exploiting a natural element of gameplay (and learning): the movement between engagement and disengagement. The results of the study revealed that integrating data across the period of engagement during game levels and disengagement during intermissions provided important insight as to the cognitive and affective state of the user. We discuss these findings in detail below.

The analysis of performance measures fully supported H1. Those in the Frustration condition performed the worst and produced more errors compared with those in the Flow and Boredom conditions. The Boredom condition created the fewest errors. The analysis partially supported H2, with the Frustration condition showing significantly fewer game-clock clicks during gameplay than Boredom and Flow. However, we saw no difference in clicks between Boredom and Flow during gameplay, even though we predicted both cognitive load and positive affect to be lower for the Boredom condition. The analysis also partially supported H3 with the results showing that participants in the Frustration condition clicked significantly more times during intermissions compared those in the Boredom and Flow conditions. The hypothesis predicted that intermission clicks for Frustration and Boredom conditions would be similar to each other and higher than the Flow condition. However, Frustration ended up being higher than both Boredom and Flow. H4 was not supported, with no differences seen in beyond-minimum gameplay time between the three levels. The analysis of the composite NASA-TLX scores supported H5, revealing that the participants in the Boredom condition had a significantly lower cognitive load score than the Flow and Frustration conditions.

Although we might have wanted to have the increased diagnosticity of also seeing a significant difference between the Flow and Frustration conditions, combining the results of the game-clock intermission to gameplay clock clicks with the NASA-TLX scores gives us a more robust, if somewhat more complex picture of engagement during gameplay.

These results—both those expected and unexpected—can be interpreted in a number of ways; however, the following is a compelling approach, based on both the framework presented initially and our findings from this study. The cognitive load measurement data supported our hypotheses and seem to indicate that both the Flow and Frustration condition participants experienced high levels of load. As this measure is overall load, what is not known is exactly how much of this load was intrinsic to the challenge of the task and how much was germane (i.e., participant initiated). Based on FT, the Flow condition participants would have had both the capacity and inclination to raise their overall load level through increased germane load. The manipulation check analysis that confirmed the relative difficulty of gameplay in the three conditions supported this supposition, with the Flow condition participants demonstrating higher performance than the Frustration condition. This rise in germane load helps explain the lack of significant difference in overall load between Flow and Frustration. We propose that an affective state that desires increased engagement in the game drives, in part, the increase in germane load. In the Frustration condition, participants might, however, have high enough intrinsic load that they experienced neither the headroom nor inclination to increase germane load. By logical extension, the Boredom condition individuals would both have lower intrinsic load and no real inclination or need to raise germane load in order to solve a game level, thus leaving them with an overall lower load. Both the manipulation check analysis and the lower cognitive load self-report scores support this supposition about the Boredom condition. Figure 4 shows a summary of this interpretation.

Cognitive load, as measured through the post hoc, self-report measure of the NASA-TLX, does have its limitations in unpacking the state of the player. To begin with, the aggregate score is limited to reporting on overall cognitive load—across all levels of gameplay within a condition and across both in-gameplay and intermission periods. It is, as designed, also only reporting on cognitive load and not on an individual's affective state. Therefore, by itself, it only provides a partial picture of engagement. The question, then, is how we might use the game-clock click data to compliment the cognitive load scores to further explain the study results and better explicate the states of engagement individuals might have been in at different points in the game.

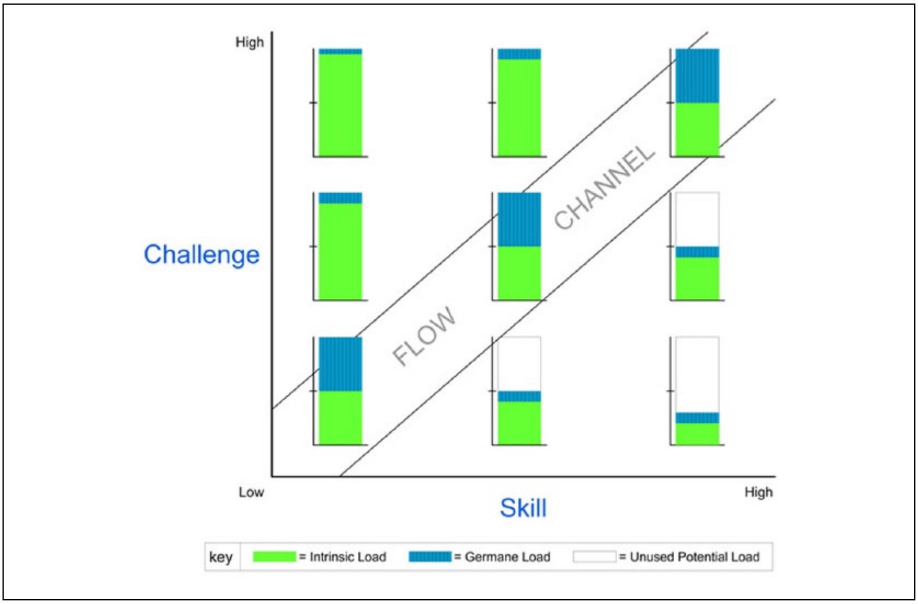

Figure 4. The flow model overlaid with stacked bar charts representing the additive intrinsic and germane loads during in-level gameplay.

It is likely that participants would move into a momentary lower level of cognitive load during intermissions. Although a measure of load such as the NASA-TLX might provide a direct measure of this change, the self-report survey nature of the instrument means a higher level of intrusiveness and the risk that the report is more retrospective of the overall experience than desired (Rubio, Díaz, Martín, & Puente, 2004). Game-clock clicks, used as a type of secondary task measure, would be less intrusive and more sensitive to changes in load. Therefore, it was useful to use this measure and analyze game-clock clicks both *during* and *between* game levels. Building on the interpretation of the cognitive load scores, the game-clock clicks—within and between game levels—revealed something about both the cognitive load and the affective state of the participant. Whereas the NASA-TLX scores revealed differences between the Boredom condition and the other two conditions (Flow and Frustration), the game-clock clicks provided a pattern of behavior that distinguished Frustration from Flow and Boredom. Looking at the intermission period, when the player should be released from most of their game-related cognitive load, the game-clock clicks should represent, primarily an affective state concerning motivation to continue gameplay. During this period, Frustration condition participants showed significantly higher clicks. One way of interpreting this result is that participants experienced low motivation to reengage with gameplay at the next level—in other words, they were "seeking an exit from the game." Thus, the resulting pattern of response of both the NASA-TLX and

Table 4. Overview of Cognitive-Affective Outcomes.

Condition	Cognitive load (NASA-TLX)	Motivation for reengagement (intermission game-clock clicks)
Boredom	Low	High
Flow	High	High
Frustration	High	Low

intermission game-clock clicks provides a way of distinguishing all three conditions. Table 4 provides a summary of this.

An appropriate follow-up question would be to seek an explanation for the significantly lower level of game-clock clicks for the Frustration condition during gameplay. One line of reasoning would be to employ the same cognitive load model used for the NASA-TLX score results. Those in the Frustration condition had the largest amount of their cognitive resource capacity tied up in intrinsic load compared with those in the Flow and Boredom conditions. CLT supports this explanation in terms of resource allocation—if the task load is too high, no additional working memory resources remain available to attend to the game-clock. When resources are freed during intermission to attend to the game-clock, the affective dimension of the participant's experience—explained by FT—can then drive whether the game-clock is attended to. Participants in the Flow condition, however, did not quite respond as expected based on this model. FT seemed to correctly predicted a more positive affect for the Flow condition that seemed to carry over to the intermission period (when cognitive load did not dominate click behavior), where Flow participants clicked on the game-clock significantly less than in the Frustration condition. However, it was predicted in H2 that in the more engaging Flow condition, players would allocate available capacity to germane load, leaving them with no more available load than the Frustration condition—a supposition backed by the NASA-TLX scores. A game-clock click pattern did not seem to occur based purely on a cognitive load explanation, with the Flow condition showing a similar pattern of in-gameplay clicking as the Boredom condition. This leads us to believe that possibly the affective dimension was also influencing the click patterns during gameplay.

This then leaves how to interpret the Boredom condition. It seems that the participants in the Flow and Boredom conditions, relative to Frustration, experienced a sustained positive affective state during intermission that suppressed game-clock clicks, even though they had the resources to do so. The Boredom condition may have created a kind of low-effort engagement. The psycho-motor manipulation of the game may have engaged the participant in a positive affective way without the need for creating a cognitively challenging flow state. This interpretation is backed by both the low cognitive load self-report rating on the NASA-TLX and the low game-clock clicks during intermission, when the affective dimension should be guiding behavior. We can think of the game over-time as another proxy for measuring affective state the same way intermission game-clock clicks does; those in a negative affective state would be seeking an exit and not willing to play the game in over-time. Although the results of the game over time dependent variable were not significant, it was trending such that it showed the inverse pattern of the intermission game-clock clicks: low over-time for the Frustration condition and higher for both Flow and Boredom. Our analysis found a large amount of variance for this dependent variable, likely contributing to the lack of significant findings.

The NASA-TLX subscales shed additional insight into the perceived differences between the three conditions (Table 2). Looking at the individual subscales, the *Performance* and *Frustration* subscales were the only measurements that produced significant differences between all three conditions. As would be expected, the Frustration condition was perceived to be more frustrating than the Flow condition that, in turn, was more frustrating than the Boredom condition. The Performance subscale showed the exact reverse pattern, with Boredom indicating the highest perceived performance and Frustration the lowest, with Flow in-between. These subscale results are parsimonious with a generally positive affective state for both the Boredom and Flow conditions relative to the Frustration condition. The Boredom participants also had a false sense of achievement based on the consistently easy game levels. Additional research into the direction of influence between these two factors could provide insight into how feelings of frustration affect performance. It could be that a game needs a certain amount of tension around frustration and performance to create the challenge of a skill/challenge ratio indicative of a flow state.

Conclusion

The data suggest that the game-clock, coupled with the NASA-TLX as a measurement of cognitive load, offers a useful analytic tool set of user-behavior related to engagement. The results of this study provided two important findings for researchers to explore further. First, we deployed the game-clock as an unobtrusive measurement tool both during gameplay and during the intermission period. It seems that the game-clock results during the intermission ended up being the most useful, as a foil to the NASA-TLX. Seen in Table 4, the NASA-TLX scores, when combined with the intermission game-clock clicks gave a combined cognitive-affective measure of user engagement in the game. Combining the results of these two measures, we could distinguish all three game conditions from each other. Just as important, a logical synthesis of CLT and FT provides us with a model that supports these findings. Although the game-clock clicks during play are not shown in Table 4, this does not mean that their data were un-interpretable or not useful in supporting these findings. They were, in the end, more complex to interpret, as they seemed to represent behaviors being influenced by both cognitive and affective dimensions. Instead, they represent useful, supporting, and triangulating data. We can say the same for the analysis of individual subscales of the TLX. Unfortunately, the amount of gameplay over the minimum time data did not provide supporting evidence. Researchers need to carry out further work in different serious games contexts to explore this as a potential measure.

Given that part of the goal of this study was to enhance our ability to measure facets of engagement, it is worth stepping back and looking, methodologically, at the strengths and weaknesses of the NASA-TLX and the game-clock clicking tool. The advantage of the game-clock's real-time monitoring system over retrospective self-report methods is that the game-clock provides an increased level of sensitivity by uncovering changes in task demands during participation, something the NASA-TLX cannot provide (Rubio et al., 2004). Workload sensitivity can be thought of as the degree to which a workload measure is able to discriminate between task-related workload differences (Eggemeier, Shingledecker, & Crabtree, 1985; Wierwille & Eggemeier, 1993). Specifically, the game-clock collects time-sensitive performance data that potentially we can use to extrapolate when cognitive overload occurs. In addition, the game-clock decreases measurement intrusiveness, which is described as a measurement tool's level of interference with a task which can potentially lead to degraded performance (Eggemeier et al., 1985). For example, for the NASA-TLX to provide a more precise workload measurement of a user's workload during specific stages in a task, we would have to administer it during the task, leading to a potential point of disengagement. However, the NASA-TLX provides a valuable triangulating measure of load that reflects the overall perception of load on the part of the participant. It also allows the researcher to help disaggregate game-clock click patterns driven by either cognitive or affective factors. Finally, the TLX subscales provide additional diagnosticity for the source of workload, along with perceptions of frustration and performance.

Future Research

For the game-clock-plus-cognitive load measure paradigm to be useful as a diagnostic tool in training or learning systems (i.e., eLearning), the training environment would need to contain discreet engagement/disengagement periods where clock clicks can be measured and compared. However, the purpose of this measurement paradigm is primarily to examine tasks that provide some degree of challenge, and is immediately useful for measuring engagement during eLearning courses that employ game-like environments. For example, the number of game-clock clicks during the period between completing a learning module and starting the next one could be compared with the number of game-clock clicks during in-module participation (Brunken, Seufert, & Paas, 2010; Wierwille & Eggemeier, 1993). This study has pointed out that measuring engagement without interfering with a task is an extremely complex undertaking. More detailed research will need to be conducted to determine when and how such a measurement tool could be used effectively. Results from this study may indicate that the amount of cognitive load required to essentially pause participation in the task and click on the game-clock could influence a person's ability to click on the clock. Task-switching literature may provide insight into how further research could be conducted in this regard. In addition, including a debriefing stage after a person has completed a serious game could provide further cognitive scaffolding that links the learning material in the game to real-world applications (Garris et al., 2002). Including such a debriefing stage was out of the scope of this exploratory research, but it would be essential to include in a follow-up learning retention study.

A promising application of the game-clock-plus-cognitive load measure is in the area of adaptive training/learning management. Rather than change aspects of a learning task after performance has degraded, potentially researchers could use a form of this combined measure to adaptively adjust the challenge/skill ratio to support a person's engagement and keep them in an optimal flow state for learning purposes. For example, considering the possibility that the game-clock can be used to predict when a person is beginning to reach the point of disengagement; aspects of the task could be changed to reduce frustration and/or cognitive load. Similarly, if the person is under-tasked yet still engaged while not entering flow (similar to our Boredom condition), challenge could be increased and thus increase learning efficiency. Understanding and measuring engagement is still a young area of research and therefore researchers need to maintain a focused, methodical approach to experimental design in order to develop a solid framework of engagement.

Author Contributions

The authors contributed equally to this article. DS conceived, designed, and performed the experiments; DS and EW wrote the final manuscript; DS wrote the first draft and did the bulk of the literature search; EW made numerous critiques and suggested specific wording; and DS designed most of the graphics and did most of the statistical analyses.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Abt, C. C. (1970). *Serious games*. New York, NY: Viking.

Admiraal, W., Huizenga, J., Akkerman, S., & ten Dam, G. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, *27*, 1185-1194.

Amazon.com. (2012). *Mechanical Turk*. Retrieved from https://www.mturk.com/mturk/

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*, 223-241.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191-215.

Behrend, T., Sharek, D., Meade, A., & Wiebe, E. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*, 800-813.

Block, R. A., George, E. J., & Reed, M. A. (1980). A watched pot sometimes boils: A study of duration experience. *Acta Psychologica*, *46*, 81-94.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*, 330-343.

Brunken, R., Paas, F., & Moreno, R. (2010). Current issues and open questions in cognitive load research. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 253-272). New York, NY: Cambridge University Press.

Brunken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 181-202). New York, NY: Cambridge University Press.

Burgos, D., Tattersall, C., & Koper, R. (2007). Re-purposing existing generic games and simulations for e-learning. *Computers in Human Behavior*, *23*, 2656-2667.

Burns, A. C., & Gentry, J. W. (1998). Motivating students to engage in experiential learning: A tension-to-learn theory. *Simulation & Gaming*, *29*, 133-151.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw Task Load Index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety* (pp. 481-485). London, England: Taylor & Francis.

Ciavarro, C., Dobson, M., & Goodman, D. (2008). Implicit learning as a design strategy for learning games: Alert Hockey. *Computers in Human Behavior*, *24*, 2862-2872.

Clarke, D. (2007). BLOXORZ [Computer game]. DX Interactive.

Cook, G., Marsh, R., & Hicks, J. (2005). Associating a time-based prospective memory task with an expected context can improve or impair intention completion. *Applied Cognitive Psychology*, *19*, 345-360.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*, 715-730.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In M. Csikszentmihalyi (Ed.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 15-35). Cambridge, UK: Cambridge University Press.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper & Row.

D'Mello, S. K., Taylor, R., & Graesser, A. C. (2007). Monitoring affective trajectories during complex learning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 203-208). Austin, TX: Cognitive Science Society.

Eggemeier, F. T., Shingledecker, C., & Crabtree, M. (1985, September 29-October 3). *Workload measurement in system design and evaluation*. Paper presented at the Human Factors and Ergonomics Society Annual Meeting Proceedings, Baltimore, MD.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 717-726.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59-109.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, *33*, 441-467.

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, The Netherlands: North Holland.

Joan Ganz Cooney Center. (2013). *National STEM video game challenge*. Available from http://stemchallenge.org

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors*, *46*, 567-581.

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, *65*, 724-736.

Kort, B., Reilly, R., & Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy—Building a learning companion. In *Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges* (pp. 43-48). Madison, WI: IEEE Computer Society.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude learning and instruction* (Vol. 3, pp. 223-253). Mahwah, NJ: Lawrence Erlbaum.

Mayer, R., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 131-152). New York, NY: Cambridge University Press.

McCarthy, J., & Wright, P. (2004). *Technology as experience*. Cambridge, MA: MIT Press.

Moreno, R., & Park, B. (2010). Cognitive Load Theory: Historical development and relation to other theories. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 9-28). New York, NY: Cambridge University Press.

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*, 938-955.

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, *61*, 50-69.

Ogden, G., Levine, J., & Eisner, E. (1979). Measurement of workload by secondary tasks. *Human Factors*, *21*, 529-548.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1-4.

Paas, F., & van Merrienboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122-133.

Paradise, A. (2008). *State of the industry report: ASTD's annual review of trends in workplace learning and performance*. Alexandria, VA: American Society for Training & Development.

Patrick, B. C., Skinner, E. A., & Connell, J. P. (1993). What motivates children's behavior and emotion? Joint effects of perceived control and autonomy in the academic domain. *Journal of Personality and Social Psychology*, *65*, 781-791.

Pavlas, D., Heyne, K., Bedwell, W., Lazzara, E., & Salas, E. (2010, September 27-October 1). *Game-based learning: The impact of flow state and videogame self-efficacy*. Paper presented at the Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*, 154-166.

Rabin, S. (2005). *Introduction to game development*. Boston, MA: Charles River Media.

Rodrigo, M. M. T. (2011). Dynamics of student cognitive-affective transitions during a mathematics game. *Simulation & Gaming*, *42*, 85-99.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, *53*, 61-86.

Schell, J. (2008). *The art of game design: A book of lenses*. Burlington, MA: Morgan Kaufmann.

Schnotz, W., & Kuerschner, C. (2007). A reconsideration of Cognitive Load Theory. *Educational Psychology Review*, *19*, 469-508.

Sharek, D. (2010). *BLOCK WALK*. Raleigh, NC: Playgraph. (725 Pierside Drive, Cary, NC 27519, USA)

Sharek, D., & Wiebe, E. (2011, September 19-23). *Using Flow Theory to design video games as experimental stimuli*. Human Factors and Ergonomics Society 55th Annual Meeting, Las Vegas, NV.

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' flow experience while browsing a web site: Its measurement, contributing factors and consequences. *Computers in Human Behavior*, *20*, 403-422.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.

Sweller, J. (2005). Implications of Cognitive Load Theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). New York, NY: Cambridge University Press.

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, *16*, 1-12.

Webster, J., & Ho, H. (1997). Audience engagement in multimedia presentations. *SIGMIS Database*, *28*(2), 63-77.

Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior*, *9*, 411-426.

Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., & Groner, R. (2008). Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, *24*, 2274-2291.

Whitton, N. (2011). Game engagement theory and adult learning. *Simulation & Gaming*, *42*, 596-609.

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior*, *26*, 474-481.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, *35*, 263-281.

Author Biographies

David Sharek makes people's lives happier by creating things that are easier to use. He holds a PhD in human factors and focuses his research on user experience, the psychological construct of engagement, and serious games development. He has created several online video games and computer programs used by thousands to conduct experimental research.

Contact: dsharek@playgraph.com.

Eric Wiebe is a professor of STEM Education at North Carolina State University. His research focuses on multi-modal communication of scientific and technical information for teaching and learning. Specific research programs include the use of intelligent tutors to support science learning in classrooms, and basic research in how instructional technologies including game-based learning shape student engagement and learning.

Contact: wiebe@ncsu.edu.