EXAMINATION OF USER ENGAGEMENT AND SUBJECTIVE STATES IN AN ONLINE TETRIS GAME

Su Shen[a], Huei-Yen Winnie Chen[a], Thomas J. Covey[c], Eduardo Mercado[b], Ann Bisantz[a]
[a] Industrial and Systems Engineering, University at Buffalo, the State University of New York
[b] Dept. of Psychology, University at Buffalo, the State University of New York
[c] Dept. of Neurology, University at Buffalo, the State University of New York

Engagement, as a multidimensional construct, fluctuates due to various factors. Previous research that examines engagement as a subjective experience was mainly concerned with the qualitative structure of engagement in different contexts (e.g., task difficulty, aesthetics). Few studies have examined the user's subjective engagement as a given task progresses over time. This paper reports findings from an online study that aimed to assess the usefulness of subjective measures in capture capturing momentary feelings related to task engagement, under varying duration (length of task) and difficulty conditions. The Short Stress State Questionnaire (SSSQ) and the User Engagement Scale (UES) were used to capture self-reported engagement during a Tetris video gameplay. The sensitivity of scales to task conditions and relationships among subscales were examined. Results showed that changes in SSSQ are sensitive to difficulty levels, and Engagement obtained from the SSSQ was highly correlated with UES subscales. SSSQ may be a particularly useful tool to capture participants' momentary feelings during a task via its Engagement, Distress, and Worry subscales.

INTRODUCTION

General references to engagement in literature emerged from decades ago such as Chapman's observation, "something that 'engages' us is something that draws us in, that attracts and holds our attention" (1997), and Laurel's statement, "Engagement is what happens when we can give ourselves over to a representational action, comfortably and unambiguously. It involves a kind of complicity" (1991). More recently, engagement has been explicitly studied in various contexts. In human-computer interaction studies, engagement is often referred to as a *perceived quality* that represents a qualitative character of subjective experience in relation to the features of the interactive activities (Montgomery et al., 2004; O'Brien & Toms, 2010), or a *state* as representations of affect, cognition, and motivation (Kim & Bae, 2018; Berka et al., 2007).

Theories and frameworks of engagement

Given the different definitions, multiple theories have been employed to operationalize engagement in relation to cognition, affect, and motivation. Engagement has been framed as conscious efforts in processes involving cognition, such as sustained attention (Matthews et al., 2010) and cognitive effort (Sharek & Wiebe, 2015). The theory of flow (Csikszentmihalyi, 1990) emphasizes the balance between task demands and a person's skills in achieving a certain level of engagement. On the other hand, engagement has been associated with more subjective states, such as the willingness to seek out an appropriate level of cognitive demand (Sharek & Wiebe, 2015). The strength of engagement has also been considered a significant aspect of motivation-related experiences. (Seddon et al., 2008; Higgins, 2006).

More generalized frameworks have also been proposed to capture the multi-dimensionality of engagement. For example, the popular User Engagement Model developed by O'Brien and Toms (2010) employed theories of flow, play, and aesthetics to describe engagement as both a dynamic process and a product (outcome of engagement versus disengagement) in computer-based contexts. Similarly, the Engagement Mode Model identified 5 engagement modes characterized on 3 dimensions: evaluation, control, and motivation. Flow experience was described as "the balance between enjoyment/acceptance mode and efficiency/productivity mode propelled by ambition/curiosity" (Montgomery et al., 2004).

Measuring engagement

While there is a general consensus that engagement is a dynamic process, the amount of engagement—or level of engagement—has often been studied as a discrete state corresponding to overall task demands, manipulated by varying task conditions such as difficulty levels (Köles et al., 2015; Sharek & Wiebe, 2011). Although this approach is effective for validating measures against manipulated conditions such as self-report measures (Keller et al., 2011) and behavioral measures (Sinha et al., 2015), it provides limited insights into the dynamic nature of engagement (Doherty & Doherty, 2019). On the other hand, when studies attempt to capture engagement continuously, adopted measures tend to rely on the more objective but indirect measures of engagement, such as the use of physiological metrics to characterize flow states (Sinha et al., 2015), behavioral responses in primary-secondary task paradigms (Sharek & Wiebe, 2015), and changes in performance over time (Lora Ariza et al., 2017). These continuous measures are primarily associated with the cognitive dimension of engagement and are unlikely to differentiate the motivational and other affective factors of engagement as a subjective experience.

Study objectives

As the first step to a larger project that seeks to assess the multiple dimensions of engagement on a continuous basis, this

paper reports an online study that aimed to validate the use of self-reported scales in assessing momentary engagement experiences. Our goal was to explore to what extent known scales can help capture dynamic changes in engagement over time. To this end, we manipulated the duration and difficulty level of an online Tetris game. Participants, assigned to a specific difficulty level and game duration, were probed about their level of engagement of the current moment during and at the end of the gaming session.

A secondary goal of this study was to examine the relationship between engagement, workload, and subjective states, such as stress. The current study thus employed the User Experience Scales and the short version of Dundee Stress State Questionnaire (SSSQ) (Helton, 2004) to study the multi-dimensionality of user engagement in the given Tetris game context. The well-known Dundee Stress State Questionnaire has identified Task Engagement, along with Distress and Worry, as a high-order factor that is the culmination of mood, motivation, and cognition (Matthews et al., 2002). The manipulation of difficulty level and game duration may act as potential stressors that could affect participants' subjective states, including engagement (Matthews et al., 2013). Finally, as mentioned in the last section, workload has been treated as a crucial factor in task engagement and is thus important to assess in this study.

METHOD

Participants

A total of 220 participants were recruited using Mechanical Turk (Mturk); however, data analysis reported in this paper was conducted on data from only 81 participants due to various issues with data collection. Participants who did not follow task instructions in providing keystroke responses (see Measures) were removed (N=123); data from second attempts were removed for those who re-entered the task upon completion (N=5); and outliers in game performance and response times based on statistical analysis were also removed (N=11). The remaining participants (Male=63, Female=18) ranged from 18 to 60 years old with the majority at 21-40 years old (76.5%). Upon completion of the study, participants received $2 through their Mturk account as compensation.

Experimental design and setup

This experiment followed a 3 (difficulty levels: easy, medium, difficult) x 3 (duration: 4 mins, 8 mins, 12 mins) between-subject design. Difficulty corresponds to the rate at which shapes "fall"—easy: every 1.2s; medium: 0.8s; difficult: 0.4s. Participants were recruited from Mturk (https://www.mturk.com) and entered the task through a link from Jatos, an online platform where the experiment was set up. After completing a web-based consent form, an instruction presentation, and a 2 mins warm-up session, participants were randomly assigned into one of the nine (3x3) conditions of Tetris. Tetris had been used in other studies to elicit an engaging experience (Ewing et al., 2016; Keller et al., 2011). Our version was modified to allow the game to continue regardless of the participant's performance: 3 top and bottom rows automatically disappear whenever the piled shapes reach the top of the screen. Given our goal to examine the subjective experiences associated with engagement, not performance, participants were instructed to enjoy the game and that their performance would not impact their participation in this study.

Measures

Two subjective questionnaires were administered: 24-item SSSQ (administered pre-task and post-task) and 18-item UES (post-task only). Items were rated on a 5-point Likert scale with 5 being strongly agree and 1 strongly disagree. Participants were asked to consider their current feelings at the time they were queried with these scales, i.e., before starting the game and at the end of the game. The factors of usability and aesthetics in the UES were not included as they were not part of the focus of this study. The perceived workload was also collected through a 5-point scale—Instantaneous Self-Assessment (ISA) (Jordan, 1992)—to assess its relationship with subjective engagement. We also attempted to capture self-report engagement during the game. Every 30 seconds, a pop-up question asked participants to rate their level of engagement on a 5- point Likert scale. Participants were asked to first hit the spacebar to signal that they have received the prompt, which resulted in response times to the queries, and then verbally provide a number response on engagement level, which were captured and uploaded to the server as audio files. However, verbal responses (self-reported engagement level during the game) were not analyzed due to the large number of missing audio files and incomplete or poor quality of recordings.

RESULTS

Keystroke response time

We analyzed means, response rates, and standard deviation (SD) of participants' keystroke response times in 3 (Difficulty) x3 (Duration) between-subject ANOVAs. SD was the only measure significantly affected by Duration, $F(2, 72) = 18.028$, $p < .001$, with the SDs of the 4 mins group significantly lower than those found in the 8 mins group (adjusted $p = .0018$) and in the 12 mins group (adjusted $p < .001$).

Survey: Comparisons of factor structure

SSSQ was previously constructed and extensively validated in stressful performance contexts (Helton, 2004; Matthews et al., 2013). For this study, due to the context of interest was not the particular performance pursuit but the overall engaging state, structural differences may exist between the context of the present study and previous studies. As for UES, previous studies found a different factor structure where Endurability, Novelty, and Involvement merged under the context of game-play (Wiebe et al., 2014) and information retrieval (O'Brien & Toms, 2013). To see if any structural differences exist, factor analysis was performed on both scales using principal axis factoring with oblimin rotation. Scree plot

and Parallel analysis were used to determine the number of factors (Horn, 1965) ensuring all extracted factors' eigenvalue of real data was higher than the eigenvalue for random data. The cutoff value for factor loadings was 0.45 (Comrey & Lee, 1992, as cited in Tabachnick et al., 2019).

A four-factor solution was formed for ***SSSQ***, which explained 59% of the variance (see Table 1). The current structure largely conforms with the original three-factor except that three items referring to performance self-efficacy departed from the original structure. A three-factor solution was settled for ***UES*** (see Table 2), which explained 61% of the variance in total. items from Novelty, Endurability, and Involvement are merged into one factor, which is consistent with previous studies (Wiebe et al., 2014; O'Brien & Toms, 2013). However, three items related to success/plan accomplishment/rewarding separated from Endurability and formed a new factor, and the item "My game experience was rewarding." was cross-loaded on two factors.

Sensitivity of scales to task conditions

As shown above, the current structures are mostly consistent with the originals. We herein ran analysis following the original structure (SSSQ: Distress, Engagement, and Worry, UES: Novelty, Focused attention, Involvement, and Endurability). The 3x3 between-subject ANOVAs were conducted separately for pre-task SSSQ as a baseline, and pre-post changes (z-score) were examined followed by post hoc analysis with Bonferroni correction.

Table 1. Structure comparison of SSSQ

Original factors	SSSQ items	1	2	3	4
Distress	I felt dissatisfied.	**0.57**	0.04	0.26	-0.35
	I felt depressed.	**0.75**	-0.17	0.14	0.3
	I felt sad.	**0.69**	0.02	0.18	0.14
	I felt impatient.	**0.54**	-0.01	0.13	-0.23
	I felt annoyed.	**0.87**	0.12	-0.06	-0.12
	I felt angry.	**0.8**	0.06	0.03	-0.02
	I felt irritated.	**0.83**	0.06	-0.1	-0.24
	I felt grouchy.	**0.87**	-0.17	-0.07	0.12
Engagement	I felt alert.	-0.08	**0.67**	0.1	-0.15
	I felt active.	-0.3	0.41	0.24	-0.09
	I was committed to attaining my performance goals.	-0.15	**0.75**	0.13	-0.03
	I wanted to succeed on the task.	0.08	**0.86**	0.06	0.12
	I was motivated to do the task.	0.03	**0.96**	0	0.08
	I felt confident about my abilities.	-0.22	0.31	-0.09	**0.59**
	I expected to perform proficiently on this task.	0.08	**0.46**	-0.08	**0.51**
	Generally, I felt in control of things.	-0.23	0.33	-0.03	**0.59**
Worry	I was trying to figure myself out.	0.01	-0.12	**0.61**	0.3
	I was reflecting about myself.	-0.08	-0.19	**0.6**	**0.52**
	I was daydreaming about myself.	-0.03	-0.36	**0.34**	0.32
	I felt self-conscious.	0.14	0.03	**0.57**	0.04
	I was worried about what other people would think of me.	0.37	-0.02	**0.5**	0
	I felt concerned about the impression I was making.	0.07	0.1	**0.73**	-0.17
	I was thinking about how others have done on this task.	0.07	0.23	**0.74**	-0.11
	I was thinking about how I would feel if I would be told how I performed.	-0.14	0.16	**0.72**	-0.1
	Proportion variance	0.22	0.15	0.14	0.08
	Cumulative variance	0.22	0.37	0.51	0.59

Note, items with correlation >.45 are bold

Table 2. Structure comparison of UES

Original factors	UES items	1	2	3
Novelty	The content of the task incited my curiosity.	**0.65**	0.15	0.15
	I would continue to game out of curiosity.	**0.58**	0.14	0.28
	I felt interested in my gaming task.	**0.85**	-0.02	-0.03
Focused attention	I forgot my immediate surroundings while doing task.	-0.01	**0.65**	0.06
	I was so involved that I ignored everything around me.	0.4	0.43	-0.31
	I lost myself in the experience.	0.02	**0.78**	0.07
	I was so involved thus I lost track of time.	0.1	**0.71**	-0.11
	The time I spent just slipped away.	0.03	**0.77**	0.08
	I was absorbed in my task.	**0.67**	0.25	-0.26
	During this experience I let myself go.	-0.06	**0.64**	0.18
Involvement	I was really drawn into my task.	**0.69**	0.24	-0.02
	I felt involved in this task.	**0.82**	0.11	-0.14
	This gaming experience was fun.	**0.73**	0.03	0.14
Endurability	Finishing this task was worthwhile.	**0.9**	-0.05	0.01
	I considered my gaming experience a success.	0.2	0.09	**0.73**
	This gaming experience did not work out the way I had planned.	0.06	-0.18	**-0.67**
	My gaming experience was rewarding.	**0.52**	-0.02	**0.49**
	I would recommend this game to my friends and family.	**0.78**	-0.23	0.23
	Proportion variance	0.33	0.18	0.1
	Cumulative variance	0.33	0.51	0.61

Note, items with correlation >.45 are bold.

Pre-task score. None of the conditions showed a significant effect on pre-task scores, which indicated that the baseline was not likely a driver of the findings in score changes.

Score changes. Mean scores for subscales of SSSQ and UES were first calculated following the original scale structure. Then, standardized SSSQ change scores were calculated for three scales in SSSQ using the formula, (Post-score – Pre-score) / σ of the Pre-scores (Helton, 2004). There was a significant main effect for difficulty on *Distress change*, F (2, 72) = 4.667, p = .012, and post hoc analysis indicated that significantly smaller changes of Distress in the easy condition (M = 0.02, SD = 0.86) than both the medium condition (M = 0.84, SD = 1.34), adjusted p = .034, and the hard condition (M = 0.81, SD = 1.08), adjusted p = .039. A marginal effect of task duration was also found in Distress change, F (2,72) = 2.859, p = .064. For *Engagement change,* a significant effect of difficulty, F (2, 72) = 3.316, p = .042, and a marginal effect of the interaction between duration and difficulty, F (4, 72) = 2.195, p = .078, were observed. Post hoc analysis showed, in the 8-min group, a significant difference from the easy condition (M = 0.82, SD = 0.62) to the medium condition (M = -0.36, SD = 0.82), adjusted p = .019, and to hard condition (M = -0.66, SD = 0.84), adjusted p = .002. When the game was easy, the 8-min player (M = 0.82, SD = 0.62) had Engagement increases at significantly larger magnitude than the decrease in the 4 mins group (M = -0.13, SD = 0.95), adjusted p = 0.047. *Worry changes* significantly with difficulty, F (2, 72) = 3.551, p = .034. Post hoc analysis indicated that people had a significantly difference in change on Worry in the medium condition (M = 0.27, SD = 0.66) than the easy condition (M = -0.21, SD = 0.66), adjusted p = .03.

Relationship between post-SSSQ and UES

Correlations were performed on the post-task SSSQ and UES scales to examine their relations (see Table 3).

Table 3. Correlations of SSSQ and UES

	ISA	SSSQ		UES				
N=81	1	2	3	4	5	6	7	8
Perceived-Workload	1							
Engagement	-0.1	1						
Worry	0.07	0.15	1					
Distress	0.18	-0.38***	0.23*	1				
Novelty	-0.06	0.77***	0.2	-0.22	1			
Focused-attention	0.17	0.52***	0.06	-0.04	0.56***	1		
Involvement	-0.04	0.73***	0.13	-0.24*	0.82***	0.66***	1	
Endurability	-0.22	0.68***	0.26*	-0.26*	0.71***	0.38***	0.74***	1

Note, *p < .05.** p < .01.*** p < .001

Results indicated the Engagement in SSSQ was significantly correlated with all subscales in UES. Worry was negatively correlated with Endurability, r (81) = .26, p = .032, and Distress were negatively correlated with Involvement r (81) = -.24, p = .031, and Endurability r (81) = -.26, p = .017. Among the SSSQ subscales, Engagement was negatively correlated with Distress r (81) = -.38, p < .001, and Worry was positively correlated with Distress r (81) = .23, p =.04. Significant intercorrelations were also found among all subscales of the UES. Reliabilities for all subscales were beyond 0.8 (see Table 4).

Table 4. Descriptives of SSSQ and UES.

Subscale	N (# Item)	M	SD	Cronbach's α
Engagement	8	4.21	0.68	0.92
Worry	8	2.79	0.93	0.84
Distress	8	2.28	1.03	0.83
Novelty	5	3.93	0.98	0.82
Focused attention	7	3.92	0.86	0.87
Involvement	3	4.27	0.88	0.83
Endurability	3	3.68	0.68	0.88

DISCUSSION

Results showed that changes of SSSQ from pre- to post-task were overall sensitive to difficulty conditions, suggesting the effectiveness of SSSQ to assess task demand-induced changes in subjective states. Engagement and Worry scores changed (from pre- to post-task) in the opposite direction in all difficulty conditions, a finding that is consistent with their respective definitions—Engagement reflects the level of focus on the task while Worry reflects the level of focus on personal concerns (Matthews et al., 2002). It is thus reasonable to observe that as engagement level increases, worry decreases, and vice versa. However, distinct patterns for Engagement and Worry were observed across different conditions. Compared to pre-task scores, Engagement decreased most at the hard condition, while the largest increase of worry state was at the medium condition. It is possible that the hard condition of the game generated a demand that was indeed a poor match for the participants' skill level, thus diminishing their focus/engagement, while sufficient enough to prevent more time for personal reflections (Matthews et al., 2006, 2013). This suggests that the association between Engagement and Worry may detangle under some circumstances.

There was no significant change in Distress in the easy condition, but Distress increased substantially in both the medium and hard conditions. Given that Distress represents low hedonic tone (i.e., unhappiness) and low confidence-control (Matthews et al., 2002), this result suggested, in the context of a video game, the higher task demand could influence affect and perceptions about self-efficacy, even when no external motivation is provided to perform well.

On the other hand, our manipulation of game duration did not have much impact on the self-reported scores. There was a marginal effect of duration on Distress (p = .06) and a marginal interaction between duration and difficulty on Engagement (p = .078). Post-hoc analysis of this interaction seems to suggest more substantial changes in Engagement between easy and medium/hard conditions for the 8 minutes group. In other words, if engagement in Tetris is treated as a dynamic construct, higher levels of engagement may be more likely to happen not too early nor too late into the game and would depend on the level of game demands.

Regarding UES, none of the subscales were sensitive to difficulty or duration. This might be due to the items of the scale being designed to assess overall feelings of interest and focus on the task, i.e. Tetris, thus not effective for assessing momentary feelings as was intended.

Relationships between engagement (UES), subjective states of stress (SSSQ), and workload (ISA) were also explored (Table 3). Not surprisingly, Engagement from SSSQ was significantly associated with all subscales of UES—Novelty, Focused attention, Involvement, and Endurability. Worry and Distress were also both associated with Endurability, which is about the perception of success in a task and willingness to return or recommend the task to others (O'Brien & Toms, 2010). Interestingly, perceived workload was not correlated with any subscales, nor was it a function of difficulty or duration levels in this study. As task demands impacted self-reported engagement but not perceived workload, one should be careful in using self-reported workload as a proxy measure for task engagement.

Overall, the SSSQ appears to be a more sensitive measure of momentary feelings to the present task duration and difficulty under the gameplay context, compared to UES, but we have also observed significant correlations among the SSSQ and UES factors. Future studies to explore dynamic changes in engagement may investigate how best to assess the different dimensions of engagement by incorporating the factors stipulated in the SSSQ and to conduct triangulations with physiological data and objective measures such as performance and response times. While we made attempts to include some objective measures in this study, we experienced many limitations associated with an online study. A large portion of recruited participants did not provide keystroke responses as instructed, and it was hard to differentiate technical issues from participation issues related to data collection quality. An experimental platform with better control is necessary for future studies to be able to provide valid real-time data to more closely investigate participants' ongoing experiences about task engagement.

REFERENCES

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). *EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks*. *78*(5), 14.

Chapman, P. M. (1997). Models of engagement: Intrinsically motivated interaction with multimedia learning software (Doctoral dissertation, University of Waterloo). (n.d.). *Models of engagement: Intrinsically motivated interaction with multimedia learning software (Doctoral dissertation, University of Waterloo)*.

Comfrey, A., & Lee, H. (1992). Interpretation and application of factor analytic results. *Comrey AL, Lee HB. A First Course in Factor Analysis*, *2*, 1992.

Csikszentmihalyi, M. (1990). Flow. *The Psychology of Optimal Experience. New York (HarperPerennial) 1990.*

Doherty, K., & Doherty, G. (2019). Engagement in HCI: Conception, Theory, and Measurement. *ACM Computing Surveys*, *51*(5), 1–39.

Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop. *Frontiers in Human Neuroscience*, *10*, 223.

Helton, W. S. (2004). Validation of a Short Stress State Questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238–1242.

Higgins, E. T. (2006). Value from hedonic experience and engagement. *Psychological Review*, *113*(3), 439–460.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Jordan, C. (1992). Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. *Report No. DRA/TM (CAD5)/92011. Farnborough: Defence Evaluation & Research Agency*.

Keller, J., Ringelhan, S., & Blomann, F. (2011). Does skills–demands compatibility result in intrinsic motivation? Experimental test of a basic notion proposed in the theory of flow-experiences. *The Journal of Positive Psychology*, *6*(5), 408–417.

Kim, S., & Baek, T. H. (2018). Examining the antecedents and consequences of mobile app engagement. *Telematics and Informatics*, *35*(1), 148–158.

Köles, M., Szegletes, L., & Forstner, B. (2015). Towards a physiology-based difficulty control system for serious games. *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 323–328.

Laurel, B. (1991). *Computers as theatre reading. Mas: Addison-Wesley Publishing Company.*

Lora Ariza, D. S., Sánchez-Ruiz, A. A., & González-Calero, P. A. (2017). Time Series and Case-Based Reasoning for an Intelligent Tetris Game. In D. W. Aha & J. Lieber (Eds.), *Case-Based Reasoning Research and Development* (Vol. 10339, pp. 185–199). Springer International Publishing.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., Grier, R., & Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*(4), 315–340.

Matthews, G., Emo, A., Funke, G., Zeidner, M., Roberts, R., Costa, P., & Schulze, R. (2006). Emotional Intelligence, Personality, and Task-Induced Stress. *Journal of Experimental Psychology. Applied*, *12*, 96–107.

Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). Profiling task stress with the dundee stress state questionnaire. Psychology of stress: New research, 1, 49-90.

Matthews, G., Warm, J. S., Reinerman-Jones, L. E., Langheim, L. K., Washburn, D. A., & Tripp, L. (2010). Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. *Journal of Experimental Psychology: Applied*, *16*(2), 187–203.

Montgomery, H., Sharafi, P., & Hedman, L. R. (2004). Engaging in Activities Involving Information Technology: Dimensions, Modes, and Flow. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(2), 334–348.

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*(1), 50–69.

O'Brien, H. L., & Toms, E. G. (2013). Examining the generalizability of the User Engagement Scale (UES) in exploratory search. *Information Processing & Management*, *49*(5), 1092–1107.

Seddon, K., Skinner, N. C., & Postlethwaite, K. C. (2008). Creating a model to examine motivation for sustained engagement in online communities. *Education and Information Technologies*, *13*(1), 17–34.

Sharek, D., & Wiebe, E. (2011). Using Flow Theory to Design Video Games as Experimental Stimuli. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1520–1524.

Sharek, D., & Wiebe, E. (2015). Investigating Real-time Predictors of Engagement: Implications for Adaptive Videogames and Online Training. *International Journal of Gaming and Computer-Mediated Simulations*, *7*(1), 20–37.

Sinha, A., Gavas, R., Chatterjee, D., Das, R., & Sinharay, A. (2015). Dynamic assessment of learners' mental state for an improved learning experience. *2015 IEEE Frontiers in Education Conference (FIE)*, 1–9.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (Seventh edition). Pearson.

Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, *32*, 123–132.