Assessment of Flow, Immersion & Engagement on Game-Based Strategies for Kinesic Cue Detection

Alexander Miranda, Jonathan Hurter, Kayla Coca, Kassidy James, and Crystal S. Maraj
Institute for Simulation and Training,
University of Central Florida, Orlando, USA
{amiranda, jhurter, kcoca, kjames, cmaraj}@ist.ucf.edu

ABSTRACT

Military training may benefit from using game-based strategies to aid in instruction. This study evaluated Game-Based Training (GBT) by investigating a control condition and two feedback conditions: leaderboard and excessive positive feedback. Leaderboards produce competition using a ranking system. Excessive positive feedback praises correct behaviors. The control condition lacked a feedback strategy. Specifically, the GBT task involved detecting kinesic cues in a virtual simulation. Detecting kinesic cues, or non-verbal gestures that convey one's true emotions, help Soldiers interpret their environment. The kinesic cues trained in this study were aggressiveness and nervousness. The research investigated correlations between involvement and performance per each strategy; as well as correlations between game usage and involvement. A Kruskal-Wallis test was also conducted to assess differences in involvement between strategies. Involvement was operationalized as flow, immersion, and engagement surveys based on Likert scales. Performance comprised detection accuracy and response time in the kinesic cue detection task. Significant correlations were found between some select flow subscales and performance; and video game usage and the flow subscale of concentration on task. The Kruskal-Wallis test revealed the flow subscale of unambiguous feedback was highest in the excessive positive feedback condition. The research findings offer design considerations relevant to select flow states and their relation to performance. Unexpected flow outcomes suggest extrinsic motivation may have played a part in increasing performance within the excessive positive feedback condition. Additional considerations for GBT may be explored for educational applications. This research lays the foundation for defining which flow subscales map to specific feedback elements. As a limitation, the immersion and engagement questionnaires may require modification to increase sensitivity.

ABOUT THE AUTHORS

Alexander Miranda is a Research Assistant at the University of Central Florida (UCF) Institute for Simulation and Training (IST). He is a fourth-year undergraduate student studying psychology at UCF. Currently, Alex also holds another research position at UCF, where he is investigating the potential prosocial benefits of meditation. Alex has devoted the majority of his education to the study and practice of meditation. He credits this passion to his traveling experiences in Nepal during the summer of 2016. Alex plans to continue his education in the field of psychology by attending graduate school and extending his research.

Jonathan Hurter is a Research Technician at UCF IST. He earned his Master's degree in Modeling & Simulation, as well as a Certificate in Instructional Design for Simulations, from UCF. His main focus is on instructional design and technical writing. As a writer, researcher, and experimenter, he has contributed to IST efforts in experiential training (including virtual simulation-based training and game-based training), avatar science, and malodor studies. He has also supported instructional design efforts at Science Applications International Corporation (SAIC).

Kayla Coca is a Research Assistant at the Institute for Simulation and Training (IST) at the University of Central Florida, where she focuses on assessing Simulation-Based Training primarily for military applications. She is a third-year undergraduate student studying psychology at UCF. She also works at the Marriage and Family Research Institute (MFRI), where her research focus is on assessing the value of a relationship education program.

Kassidy James is a Research Assistant for the Institute for Simulation and Training (IST) at the University of Central Florida (UCF). She graduated at the top of her class with her Bachelor's degree in Psychology from UCF in addition to a certificate in Intelligence and National Security. James' work at IST focuses on assessing and improving the efficacy of simulation-based training primarily for various military applications. James will be

returning to UCF to earn her Master's degree in School Counseling through the Counselor Education graduate program. In her graduate career, she hopes to continue her research efforts in the areas of suicidal behavior and non-suicidal self-harm.

Crystal S. Maraj is a Research Associate employed by the Institute for Simulation and Training (IST) at the University of Central Florida. Concurrently, she also works for the Simulation & Training Technology Center (STTC) as a researcher on medical simulation technology projects. She has attained her Bachelor's degree in Psychology, as well as her M.S. and Ph.D. in Modeling and Simulation (M&S) from UCF. Dr. Maraj's focus on developing and implementing empirically-based research experiments to assess the effectiveness of evolving Simulation-Based Training (SBT) platforms as well as provide data-driven SBT recommendations to the training and acquisition communities. Current project focus on training effectiveness evaluation of virtual world platforms for military training; comparing usability platforms to offer scientific recommendations, and assessing the technical feasibility of using alternative simulations for training military medical personnel. Finally, Dr. Maraj publish research findings to inform the scientific and training communities to improve trainee performance and training system utility.

Assessment of Flow, Immersion & Engagement on Game-Based Strategies for Kinesic Cue Detection

Alexander Miranda, Jonathan Hurter, Kayla Coca, Kassidy James, and Crystal S. Maraj
The Institute for Simulation and Training at the University of Central Florida
Orlando, Florida
{amiranda, jhurter, kcoca, kjames, cmaraj}@ist.ucf.edu

INTRODUCTION

Game-Based Training (GBT) is the incorporation of game mechanics (e.g., points, badges, and ranking systems) into an activity that attempts to teach a skill or change user behavior (Meng & Khe Foon, 2016). GBT exposes trainees to real-world scenarios in a Virtual Environment (VE) while being cost-effective and safe (Szczepkowski, Santarelli, Stagl, Glenn, & Paulus, 2011). One of the most important aspects of games is their ability to increase engagement in learners (Oblinger, 2004). Researchers use engagement, along with flow and immersion, as psychological constructs to measure the degree of an individual's involvement in a task. Despite the prevalence of these constructs, the relationships between them, video game usage, and performance are inconsistent and unclear.

Recently, the Military has increased the use of GBT to train and test a wide variety of important skills (see Yildirim, 2010 for an in-depth review). GBT can be used to train military tactics and strategies (e.g., deception, bat and bleed, and boxing maneuver), train skills (e.g., rifle handling and use of robots), and increase traits the military deems essential (e.g., teamwork and responsibility; Yildirim, 2010). Training can be based on the role a particular individual will hold, and the specific situations they are likely to encounter (Yildirim, 2010). In addition to military training, GBT is also used in various other domains (e.g., higher education, medical training, police training, and business training).

This GBT study examined a kinesic cue detection task. Kinesics is the study of how people communicate through body motions, such as gestures and facial expressions (Birdwhistell, 1970). Kinesic detection skills are of great importance to Military Soldiers, who need to make quick assessments of their surroundings in the battlefield in order to identify potential threats. This task tested individual's recognition of two aggressive (i.e., clenched fist and slap hands) and two nervous kinesic cues (i.e., check six and wring hands). This study examines how performance on a kinesic cue detection task using different GBT strategies is related to three measures of involvement: flow, immersion, and engagement. The two GBT strategies assessed were leaderboard and excessive positive feedback; these will be described in the following paragraphs.

Game-Based Training (GBT) Strategies

Leaderboards

In GBT, leaderboards incorporate the game design element of a ranking system that promotes competition through social comparison and goal-directed behaviors (Chernbumroong, Sureephong, & Muangmoon, 2017; Sailer, Hense, Mayr, & Mandl, 2017; Hanues & Fox, 2015). Social comparison involves comparing one's performance to other participants, by presenting scores from highest to lowest, to promote a competitive environment (Hanues & Fox, 2015; Sailer et al., 2017; Landers & Landers, 2014). Competition is both a product of social comparison and a means for motivation (Garcia, Tor, Gonzalez, 2006; Landers & Landers, 2014). Past research demonstrates the effectiveness of leaderboards on improving performance and establishes leaderboards as a viable training strategy (Landers & Landers, 2014).

Excessive Positive Feedback

The excessive positive feedback strategy employed in this study has two primary dimensions: excessive frequency and positive sign. Frequency refers to how often the feedback occurs (Johnson, 2016). In contrast, sign refers to the given type of criticism: either positive praise of goal attainment, or negative responses for lack of goal attainment (Rivera, 2011). This experiment only used positive sign for feedback. Previous research on frequency and sign show varying results: although some evidence indicates the ineffectiveness of praise on performance (Hattie & Timperley, 2007), there appears to be no significant effect for sign (Rivera, 2011; Thurlings, Vermeulen, Kreijns, Bastiaens, & Stijnen, 2012). Previous research has also shown negative effects of frequency on performance (Lurie & Swaminathan, 2009; Ackerman & Gross, 2010). However, the present integration of feedback and sign dimensions may produce a unique interaction effect. In this condition, participants received praise for every correct behavior detected in the VE during training. Participants also received praise for a sequence of three correct behaviors detected. The praise was administered as a text message at the top of participants' screens; for example, for every single cue detected, one would be told, "Correct Cue Identified. Good Job, Keep Going."

Involvement

Flow

Flow is an optimal state of consciousness where performance is heightened (Csikszentmihalyi, 1990). Flow is measured by nine dimensions (see Table 1; Jackson, Martin, & Eklund, 2008). Kiili (2006) separated the nine dimensions into two categories: flow antecedents and flow characteristics. The five antecedents that contribute to flow experiences are challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, and sense of control. The four characteristics that describe flow experiences are concentration on the task at hand, loss of self-consciousness, the transformation of time, and autotelic experiences.

Table 1. Operational Definitions for the Nine Dimensions of Flow

Term	**Definition**
Challenge-skill balance	How competent a participant believed they were in a high-demand situation
Action-awareness merging	How aware a participant was of the decision-making process
Clear goals	How aware the participant was of what they were trying to accomplish
Unambiguous feedback	How much a participant felt like they knew how they were doing
Sense of control	How in control of a situation a participant felt
Concentration on task at hand	How focused a participant felt
Loss of self-consciousness	How much a participant considered the impressions they made on others
Transformation of time	How much a participant felt like time passed differently during the training
Autotelic Experiences	How rewarding the training was for a participant

Past GBT research indicates that flow is related to various metrics, including performance, learning outcomes, enjoyment, and perceived learning (For further reviews, see Perttula, Kiili, Lindstedt, & Tuomi, 2017; Pavlas, Heyne, Bedwell, Lazzara, & Salas, 2010; Barzilai & Blau, 2014). This is congruent with Csikszentmihalyi's (1988) notion that higher flow scores correlate with improved performance. However, other studies have shown conflicting results indicating no significant relationships between flow experiences and learning (Kiili & Lainema, 2008). The flow experience is also thought to be tied directly to intrinsic motivation (Rodriguez-Ardura & Meseger-Artola, 2016). Given the inconsistent findings, this research seeks to clarify the relationship between flow and performance using different GBT strategies.

Immersion

Despite immersion's prevalence in GBT literature, researchers have yet to accept a straightforward definition of the term. Murray (1997) described immersion as being completely engulfed in an environment to the point where an individual is no longer aware of the external world. Brown and Cairns (2004) classified three levels of immersion from lowest to highest: engagement, engrossment, and total immersion. This research, however, defines immersion as how absorbed individuals are during gameplay; it is a seamless experience where individuals are temporarily relieved from self-referential thoughts (Jennett et al., 2008). Both immersion and flow involve a strong union with the present moment, but immersion differs from flow in that it is not contingent on skill level (Jennett et al., 2008). In the literature, the relationship between immersion and performance in GBT is unclear. Previous studies indicate a positive relationship between immersion and performance on GBT tasks, as well as immersion and learning outcomes (Cheng, She, & Annetta, 2015; Meng-Tzu, Yu-Wen, & Hsiao-Ching, 2017), whereas others indicate no effect on learning (Hamari, Shernoff, Rowe, Coller, & Edwards, 2016). One possible explanation for these inconsistent findings is that the researchers used different measures for immersion. Given the contrasting results, this research aims to investigate the relationship between immersion and performance using different GBT strategies.

Engagement
Engagement is a psychological construct used to measure an individual's willingness to continue an activity (McMahan, 2003). To be engaged is an effortful, volitional experience where the individual feels that a task is worthwhile. Using the Game Immersion Questionnaire (GIQ), Meng-Tzu et al. (2017) found that engagement was crucial for producing learning outcomes in GBT. Engagement is also linked to positive attitudes towards learning, such as enthusiasm, curiosity, and interest (Skinner & Belmont, 1993). Various game mechanics such as badges, points, and leaderboards have been shown to significantly improve student engagement (Meng & Khe Foon, 2016). On the continuum of involvement, engagement occurs at the lowest level, immersion occurs at a level higher, and flow marks the highest level of involvement. This research seeks to examine the role of engagement in the leaderboard and excessive positive feedback strategies as well as a control group.

Purpose

Flow, immersion, and engagement have been inconsistently linked to motivation, performance, and learning outcomes. This research seeks to answer a question: how is performance on a kinesic cue detection task using different game-based strategies related to flow, immersion, and engagement? By answering this question researchers can better understand the GBT strategies.

METHOD

Participants

This study recruited 91 participants from the University of Central Florida and its surrounding communities. The inclusion/exclusion criteria for participants included normal or corrected-to-normal vision, a lack of prior participation in behavior-cue detection studies, and US citizenship. Participants were also required to be between 18 and 40 years of age. The control group consisted of 16 females and 14 males, between the ages of 18 and 29 (M = 21.40, SD = 2.31). The excessive positive feedback group consisted of 12 females and 18 males, between the ages of 18 and 39 (M = 21.17, SD = 4.58). The leaderboard group consisted of 16 females and 15 males, between the ages of 18 and 37 (M =21.84, SD = 3.94).

Experimental Design

A between–subjects design was used with one independent variable: feedback type. There were two levels of feedback: leaderboard and excessive positive feedback. In the leaderboard group, participants saw a leaderboard that compared individuals' performance scores to computer generated scores using a three-letter identifier both before and after the scenario vignettes. In the excessive positive feedback group, participants received a positive comment after every correct detection, as well as when they were on a streak of three correct detections (i.e., correct identifications). There was also a control group where participants did not receive any feedback. The dependent variables were flow, immersion, engagement, post-test detection accuracy, and post-test response time.

Metrics

Objective
This study collected objective data on post-test detection accuracies and response times. Post-test detection accuracy was represented as a percentage of kinesic cues correctly detected out of the total number of target cues. Post-test response time was measured as the length of time in milliseconds between when a cue animation was initialized and when participants clicked on the respective cue event on screen, before classifying if the cue was either aggressive or nervous.

The demographics questionnaire gathered basic information about each participant, including participants' age, sex, handedness, education, familiarity with various technological components, and hours per week spent playing videogames.

Subjective

Flow was measured using the Flow State Short Scale (Jackson, Martin, & Eklund, 2008). Each of the nine questions assessed a different flow subscale using a Likert scale that ranged from one to five, where one was strongly disagree and five was strongly agree. The flow subscales were challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration on task at hand, sense of control, loss of self-consciousness, transformation of time, and autotelic experience.

Immersion was measured using an eight-item questionnaire, where participants rated each item on a five-point scale, where one was strongly disagree and five was strongly agree (Jennett et al., 2008).

Engagement was measured using a seven-item questionnaire, where participants rated each item on a five-point, where one was strongly disagree and five was strongly agree (Charlton & Danforth, 2005).

Testbed

The Virtual Battlespace 2 (VBS2) Version 2.0 development software was used for this study (VBS2 was the standard for military simulation training at the time of data collection). The software allows instructional designers to create various simulations in VEs. VBS2 offers a high level of customizability and extensive feedback capabilities, such as mission playback, and presents multiple camera angles to view the VE (Green, Leibrecht, Fite, 2011).

Procedure

Upon arrival, the experimenter greeted each participant, and confirmed his or her eligibility based on the inclusion/exclusion criteria. The experimenter then administered the Ishihara Color Blindness assessment. Each participant was then provided with the informed consent document to read and sign. The experimenter administered the demographics questionnaire to the participant, and then randomly assigned participants to one of three conditions: excessive positive feedback, leaderboard, or control. Then, the experimenter presented the interface training using PowerPoint, which allowed the individual to become familiar with the software and practice the classification task. The interface training required participants to identify colored barrels and click on them to become familiarized with the VE. If a participant did not meet the proficiency requirements (i.e., at least nine points), the individual was given a second attempt. If the participant failed a second time, the individual was dismissed. Each participant was then shown the pre-test training slides in PowerPoint and completed the pre-test scenario in VBS2. During this scenario participants were required to detect whether virtual characters displayed nervous or aggressive cues, based on previous knowledge. If a character did portray these cues, participants had to click on the character and then classify if the cue was nervous or aggressive. After this, the participant viewed the kinesics training PowerPoint presentation, which covered the purpose of behavioral cue detection and the types of target behavior cues. Next, each participant completed the scenario vignettes by experimental group. The scenario vignettes were similar to the pre-test but included the GBT strategies (see Figures 1 and 2). Afterwards, the experimenter provided the Flow State Short Scale, Immersion Measure, and Engagement Measure questionnaires. The experimenter then showed a post-test training PowerPoint presentation to introduce the task, and each participant completed the post-test scenario, which was identical to the pre-test. Finally, participants were debriefed and dismissed.

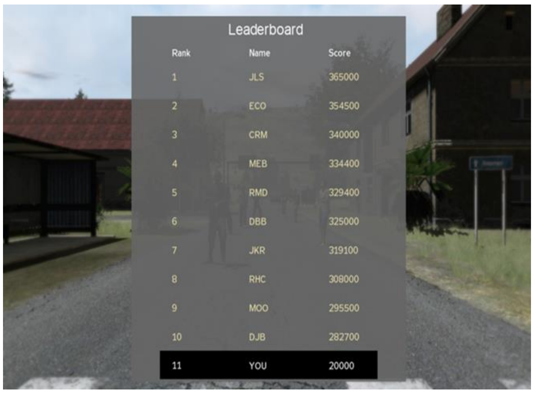

Figure 1. Leaderboard Condition

Figure 2. Excessive Positive Feedback Condition

RESULTS

Biographical Analysis

Demographics were used to inform biographical analysis. A split-file Mann-Whitney U Test was used to look at gender differences on the involvement and performance metrics. Kruskal-Wallis tests were also used to look at differences on the involvement and performance metrics based on comfort with computers and familiarity with a mouse; the latter two were both ranked on a five-point scale. None of these tests showed any significant differences between groups. There were also no significant correlations found for game usage, which was measured in game genre categories.

Preliminary Data Analysis

A Cronbach's alpha test determined the reliability of the experimental questionnaires (i.e., flow, engagement, and immersion) for the leaderboard, excessive positive, and control conditions. The analysis revealed the results shown in Table 2.

Table 2. Cronbach's Alpha Measures

Cronbach's Alpha (α)	
Survey	**Cronbach's Alpha (α)**
Flow	.69
Immersion	.74
Engagement	.79

Leaderboard

The data were split by experimental groups and were analyzed using Spearman rho correlations, to look for relationships between the flow subscales, immersion, engagement, and post-test performance. The leaderboard group had the most significant correlations. There was a moderate negative correlation between challenge-skill balance and response time (see Table 3), which showed that when participants felt the task matched their skill level, they had a faster response time. There was also a moderate negative correlation between concentration on task and response time (see Table 3 and Figure 3), which showed that the more concentrated participants felt, the faster their response time was. There was also a strong positive correlation between concentration on task and detection accuracy (see Table 3 and Figure 4), showing that increased concentration was linked to detection accuracy. Game hours only had one significant correlation, and it was in the leaderboard group only. There was a weak positive correlation between game hours and concentration on task (see Table 3), which showed that in the leaderboard group, the more hours a participant spent playing video games during the week, the better the participant felt they concentrated (see Table 3).

Table 3. Leaderboard Correlations

Spearman's Correlations in Leaderboard Condition			
	1	**2**	**3**
1. Post-Test Response Time	–		
2. Post-Test Detection Accuracy	-.63**	–	
3. Flow – Concentration	-.50**	.60**	–
4. Flow – Challenge-Skill Balance	-.40*	.17	–
5. Game Hours	-.18	-.04	0.36*
*Note *$p < .05$, ** $p < .01$			

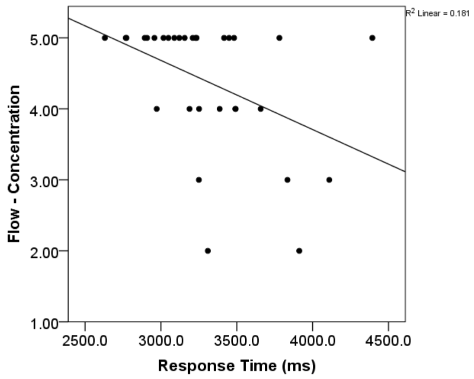

Figure 3. Response Time Scatterplot

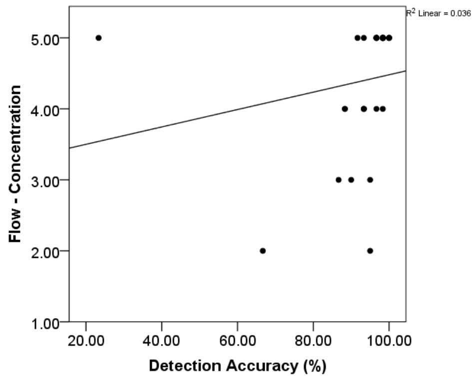

Figure 4. Detection Accuracy Scatterplot

Excessive Positive Feedback

In the excessive positive feedback group, there was a moderate negative correlation between autotelic experience and detection accuracy (see Table 4), which shows that the more intrinsically enjoyable participants found the task, the worse their detection accuracy was.

Table 4. Excessive Positive Feedback Correlations

Spearman's Correlations in Excessive Positive Feedback Condition		
	1	**2**
1. Post-Test Response Time	–	
2. Post-Test Detection Accuracy	-.43*	–
3. Flow – Autotelic Experience	.01	-.43*
*Note *$p < .05$		

Control

In the control group, there was one significant correlation between clear goals and detection accuracy (see Table 5), showing that the more clearly a participant understood the goals of the task, the better a participant's detection accuracy.

Table 5. Control Correlations

Spearman's Correlations in Control Condition		
	1	**2**
1. Post-Test Response Time	–	
2. Post-Test Detection Accuracy	-.63**	–
3. Flow – Clear Goals	-.28	.41*
*Note *$p < .05$, ** $p < .01$		

Cumulative Condition Analysis

Within all three conditions, detection accuracy was negatively correlated with response time (see Tables 3, 4, and 5). Additionally, a Kruskal-Wallis test was conducted for performance, engagement, immersion, and the flow subscales. Performance, engagement, immersion, and seven of the flow subscales were not found to vary significantly across groups. There was a significant effect for flow subscales of unambiguous feedback ($\chi^2(2) = 6.69$, $p < .05$) and sense of control ($\chi^2(2) = 13.45$, $p < .001$).

DISCUSSION

This study investigated the impact of flow, immersion, and engagement on performance to offer recommendations for improving GBT. The findings both support and refute previous theories on the effects of engagement, immersion, and flow. In the following sections, the relationships between the three conditions (leaderboard, excessive positive feedback, and control) and five flow subscales (concentration on task at hand, autotelic experience, unambiguous feedback, sense of control, and clear goals) will be examined.

Leaderboard

Within the leaderboard condition, higher concentration correlated with shorter response time and better detection accuracy. After each training scenario, participants viewed their score at the bottom of a leaderboard. It is possible that comparing their scores to the other scores instilled a motivation to perform better on the task. This suggests the use of leaderboards as a game-based strategy may leverage participant's concentration through competition and social comparison. Therefore, this strategy may be used for training in areas that require intense focus and immediate responses with accurate results (e.g., combat medics, first responders, medical personnel, and air traffic controllers).

Excessive Positive Feedback

Within the excessive positive feedback condition, participants had significantly higher scores on the unambiguous feedback subscale compared to the leaderboard and the control condition. This finding may be due to the sheer amount of feedback received. The leaderboard condition received only a few moments of feedback at the end of their scenarios and the control condition received no feedback; whereas participants in the excessive positive feedback condition were notified every time a target cue was detected (i.e., correctly identified).

Additionally, participants who found the task more intrinsically rewarding (i.e., had an autotelic experience) had lower scores for detection accuracy. This is unexpected considering past research has shown that intrinsic motivation has stronger positive effects on performance compared to extrinsic motivation (see Cerasoli, Nicklin, & Ford, 2014, for further review). According to Csikszentmihalyi (1990), one of the characteristics of an autotelic experience is that individuals do not concern themselves with the outcome of their performance on a task. It is possible that this outcome was facilitated by the exposure focused on positive feedback, which promoted interest in the activity over the outcome. This may have undermined the intrinsic motivation to perform well, resulting in a negative correlation with detection accuracy. This supports past research on the negative effects of excessive feedback (Lurie & Swaminathan, 2009; Ackerman & Gross, 2010). Considering the converse to this correlation, it is also possible that participants who experienced less intrinsic motivation experienced more extrinsic motivation by placing greater value on the praise received. This may have acted as an extrinsic incentive that positively impacted performance in this group.

Control

In the control group, clear goals were correlated with better detection accuracy. One possible explanation is that participants may have been primed more to the task in this condition because no additional feedback was added after the pretest. These results also point to the potential effectiveness of a goal-setting strategy in the absence of feedback strategies. Without any feedback to direct behavior, participants are forced to rely on the goals of the task. Because the participants who felt they did not understand the goals of the task scored lower in detection accuracy, it is imperative that GBT designers instill a clear understanding of the goals for a task in participants. The finding that clear goals was not related to performance in the leaderboard or excessive positive feedback group suggests that a simulation-based approach (here, no feedback strategy) may be effective in fields that require a higher level of fidelity with a realistic environment. For example, in the real world, an air traffic controller will not have a leaderboard to reinforce his or her progress in successfully identifying anomalies on a radar.

Cumulative Condition Analysis

The flow sense of control dimension varied significantly across the three conditions. Participants' sense of control was the highest in the excessive positive feedback condition, second highest in the leaderboard condition, and lowest in the control condition. Therefore, it appears that quantity of feedback may encourage a sense of control. The results also partially support the notion that flow may be positively related to performance, because four of the nine subscales of flow were significantly related to performance (Perttula et al., 2017; Pavlas et al., 2010; Barzilai & Blau, 2014, Kiili & Lainema, 2008). The finding that five subscales (action-awareness merging, loss of self-consciousness, unambiguous feedback, sense of control, and transformation of time) were unrelated to performance supports that not all dimensions need to be present to experience flow (Swann, Keegan, Piggot, & Crust, 2012). It may be the case that some dimensions are more important than others in different contexts: Csikszentmihalyi (1990) noted that an artist's sense of time may be completely dissolved during a flow experience, but other activities such as surgery or marathon running require a heightened sensitivity to the passage of time. Therefore, in some instances it may be impractical to score highly on all nine dimensions. Future research should investigate the importance of each the nine flow dimensions in relation to military GBT tasks.

The research also showed that participants in all three groups had similar immersion and engagement scores. Contrary to previous research that showed the benefits of immersion and engagement, both metrics were unrelated to performance in this study. It is unclear why this finding occurred, as participants in all three groups reported feeling both immersed and engaged; however, the measures had no bearing on overall participant performance in the task. One possible explanation is that the questionnaires may have measured immersion and engagement in relation to the VE rather than the task.

Limitations

There were a few noted limitations in this study. One was an inconsistency in the leaderboard condition. The training slides stated that the leaderboard would be shown after each vignette, however it only appeared before the first vignette and after the last one. This limited the amount of feedback that participants received. Future research should incorporate the leaderboard better by showing it after each vignette, to provide participants with more frequent feedback. The leaderboard also compared participants to computer generated scores, which were much higher than the participant scores, and which may have made participants feel they performed poorly. Additionally, this research had a small sample size, which limited the use of additional analysis beyond correlations. Future research would benefit from increasing the sample size to allow for the use of further analysis, such as regressions. This study also used a shortened version of the immersion and engagement measures. It is possible that using longer and more comprehensive questionnaires may have increased the validity of results; however, due to time constraints and participant workload, longer questionnaires were not feasible. Additionally, the lack of specificity of some of the questionnaires may have compromised construct validity. For example, this study did not measure engagement in accordance to its various subcomponents (i.e., cognitive engagement, behavioral engagement, and affective engagement; Fredricks, Blumenfeld, & Paris, 2004). Future research may want to consider a more comprehensive approach by measuring the various subcomponents of engagement.

Another concern is how GBT may only be teaching trainees how to succeed at the game, as opposed to teaching them the skills they need for the real world. This study attempted to partially address the issue of flow's relation to learning by administering the involvement questionnaires before the non-GBT, post-test assessment. Since the post-test lacked feedback (as it was the same as the pre-test), participants were left to detect and classify cues in a more realistic context. This experimental manipulation revealed that when flow was experienced during GBT, flow was, in general, positively related to performance variables (e.g., detection accuracy) outside of GBT. Here, GBT seems to relate to the way participants performed on a transfer-of-context task outside of the game. Another issue arises with the possibility of users becoming disengaged when GBT strategies are removed. Further, leaderboards may be problematic because participants might be willing to take extra risks (e.g., accepting increased false positive detections), that are potentially catastrophic in real-world jeopardy, to obtain a higher score. A similar problem arises with trainees seeking more positive feedback, in order to be rewarded.

Future Research

In addition to the brief recommendations noted in prior sections, to increase the accuracy of flow, immersion, and engagement scores, future research should consider incorporating objective physiological measurements in addition to the traditional subjective scales. Notable examples would include Jennett et al.'s (2008) use of eye-tracking technology to measure immersion and Tsai, Huang, Hou, Hsu, and Chiou's (2016) use of eye-tracking to measure flow. Furthermore, as neuroscience continues to explore the neurophysiological mechanisms of flow, researchers should devise new ways to assess flow states by looking at brain activity. Future research should also consider measuring motivation, which may have been an extraneous variable in this study. Doing so may provide additional insight into construct relationships, such as the autotelic experience correlation with detection accuracy in the excessive positive feedback condition. There is also a need to further assess the impact of GBT on learning, by investigating both changes in context (e.g., in live simulation or real-world situations) and across timescales. A finer-grain study would also investigate trends across signal detection measures, including false positives. Ultimately, researchers should further examine the flow subscales in relation to performance in GBT. Continuing to map these subscales within different GBT strategies may offer suggestions as to which strategies are most effective for different types of training.

CONCLUSION

In summary, this research looked at how flow, immersion, and engagement correlated with different performance metrics on a kinesic cue detection task using leaderboards and excessive positive feedback. The findings suggest that certain flow subscales, particularly challenge-skill balance, concentration task at hand, autotelic experience, and clear goals, may have an impact on training performance and should be considered by individuals who are designing training programs. Future research should continue to map the flow subscales to various GBT strategies. Here, GBT may be injected into military tasks, including detection of improvised explosive devices (via eyesight or detector tools), maintaining separation in air traffic control, applying medical devices, or understanding topographical map features. Although this paper focused on leaderboards and feedback mechanisms, other GBT strategies exist: making certain information scarce (e.g., forcing attentive search strategies in detection tasks), using time-pressure, testing memory, and creating scores based on chained successes (e.g., expanding a scoreboard, where repeated correct answers create a score multiplier) are potential avenues to investigate.

REFERENCES

Ackerman, D. S. & Gross, B. L. (2010). Instructor Feedback: How Much Do Students Really Want? *Journal of Marketing Education, 32*(2), 172-181.

Barzilai, S. & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education, 70*, 65-79.

Birdwhistell, R.L. (1970). *Kinesics and Context*. University of Pennsylvania Press, Philadelphia.

Brown, E. & Cairns, P. (2004). A grounded investigation of game immersion, *proceedings of the International Conference for Human-Computer Interaction*, (pp. 1279-1300). New York, New York: ACM Press.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014, February 3). Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis. *Psychological Bulletin*. Advance online publication.

Charlton, J., & Danforth, I. (2005). Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior, 23*(3), 1531-1548.

Cheng, M. T., She, H. C., & Annetta, L. A. (2015). Game immersion experience: its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning, 31*(3), 232–253.

Chernbumroong, S., Sureephong, P., & Muangmoon, O. (2017). The Effect of Leaderboard in Different Goal-Setting Levels, *proceedings of the International Conference on Digital Arts, Media and Technology*. Chiang Mai, Thailand: IEEE.

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In *Optimal experience: Psychological studies of flow in consciousness* (pp. 5-35). New York: Cambridge University Press.

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York: Harper-Collins.

Dunlap, W. P., Chen, R., & Greer, T. (1994). Skew reduces test-retest reliability. Journal of Applied Psychology, 79(2), 310-313.

Fredricks, J., Blumenfeld, C., & Paris, A. (2004). School engagement: potential of the evidence. *Review of Educational Research, 74* (1), 59-109.

Garcia, S. M., Tor, A., & Gonzalez, R. (2006). Ranks and Rivals: A theory of Competition. *Personality & Social Psychology Bulletin*, 32, 970-982.

Green, C., Leibrecht, B. Fite, J. (2011). *After Action Review Guide for Trainers of Virtual Battlespace-2 Missions*. United States Army Research Institute for the Behavioral and Social Sciences.

Guan, N. C., Yusoff, M. B., Zainal, N. Z., & Yun, L. W. (2012). Analysis of two independent samples with non-normality using non parametric method, data transformation and bootstrapping method. *International Medical Journal, 19*(3), 218-220.

Hamari, J., Shernoff, D. J., Rowe, E., Coller. B., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170-179.

Hanues, M. D. & Fox, J. (2015). Assessing the Effects of Gamification in the Classroom: A Longitudinal Study on Intrinsic Motivation, Social Comparison, Satisfaction, Effort, and Academic Performance. *Computers & Education*, *80*, 152-161.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81-112.

Jackson, S. A., Martin, A. J., & Eklund, R. C. (2008). Long and short measures of flow: The construct validity of the FSS-2, DFS-2, and new brief counterparts. *Journal of Sport and Exercise Psychology*, *30*(5), 561–587.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human - Computer Studies*, *9*, 641.

Johnson, M. (2016). Feedback Effectiveness in Professional Learning Contexts. *Review of Education*, *4*(2), 195-229.

Kiili, K. (2006). Evaluations of an experiential gaming model. *Human Technology*, *2*(2), 187–201.

Kiili, K., & Lainema, T. (2008). Foundation for measuring engagement in educational games. *Journal of Interactive Learning Research*, *19*(3), 469–488.

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*, *45*(6), 769-785.

Lurie, N. H. & Swaminathan, J. M. (2009). Is Timely Information Always Better? The Effect of Feedback Frequency on Decision Making. *Organizational Behavior and Human Decision Processes*, *108*, 315-329.

McMahan, A. (2003). Immersion, Engagement and Presence: A Method for Analyzing 3-D Videogames. M. J. P. Wolf and B. Perron (Eds.) *The Videogame Theory Reader* (pp. 67-86). London: Routledge.

Meng, T. & Khe Foon, H. (2016). Incorporating meaningful gamification in a blended learning research methods class: examining student learning, engagement, and affective outcomes. *Australasian Journal of Educational Psychology*, *32*(5), 19-34.

Meng-Tzu, C., Yu-Wen, L., Hsiao-Ching, S., & Po-Chih, K. (2017). Is immersion of any value? Whether, and to what extent, game immersion experience during serious gaming affects science learning. *British Journal of Educational Technology*, *48*(2), 246-263.

Murray, J. (1997). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. The MIT Press, Cambridge, MA.

Oblinger, D. (2004). The next generation of educational engagement. *Journal of Interactive Media in Education*, *8*, 1-18.

Pavlas, D., Heyne, K., Bedwell, W., Lazzara, E., & Salas, E. (2010). Game-based learning: the impact of flow state and videogame self-efficacy. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54* (pp. 2398-2402). Orlando, FL.

Perttula, A., Kiili, K., Lindstedt, A., & Tuomi, P. (2017). Flow experience in game based learning - a systematic literature review. *International Journal of Serious Games*, *4*(1), 57-72.

Rivera, I. D. (2011). *Feedback in Videogame-based Adaptive Training* (Technical Report 1287). United States Army Research Institute for the Behavioral and Social Sciences.

Rodríguez-Ardura, I. & Meseguer-Artola, A. (2016). E-learning continuance: The impact of interactivity and the mediating role of imagery, presence and flow. *Information & Management*, *53*, 504-516.

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An Experimental Study of the Effects of Specific Game Design Elements on Psychological Need Satisfaction. *Computers in Human Behavior*, *69*, 371-380.

Swann, C., Keegan, R. J., Piggott, D., & Crust, L. (2012). A systematic review of the experience, occurrence, and controllability of flow states in elite sport. *Psychology of Sport and Exercise*, *13*(6), 807-819.

Szczepkowski, M., Santarelli, T., Stagl, K., Glenn, F., & Paulus, J. (2011).*Virtual environments for soldier training via editable demonstrations* (VESTED). Defense Technical Information Center.

Thurlings, M., Vermeulen, M., Kreijns, K., Bastiaens, T., & Stijnen, S. (2012). Development of the Teacher Feedback Observation Scheme: evaluating the quality of feedback in peer groups. *Journal of Education for Teaching*, 38(2), 193-208.

Tsai, M. J., Huang, L. J., Hou, H. T., Hsu, C. Y., & Chiou, G. L. (2016). Visual behavior, flow and achievement in game-based learning. *Computers & Education*, *98*, 115-129.

Yildirim, S. (2010). Serious game design for military training. *Proceedings of the Games: Design and Research Conference* (pp. 3-4). Volda, Norway, Volda University College.