Human Factors and Simulation, Vol. 83, 2023, 120–130
https://doi.org/10.54941/ahfe1003569

Participant Game Play Experience Predicts Mental Model Formation, Task Performance, and Teaming Behavior in Simulated Search and Rescue

Rhyse Bendell[1], Jessica Williams[1], Stephen M. Fiore[2], and Florian Jentsch[1]

[1]Team Performance Laboratory, Institute for Simulation and Training, University of Central Florida, Orlando, FL 32826, United States
[2]Cognitive Sciences Laboratory, Institute for Simulation and Training, University of Central Florida, Orlando, FL 32826, United States

ABSTRACT

Video gaming experience has been found to impact behavior and performance on experimental tasks, can influence cognitive processes, and may even transfer to tasking proficiency. The purpose of the investigations reported here was to examine the relationships between video game experience and mental model formation as well as experience and gameplay behaviors in the context of a game-based urban search and rescue mission. We hypothesized that differences in video game play experience would influence the formation of mental models, and that experience would also be associated with different behavioral tendencies during tasking. Results of Study 1 support our hypothesis that greater video game experience was associated with more convergent mental models related to the game-based experimental task. Results of Study 2 indicate that participants with greater experience showed evidence of better overall performance and more strategic behavior. These findings suggest that video gaming experience impacts both the formation of task-related mental models as well as task performance and teaming behaviors.

Keywords: Individual differences, Psychometrics, Teaming, Human-autonomy teaming, Simulated task environment

INTRODUCTION

Video gaming experience has been found to impact behavior and performance on experimental tasks (Keebler et al., 2014; Sanchez & Langer, 2020), can influence cognitive processes (Litern & Boot, 2021), and may even transfer to tasking proficiency (Yang & Chen, 2020). Research efforts to shed light on the influences of video game experience are ongoing and have produced mixed results (Yang & Chen, 2020). However, the belief that individual differences may account for variations in performance, learning, and reactions, when encountering game-like tasks or methods is consistent throughout the domain (Sanchez & Langer, 2020). Understanding how video game experience transfers to tasks both inside and outside of the laboratory is a research challenge that grows in relevance as the use of video games for entertainment as well as experimental purposes, game-based assessment, training, and education expands (e.g., Corral et al. 2021). A fundamental challenge is identifying theoretically valid differences in facets of video game experience that may impact a participants' performance on experimental tasks. Further, what is also lacking is a standardized approach to capturing that experience. Previously, we have reported efforts aimed towards establishing shared methods for scoping aspects of video game experience that may be relevant to accounting for variance in performance (Bendell et al., 2021a). Additionally, we have demonstrated the utility of accounting for experiential elements such as video game experience in distinguishing the mental model formation tendencies (Bendell et al., 2021b) and task performance patterns of participants (Williams et al., 2021). Here, we build on that foundation of work by applying those conceptual approaches to the investigation of mental model formation and execution performance of a video game-based and virtually administered urban search and rescue task.

PURPOSE

We study how video game experience can alter knowledge structures associated with game-based experimental tasks and the potential for those differences in conceptualization to transfer to differences in behavior and performance. The purpose was to examine the relationships between video game experience and the mental models, as well as gameplay behaviors associated with a Minecraft-based team task simulating urban search and rescue operations. Our primary hypotheses were as follows: first, differences in video game play experience would influence the formation of mental models related to the game-based experimental task, such that participants with greater experience would demonstrate more convergent mental models, and, second, that experience would be associated with different behavioral tendencies when participants performed the gamified task, such that participants with greater experience would demonstrate more strategic decision-making and better performance overall. To test our hypotheses, we first conducted an investigation to evaluate the relationship between video game play experience and mental model formation and second drew on data collected under DARPA's Artificial Social Intelligence Supporting Teams program to examine the influence of experience on game play behaviors.

Gamed-Based Experimental Task: Minecraft Urban Search and Rescue

The game-based experimental task referenced throughout this report was developed in service of DARPA's Artificial Social Intelligence for Successful Teams (ASIST) program (DARPA, 2019). The structure of the task was inspired by urban search and rescue (USAR) operations and involves teams of players performing vital roles to locate and rescue victims of a building collapse. To facilitate data collection and maintain safety, the task was designed to be experienced and completed in a gamified virtual environment that was developed based on the foundational platform provided by Minecraft

(Mojang, 2015). The sandbox nature of Minecraft provided a serviceable framework for ASIST developers to design and instantiate a simulated incident site within which participants control avatars to execute actions in service of the primary goal of rescuing victims. The version of the task used in the studies reported here required teams of three participants performing three distinct roles: Medical Specialist, Engineering Specialist, and Search Specialist. Each role had unique capabilities, and interdependencies were intentionally designed into the task such that no one participant could effectively succeed at the mission on their own. Medical Specialists, for example, were given the unique skill of healing victims using their first aid kits; however, Engineering Specialists were separately provided a sledgehammer with which they could clear debris which allowed Medics to access victims. Search Specialists were given a stretcher that allowed them to pick up and move victims. To promote teamwork, each role was designed to move at a different speed: Searchers moved fastest to facilitate the transportation of victims, Engineers moved slowest to incur a cost of travel to debris piles, and Medics moved at a middling speed. Additionally, the task featured two distinct victim types: regular victims which could be healed by the Medic without assistance from other team members (provided those victims were not trapped by debris piles), and critical victims that required the presence of the whole team to first stabilize the victim followed by healing using the Medic's first aid kits. Gamification of the experimental task was accomplished by allocating points for healing victims, and teamwork was promoted both by allotting five times as many points for critical victim rescues as for regular victim rescues, and by displaying the team's overall score on a persistent scoreboard. The version of the Minecraft urban search and rescue task employed in the studies reported here allowed participants to change their role (that is, switch between Medic, Engineer, and Searcher) by returning to a home base and swapping their primary tool. This role switching component of the task was vital for allowing teams to balance their spatial and temporal distribution of unique skills to meet the challenges presented during rescue missions. Each team completed a hands-on tutorial prior to executing two different 10-minute rescue missions.

Measuring Video Game Experience

There is currently no standard approach to measuring the video game play experience of a participant (Bendell et al., 2021a; Williams et al., 2021). Methods tend to differ based on researchers' needs, so measures range from short – often single item – scales focused on length of general exposure to thorough batteries aimed at capturing nuances of motivation, play style preferences, and proficiency (often self-reported, though only occasionally measured). Here, we have employed a video game experience measure that blends general video game experience queries with items that specifically target experience related to our gamified experimental task. Our experience measure has been refined over the course of multiple studies and in collaboration with industry professionals (including from partners in the program, e.g., IHMC and Gallup) to focus on three primary facets of an individual's gaming experience: length/duration of exposure, intensity/frequency of exposure, and self-perceived proficiency/skill. By tapping both general and task specific experience, we seek to capture a holistic picture of participants' experience without the need for extensive interactive batteries. A full version of the refined video game experience measure can be reviewed at the ASIST study 2 project's online open science framework record (Huang et al., 2021); however, examples of length/duration queries include "General: Years using a computer to play video games" "Specific: Years playing Minecraft (any expansion or version)" intensity/frequency include "General: Please indicate how regularly you: Play video games which require participation in a team" and "Specific: For the following, please indicate how regularly you: Play Minecraft" and perceived skill/proficiency includes "Specific: Indicate the level of mastery you have over the following skills as compared to other video gamers that you know or have played with: Maintaining an awareness of game/task parameters (e.g., time limits, point goals, etc.)." The measure administered to participants in the studies described here contained 34 items in study 1 and 24 items in Study 2 due to reduction and refinement of items. Only items contained in both versions were employed in the calculation of experience used in our analyses.

STUDY 1

This study was conducted fully online and did not require participants to perform the Minecraft-based urban search and rescue task. Rather, they reviewed the training materials associated with that task, watched gameplay footage of an experienced team completing the gamified task, and responded to a set of surveys and mental model elicitation measures. The primary goal of this study was to determine the relationship between participants' characteristics such as personality, video game experience, sense of direction, etc. and their conceptualization of the Minecraft-based experimental task.

Methods

Participants. 95 undergraduate students from a large, South-eastern University participated in Study 1 in exchange for course credit. Being conducted online, the study was vulnerable to malingering; however, we employed three malingering checks that led to the removal of 32 participants. The resulting sample of 63 participants (29 male, 34 female) was analyzed to compute video game experience measure scores and associated median split. Of those 63, only 47 provided complete responses to the mental model elicitation measure reported here. The resulting set of 47 consisted of 24 female and 23 male participants, all between the ages of 18–24 years, except one who was between 25 and 30 years old.

Procedure and Materials. The procedures and materials used in this study were reviewed and approved by our Institutional Review Board. All materials and measures employed in this study were administered through Qualtrics and were accessed by participants through our university's SONA Psychology Research Participation System. Participants reviewed an informed consent document, indicated that they fully understood and agreed to volunteer, then completed a demographics questionnaire, the extra short form of the Big Five

Personality Index, the Santa Barbara Sense of Direction questionnaire, and our video game experience (VGE) measure. After completion of those surveys, participants reviewed the training slides for the Minecraft USAR task. After reviewing the slides, participants were requested to watch 7 minutes and 22 seconds of gameplay footage that had been captured during an experienced team's completion of a Minecraft USAR mission. Participants then completed a brief set of questions related to the events that occurred in the recorded mission. After, they completed two mental model elicitation measures: one related to the skills used by the recorded team members, and the second related to the components and structure of the Minecraft USAR task. Finally, participants completed a motivation to play online games measure.

Mental Model Elicitation: Minecraft USAR Task. The mental model elicitation measure employed to capture participants' conceptualization of the gamified experimental task required participants to sort 29 "cards" into up to 10 piles, and to label those piles. Participants could sort cards into as few as 1 pile or as many as 10, but without duplication. The 29 items available for participants to group included objects vital to the completion of the urban search and rescue task such as "stretcher" and "first aid kit" in addition to concepts such as "find victims," "locate objectives," or "partnering" as well as mission monitoring elements such as "scoreboard," "countdown timer," or "rescue points."

Analysis

Video Game Experience. Participants' scores on the video game experience measure employed in this study were calculated based on responses to 11 items which examined general length/duration of video game play, general frequency/intensity of video game play, Minecraft-specific length/duration of game play, and Minecraft-specific frequency/intensity of game play. Scores were calculated by normalizing response quantities to a 100-points scale and averaging across the sum of all items. Resulting scores for the 63 datasets remaining after malingering check removals ranged from 51.8 to 77.5, with an average of 64.9 and standard deviation of 7.4. The median split cutoff for the set was 66.9, and, after removal of datasets that did not contain complete responses to the mental model card sort task, we were left with a total of 47 participants, 27 categorized as high in video game experience, and 20 categorized as low in video game experience.

Mental Model Card Sort. Analysis of responses to the card sort task was accomplished by examining the pairing of cards by each participant. Each possible card pair was assigned a value of zero (0) if the participant had not paired those cards (that is, had not placed them in the same group) or a value of one (1) if they had paired them. The resulting pair response sets were sorted into two groups: one matrix representing participants rated high in video game experience, and a second matrix representing participants rated low in video game experience. The cross-participant response correlations for the low and high experience matrices were calculated, and the resulting correlations for each pair were compared between the two video game experience groups to determine relative convergence of mental models for each group. Additionally, we examined the discrepancy in average correlation for each response pair by taking the difference between the average correlation of the high experience group responses and the low experience group responses. Pair responses associated with a correlational difference greater than 0.2 were further analyzed to determine whether convergence differed significantly between low and high video game experience groups.

Results

Results of a one-way ANOVA conducted to examine the differences in means between the matrix of pair response correlations drawn from the video game experience groups provide evidence of significantly greater convergence between participants with high video game experience, M_LowVGE = 0.098 (SD = 0.115), M_HighVGE = 0.228 (SD = 0.153), $F(1, 539) = 105.319$, $p < .001$. We interpret this to indicate that participants who scored higher on the video game experience measure demonstrated much greater similarity in the way in which they conceptualized the Minecraft USAR task.

Next, we explored the differences in specific card pairings to determine if participants who scored higher on the video game experience measure not only had more convergent mental models, but also demonstrated different conceptualization of particular aspects of the game-based task. Our exploratory analysis revealed that the points on which the two groups differed most notably were related to the duties of the Searcher role and the criticality of monitoring and coordinating with one's team. We are not reporting the full set here due to space limitations; however, some notable 'X – Y" card pairs" that were found to be significantly different include "locate objectives – partnering, M_LowVGE = 0.050 (SD = 0.224), M_HighVGE = 0.296 (SD = 0.465), $F(1, 45) = 4.707$, $p = .034$, "your location – teammate locations," M_LowVGE = 0.500 (SD = 0.513), M_HighVGE = 0.926 (SD = 0.267), $F(1, 45) = 13.609$, $p <. 001$, "rescue victims – access victims," M_LowVGE = 0.350 (SD = 0.489), M_HighVGE = 0.667 (SD = 0.480), $F(1, 45) = 4.914$, $p = .032$, and "locate objectives – client map," M_LowVGE = 0.100 (SD = 0.308), M_HighVGE = 0.370 (SD = 0.492), $F(1, 45) = 4.668$, $p = .036$. Further, high video game experience participants also differed from low experience participants in their conceptualization of how to track their team's progress as well as how to utilize the tools at their disposal to achieve a high rescue score. Examples of these differences include significantly more convergent mental models for high experience participants regarding "rescue points – scoreboard," M_LowVGE = 0.450 (SD = 0.510), M_HighVGE = 0.778 (SD = 0.424), $F(1, 45) = 5.776$, $p = .020$, and "scoreboard – countdown timer," M_LowVGE = 0.550 (SD = 0.510), M_HighVGE = 0.889 (SD = 0.320), $F(1, 45) = 7.796$, $p = .008$.

These results indicate that video game experience influenced participants' conceptualizations of the experimental task, but do not provide insight into the influence of experience on task behaviors and proficiency. The task execution data collected in Study 2 provided an avenue to extend the above findings and determine whether differences in experience and task conceptualization, impact behavior.

STUDY 2

The data analyzed here was collected through the ASIST project's second human subjects research experiment and is one part of a much larger set. The study incorporated the administration of multiple survey measures spanning demographics, experience, personality traits, social intelligence, and more in addition to training teams of three participants to complete virtual, simulated urban search and rescue missions as well as testing those teams' performances in two separate missions (see "Shared Methods" above). Given our focus on the explanatory and predictive value of capturing participant video game experience, we attend primarily to the video game experience measure outcomes and individual participant task performance metrics.

Methods

Participants. The full ASIST study was conducted with 67 teams of three individuals (201 participants); however, the datasets from three teams were removed due to errors in experiment administration, leaving 192 participants in the remaining sets. Missing data due to incomplete experience measures or errors in performance metric extraction yielded 174 participants for the analyses reported here. The sample was predominantly male, 128 males and 46 females, and young, minimum 18 years old and maximum 49 years, with an average of 21.9 years

Procedure and Materials. A full description of the procedures and materials employed in this study may be found at the ASIST study 2 project's online open science framework record (Huang et al., 2021). Briefly, the components of the study relevant to understanding the extension of the above findings are that participants completed a refined version of the video game experience measure from study 1, were trained on a game-based team urban search and rescue mission, and, unlike study 1, performed two missions as a member of a three-person team. The refined version of the video game experience measure contained all 11 items used in the earlier study 1 version as well as an additional 13 items (for a total of 24) though only 4 of those were employed for capturing video game experience as reported here (for a total of 15 as compared to the 11 in study 1). Participants completed the video game experience as part of a set of surveys during the first phase of the study. Following the completion of those surveys participants underwent task training including hands-on training and tests of comprehension and competency. Then teams of participants were tasked with completing two, 10-minute Minecraft USAR missions (see brief on mission characteristics in "Game-based Experimental Task: Minecraft USAR" above).

Analysis

Video game experience. Scores on the video game experience measure were calculated using the same approach as study 1 with the exception of including the four new items. Resulting scores ranged from 37.53 to 95.13 with an average of 76.16. The grouping method required for investigating mental model convergence in study 1 was not necessary here, so we did not separate participants using a median split and instead utilized their experience scores to conduct GLM analyses.

Task execution behaviors. A full description of the metrics that we extracted related to task execution behaviors as well as detailed hypotheses grounded in task elements can be reviewed at our university's ASIST study 2 open science framework project preregistration page (Bendell, 2021). Here, we analyzed only performance data from participants' second missions because we found evidence of medium-to-strong learning effects between first and second mission executions. Due to space limitations, we focus only on performance of one of the team support roles: Searcher (see Searcher role brief in "Game-based Experimental Task: Minecraft Urban Search and Rescue" above). As it was possible to never perform the Searcher role, only the 90 participants who did were included in this analysis. To examine participants' behaviors, we measured the quantity of role specific tasks they successfully completed (victim transports, all), calculated role performance efficiency (role efficiency, all), and delved into the particulars of task execution (role efficiency, interdependent & victim transports supporting accessibility) to determine whether participants were performing relatively more in the service of their team.

Results

Role performance: team service agnostic. We first examined the relationship between participants' video game experience and their success at completing one of the primary objectives of the Searcher role: overall transportation of victims to make them accessible for subsequent rescue. Outcomes indicated that as participant video game experience increases, so too does their number of overall transportations, $F(1, 88) = 7.731$, $R^2 = 0.080$, $p = .007$. Next we analyzed participants' efficiency scores and found that not only did participants with higher video game experience execute more transportations but they did so more efficiently, $F(1, 88) = 8.457$, $R^2 = 0.077$, $p = .005$.

Role performance: team service focused. Outcomes of a linear regression showed that participants with higher video game experience scores had a greater tendency to focus on victim transportations that did not require assistance from team members and supported them by making victims more accessible for rescue (that is, by the Medic role), $F(1, 88) = 5.644$, $R^2 = 0.060$, $p = .020$. Analysis of role efficiency associated with team interdependency-related victims further showed that participants with greater video game experience scores displayed better efficiency in executing their transportations, $F(1, 88) = 6.523$, $R^2 = 0.069$, $p = .012$

DISCUSSION

The purpose of the investigations reported here was to examine the relationships between video game experience and mental model formation as well as experience and gameplay behaviors in the context of a game-based urban search and rescue mission. Our primary hypotheses were that differences in video game play experience would influence the formation of mental models, and that experience would be associated with different behavioral tendencies when participants performed missions. Results of Study 1 support our hypothesis that greater video game experience was associated with more convergent mental models related to the game-based experimental task. Results of Study 2 indicate that participants with greater experience showed evidence of better overall performance and more strategic behavior with respect to providing service to their team. One critical takeaway from the results of these studies is that some aspects of video game experience may transfer to novel task performance.

Having found evidence of transfer here is particularly informative because the Minecraft-based USAR task was only based in Minecraft in the sense that the game provided the sandbox foundation for development and involves almost no task elements that appear in standard Minecraft survival or other modes. The video game experience measure, on the other hand, tapped general as well as Minecraft specific experience with respect to duration, frequency/intensity, and self-reported skill.

Besides the importance of accounting for variance in participant task-related behaviors, other practical considerations associated with video game experience may warrant attention from human subjects researchers. Motivation and participation biases are key among these. Readers may note that the samples in study 1 and 2 demonstrated different demographic and experience profiles such that study 2 was completed primarily by males (74% to 48% in study 1) and further by individuals who indicated greater video game experience (76.2 to 64.9 average score). We speculate that this may be a result of the studies' recruitment approaches considering that study 1 was conducted over several months and presented as "complete surveys to help us develop better AI" whereas study 2 had to be accomplished rapidly and relied on participants having some video gaming proficiency so it was presented as "come play Minecraft for money." It seems that employing a video game-based experimental task or recruiting by appealing to video game players may incur the side-effect of artificially boosting the confounding power of participants' video game experience by skewing the sample. Experience may also color participants' approaches to study measures as well as willingness to complete. We noted that all of the participants who did not provide complete responses to the mental model card sort measure in study 1 went on to faithfully complete the remaining study measures, and the majority of those participants (12 out of 16) were classified as low in video game experience. This may indicate that even measures that are not explicitly video game-based may be perceived differently by participants or be more readily completed by those of particular experience profiles.

This work can be used to further research in cognitive engineering along a number of lines. From the standpoint of research in synthetic task environments designed for experimentation mimicking complex operational environments, we provide a method for understanding variability in performance characteristics. Accounting for this kind of variance allows researchers to better understand performance differences more likely due to experimental research questions than video game experience. Second, from the standpoint of understanding performer characteristics, these kinds of measures can be used to create profiles of players (Bendell et al., 2021a). These profiles can be used to study experience-treatment interactions, the infusion of technology into some operational setting that might be related to VGE, or even be used to inform artificial intelligence how to monitor and interact with teammates (Bendell et al., 2021a; Williams et al., 2022). For example, imbuing an AI assistant with the understanding that a particular teammate tends to be individualistic and may avoid coordinating closely with the rest of their team would allow an autonomous agent to address that gap by either encouraging coordination at the team-level or directly working with the individual to ensure they are aware of the benefits and drawbacks of their choices. Further, AI teammates may be well positioned to maintain awareness of disparities in cognitive abilities and individual skills, and provide insight into optimal task distribution as well as appropriate distribution of information across human team members.

Lastly, use of the VGE can help guide researcher decisions whenever gamified experimental tasks are employed. Consideration of requisite task training, appropriate sample selection, and interpretation of performance results may be informed first by awareness of relevant experience factors and second by measurement of those factors to account for the variability introduced by the growing prevalence of video gaming experience.

ACKNOWLEDGMENT

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. W911NF-20-1-0008. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA or the authors' affiliated University.

REFERENCES

Bendell, R. (2021, June 18). UCF TA2 – ASIST Study 2 Hypotheses Preregistration. https://doi.org/10.17605/OSF. IO/8FPRV

Bendell, R., Williams, J., Fiore, S. M., & Jentsch, F. (2021a). Supporting Social Interactions In Human-Ai Teams: Profiling Human Teammates From Sparse Data. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 665–669). Sage CA: Los Angeles, CA: SAGE Publications.

Bendell, R., Williams, J., Fiore, S. M., & Jentsch, F. (2021b). Towards Artificial Social Intelligence: Inherent Features, Individual Differences, Mental Models, and Theory of Mind. In *International Conference on Applied Human Factors and Ergonomics* (pp. 20–28). Springer, Cham.

Corral, C. C., Tatapudi, K. S., Buchanan, V., Huang, L., & Cooke, N. J. (2021). Building a synthetic task environment to support artificial social intelligence research. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp.660–664). Los Angeles, CA: SAGE Publications.

DARPA (2019). Artificial Social Intelligence for Successful Teams (ASIST). Retrieved from https://www.darpa.mil/program/artificial-social-intelligence-for-successful-teams

Huang, L., Freeman, J., Cooke, N., Dubrow, S., Colonna-Romano, J., Wood, M. D., … Caufman, S. J. (2021, June 7). ASIST Study 2 June 2021 Exercises for Artificial Social Intelligence in Minecraft Search and Rescue for Teams. https://doi.org/10.17605/OSF. IO/GXPQ5

Lintern, G., & Boot, W. R. (2021). Cognitive training: transfer beyond the laboratory?. Human Factors, 63(3), 531–547.

Keebler, J. R., Jentsch, F., & Schuster, D. (2014). The effects of video game experience and active stereoscopy on performance in combat identification tasks. *Human factors, 56*(8), 1482–1496.

Mojang. (2015). Minecraft. Retrieved from https://www.minecraft.net

Sanchez, D. R., & Langer, M. (2020). Video game pursuit (VGPu) scale development: Designing and validating a scale with implications for game-based learning and assessment. *Simulation* & *Gaming, 51*(1), 55-86.

Williams, J., Bendell, R., Fiore, S. M., & Jentsch, F. (2021, September). Towards a Conceptual Framework of Comprehensive Video Game Player Profiles: Player Models, Mental Models, and Behavior Models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 807–811). Sage CA: Los Angeles, CA: SAGE Publications.

Yang, J. C., & Chen, S. Y. (2020). An investigation of game behavior in the context of digital game-based learning: an individual difference perspective. *Computers in Human Behavior, 112*, 106432.