Gender differences in achievement, behavior, and STEM interest among learners using Minecraft

Maricel A. Esclamado [1,2*] and Maria Mercedes T. Rodrigo [1]

*Correspondence:
maricel.esclamado@obf.ateneo.edu
[1] Ateneo de Manila University,
Loyola Heights, Quezon City,
Philippines
[2] University of Science and
Technology of Southern Philippines,
Cagayan de Oro City, Philippines
Full list of author information is
available at the end of the article

Abstract

In this paper, we analyze how male and female learners differ in their in-game behaviors, knowledge assessment outcomes, and STEM interest using the What-If Hypothetical Implementations using Minecraft (WHIMC). We also investigate how male and female learners' self-reported levels of frustration and boredom relate to these outcomes. We examine in-game data, out-of-game assessment data, self-reported frustration and boredom, and results of the STEM interest questionnaire (SIQ) from 175 Grade 8 learners from a school in the Philippines. We found that male learners tend to explore more than female learners. Both genders learn more through more exploration, making more observations, and completing tasks, but only female learners benefit from having more idle time. Male learners had a higher increase in STEM interest compared to female learners, while there is no significant difference in assessment scores between the genders. This study also found that boredom and frustration have a negative impact on academic outcomes, behavior, and STEM interest, especially among female learners. Bored female learners tend to do less well on post-game assessments, explore less, and make fewer observations. Frustrated female learners tend to not complete the task. Female learners who expressed frustration or boredom tend to have a decreased interest in SIQ in the Interest category after playing WHIMC.

Keywords: Minecraft, WHIMC, Gender difference, Frustration, Boredom, Behaviors, STEM interest, Philippines

Introduction

Minecraft is an open-ended, sandbox-type video game first launched in 2009 (Bitner, 2021). It is open-ended in that it gives players freedom in exploring the environment and players can also mine resources and craft or build objects, often in collaboration with other players. Having reached 140 million active players in 2021, it is one of the most popular games of

all time (Sinclair, 2021). It is no surprise therefore that Minecraft has found its way into education as a platform to demonstrate concepts that might otherwise be abstract.

Open-ended games like Minecraft typically do not enforce a learning sequence. Learners have the freedom to choose what tasks to complete and in what sequence. Among the many challenges in using a game like Minecraft as a learning platform is that of assessment. How can we determine whether students are learning and what they have learned? How can we assess students' experience of the learning process? How do individual differences such as gender influence both what was learned and the learning experience?

Gender difference is a significant factor affecting learning outcomes in game-based learning (Paraskeva et al. 2010). Ünlüsoy et al. (2010) found that males outperformed females in learning abilities in game-based learning. Peters et al. (2021) found that male learners were more likely to have better ability scores in game-based learning particularly in Minecraft than female learners. Other previous studies found that female students outperformed male students in achievement and engagement (Khan et al., 2017; Lukosch et al., 2017). In addition, the gender differences become more significant when considering the role of affect, because females and males have different patterns of affective states and behavior in game-based learning which may also affect their STEM interest (Chiu, 2020). Glynn et al. (2006) found that there is a stronger relationship between affect and both achievement and participation in advanced STEM studies for female students than for male students. Chiu (2020) found that female learners were more likely to study STEM when they experienced less boredom and more off-task behavior, and male learners were more likely to study STEM when they felt less frustration and were more concentrated.

Prior work that investigated gender differences in learners' behaviors and outcomes in game-based learning including Minecraft are limited to behaviors such as time spent on gameplay and being off-task and these studies have only used out-of-game data in identifying learners' behavior. In addition, there is little work on analyzing gender differences in Minecraft.

In this study, we investigate gender differences in in-game behaviors of students in Minecraft, an open-ended, less-structured game. We also investigate gender differences in knowledge assessment outcomes and STEM interest of the learners. We also investigate how male and female learners' self-reported levels of frustration and boredom relate to some of these outcomes. Our specific research questions are as follows:

RQ1:   How do the in-game behaviors of male and female learners differ?

RQ2:   How do the assessment outcomes of male and female learners differ?

RQ3:   How does the change in STEM interest of male and female learners differ after using WHIMC?

RQ4:   What is the relationship between the in-game behaviors and assessment outcomes of male and female learners?

RQ5:   What is the relationship between the learners' self-reported frustration and boredom and their assessment outcomes of male and female learners?

RQ6:   How do in-game behaviors of the male and female learners who self-reported frustration and boredom differ from those who did not?

RQ7:   How does the change in STEM interest of the male and female learners who self-reported frustration and boredom differ from those who did not?

Related literature

Game-based learning refers to the use of digital games for educational purposes. Educators leverage on games' ability to strike a balance between teaching/learning and fun gaming is a characteristic leveraged by educators to support the learning process (Prensky, 2003). Games used for education can act as rich primers for engaged learning and have been reported as effective in improving learning outcomes (Yu, 2019). Game-based learning environments have become popular learning tools because of their potential to deliver knowledge while keeping the experience fun for the learners (All et al., 2016). Playing educational games generates positive moods which then enhances interest, resulting in improved learning performance (Kang et al., 2017).

Minecraft

One category of game-based learning is the exploratory learning environment. This refers to software or platforms that provide learners with the freedom to discover and experiment with concepts and ideas on their own, within the context of the game (Mavrikis et al., 2016). One type of exploratory learning environment presents learners with a virtual world or simulation and encourages learners to interact with elements of this world to achieve goals, solve problems, or deepen their understanding of otherwise abstract concepts. Learners are actively engaged in the learning process which motivates them to learn. Minecraft is an example of an exploratory learning environment. As mentioned in the introduction, Minecraft is a sandbox-type game that provides players autonomy on how to do the tasks in the learning environment (Felecia, 2020). Over the last decade, Minecraft has been utilized by educators as a tool to teach subjects like math, science, social science, and language (Baek et al., 2020; Felecia, 2020; Manahan & Rodrigo, 2022) that can help students develop critical thinking skills, problem-solving abilities, and creativity while also making learning fun and interactive.

WHIMC

What-If Hypothetical Implementations in Minecraft (WHIMC; https://whimcproject.web.illinois.edu/) is a set of simulations that learners can explore in order to learn more about science, technology, engineering, and mathematics (STEM)

(WHIMC, n.d). It is composed of a Rocket Launch Facility on Earth, Lunar Base LeGuin, a Space Station, alternate versions of Earth, exoplanets, and others. The Rocket Launch Facility is modeled after that of NASA. Here, learners can visit mission control and Mars rover test sites as well as talk to simulated NASA scientists. Lunar Base LeGuin is a tutorial level in which learners practice using in-game science tools to measure and record temperature, oxygen, pressure, and wind speed. The Space Station includes an unaltered version of Earth in which learners practice making different kinds of observations. The Space Station is also the jump-off hub from which learners travel to different worlds.

WHIMC utilizes Minecraft Java Edition as an interactive learning environment for students to explore the alternate versions of Earth using "what-if?" scenarios. For example, what if Earth had no moon (See Figure 1)? There would be no seasons, days would be shorter, and winds would be stronger. What if Earth was tilted by 90 degrees, that is, with the North Pole pointed towards the sun? There would be very different day/night cycles. What if Earth had a slightly colder sun? Water might only be able to exist in liquid form in a limited strip of green, and this is where we would all be forced to live. The alternate versions of Earth present learners with opportunities to observe the planet under altered conditions. Although the worlds are fictional, they are created in consultation with scientists. They accurately depict conditions on Earth under these circumstances.

In each of these alternate Earths, learners explore the terrain, describe the environment, report observations about how life on Earth is affected by these circumstances, and possibly create habitats that will enable them to survive. By immersing learners in these activities, WHIMC hopes to generate interest in and excitement for STEM among participating learners. Lane et al. (2022) investigated data from learners who participated in a summer camp that incorporated WHIMC and found that the learners' interest in STEM was triggered by their learning experiences in the game.

Fig. 1 Earth with No Moon

Assessing learning experiences in game-based learning

One of the challenges in using exploratory learning environments is that of assessment. Game environments such as Minecraft allow students to employ a wide range of approaches to achieve goals. Hence, a single problem may have multiple solutions and multiple correct answers, if solutions and right answers exist at all.

Minecraft activities are typically evaluated through assessment methods such as pre- and post-tests (Tangkui & Keong, 2021), self-reports (Callaghan, 2016; Melián Díaz et al., 2020), and observations by teachers or researchers (Callaghan, 2016). Some researchers suggested the use of game analytics, that analyzes in-game data, to offer insights into the learning process (Horn et al., 2016). In-game data can provide insight regarding students' achievement, behavior, emotional responses to the game, and STEM interest.

In the context of WHIMC specifically, previous studies have attempted to assess students' learning experiences using in-game and out-of-game data, contributing both to methods of analysis and insights regarding indicators of student outcomes. With regards to the former, Mahajan et al. (2021) made use of computer vision and natural language processing to assess the accuracy of learner observations and descriptions of science-related images. Hum et al. (2022) used Bayesian Knowledge Tracing (BKT) to assess the structure and skill level of learners' scientific observations. With regards to the latter, researchers have found that learner traversals correlate with assessment outcomes early in gameplay (Esclamado & Rodrigo, 2022), and STEM interest relates to student assessment outcomes (Casano & Rodrigo, 2022) and influences the scientific observations made by the learners (Gadbury & Lane, 2022).

Gender and affect in game-based learning

Researchers have also examined how other factors such as emotions and gender influence student outcomes. Sabourin and Lester (2014) found that frustration and boredom are among the most frequent emotional states in open-ended learning. Frustration is an emotion that results from interference that prevents achievement of a goal (Rajendran et al., 2013). Boredom, on the other hand, occurs when individuals experience low arousal and dissatisfaction or disinterest in response to low arousal (Vogel-Walcutt et al., 2012). Previous work found that frustration, boredom, and other negative emotions are associated with negative learning and behavioral outcomes (Liu et al., 2013; Sabourin & Lester, 2014; Tze et al., 2015). For example, frustrated students tend to have a history of incorrect actions and help requests. Bored students tend to work slowly (Baker et al., 2012), have more idle time which is also an indicator of off-task behavior (Pardos et al., 2014), give up on a task (Andres et al., 2015), or engage in non-learning behaviors such as gaming the system (Baker et al., 2010). In the context of WHIMC, Esclamado et al. (2022) analyzed the relationship between behavioral and achievement outcomes and self-reported frustration

and boredom levels of students (Esclamado et al., 2022). They found that bored learners tended to have lower assessment outcomes and students who expressed boredom or frustration tended to demonstrate minimal engagement with STEM.

According to Riding and Grimley (1999), gender difference has an impact on performance in game-based learning. One key factor is that males and females have different approaches in receiving and processing information. They suggested that females tend to have a more thorough understanding of information as they are willing to take the time to process information and link the received information to existing knowledge. On the other hand, males may have a more superficial understanding of new information but are able to process more information than females. This difference is reflected in their game-based learning behaviors. Females prefer exploratory games while males prefer games that involve strategy and direct instructions from the system. Homer et al. (2012) also found that male learners prefer action games while female learners prefer to play virtual or simulation and puzzle games. Females also tend to prefer autonomous learning while men view digital games as a way to socialize and develop skills, which is why males are more engaged in game-based learning. Homer et al. (2012) found that male learners tend to be more engaged in digital-based learning and usually they spent more time on gameplay than female learners. Hamlen investigated the relationship between success in gameplay and time spent playing games in a study on gender difference, with the assumption that more time spent playing could lead to success in gameplay (Hamlen, 2010).

Gender difference in game-based learning affects learning outcomes (Dorji et al., 2015). A previous study found that males outperformed females in learning abilities in game-based learning since males were more interested in digital games (Ünlüsoy et al., 2010). Other previous studies found that female students outperformed male students in achievement and engagement (Khan et al., 2017; Lukosch et al., 2017).

Females and males may also have different patterns of affective states and behavior in game-based learning which may also affect their STEM interest. Chiu (2020) found that female learners were more likely to study STEM when they experienced less boredom and more off-task behavior, and male learners were more likely to study STEM when they felt less frustration and were more concentrated (Chiu, 2020).

While Minecraft is popular among both boys and girls, several studies have shown that more boys than girls played the game. Beavis et al. (2014) and Marsh et al. (2015) found that boys were more likely to play Minecraft and spent more time playing the game than girls. Peters et al. (2021) found that male learners were more likely to have better ability scores in Minecraft than female learners. Hughes et al. (2021) found no significant differences in achievement by gender.

Summary

Prior work has identified differences among male and female learners' behaviors and outcomes in open-ended learning environments in general and Minecraft in particular. However, our review of the literature finds that these studies are limited to gender differences in behaviors such as time spent on gameplay and being off-task. The studies that have examined gender differences in behavior such as being off-task have only used out-of-game data in identifying this behavior. There is little work on analyzing gender differences in Minecraft. With regards to WHIMC specifically, we did not find any previous studies that investigated gender differences in learners' achievement and learning experiences. We therefore hope to contribute to filling these gaps by examining the differences in in-game behavior, knowledge assessment, and STEM interest between male and female learners using WHIMC. We included several features for the in-game behavior such as exploration features and features which are indicators of off-task behavior. Additionally, we explore how the self-reported levels of frustration and boredom of male and female learners are related to these outcomes.

Methods

The study used a quantitative method to analyze the difference in in-game behaviors, assessment outcomes, and STEM interest between male and female learners using WHIMC. It also aimed to examine the difference in these outcomes among learners who expressed frustration or boredom and those who did not, both male and female. Quantitative data were gathered from knowledge assessments, and from the results of the Game Experience Questionnaire (GEQ) and the STEM interest questionnaire (SIQ). In addition, quantitative data representing students' in-game behaviors in WHIMC were derived from the interaction logs. We performed quantitative analysis of the male and female learners' in-game behaviors, assessment outcomes, and change in STEM interest after using WHIMC to determine the gender difference on these outcomes. We also performed quantitative analysis to compare these outcomes between learners who expressed feelings of frustration or boredom and those who did not, examining the differences in these outcomes among both male and female learners based on their reported levels of frustration and boredom.

Participants

The data analyzed for this paper consisted of in-game data, out-of-game assessment data, and results of the Game Experience Questionnaire (GEQ) and STEM interest questionnaire (SIQ) from 175 Grade 8 learners, 13 to 15 years old, from a school in the Philippines. There were 77 male learners and 98 female learners.

Data collection procedures

Partner teachers from the school developed two (2) learning modules that integrated the use of WHIMC in classroom lessons about biotic/abiotic components of the environment and adaptation.

WHIMC is instrumented to collect data about learner actions as they use the game. WHIMC logs learner positions in intervals of 6 seconds as they explore the game, or whenever the learner makes an observation or access a science tool. Details of the interaction logs included the timestamp, coordinates, the world explored, observations made, and the science tool used.

Knowledge was assessed out-of-game, after every module, using teacher-constructed quizzes. These quizzes consisted of objective questions and short essay questions. This paper does not analyze the validity of these quizzes. Rather, it relies on the test scores provided by the teacher.

The GEQ assessed players' levels of immersion, challenge, and negative / positive affect among others. The GEQ instrument used in this study is an abridged version of the instrument developed by IJsselsteijn et al. (2013). Two (2) items from GEQ: "I felt frustrated" and "I felt bored", were used for the data on student self-reported frustration and boredom respectively.

The SIQ is an instrument that is used to describe and measure the propensity of pursuing (or not pursuing) a career in STEM (see Table 1). Questions in the SIQ assessed students' self-efficacy as well as how they regarded the relevance and usefulness of science.

At the start of the data collection, the students were asked to answer the SIQ. They then explored WHIMC, following the instructions in the teacher-prepared learning modules. After the exploration, the learners answered the knowledge assessment, the GEQ, and the SIQ again.

Table 1 Items in SIQ

SCCT Constructs		SIQ Item
Self-Efficacy (SE)	SE1	I know I can do well in science.
	SE2	I think Science is challenging to learn.
Outcome Expectations (OE)	OE1	After I finish high school, I will use Science often.
	OE2	I believe that I can use Math and Science to solve problems in the future.
Interests (IN)	IN1	I enjoy Science activities.
	IN2	I enjoy solving Science and Math problems.
Choice Goals (CG)	CG1	Learning Science will help me get a good job.
	CG2	Knowing how to use Math and Science together will help me to invent useful things.
	CG3	Understanding engineering is not important for my career.
Choice Actions (CA)	CA	I try to get a good grade in science because I have an interest in science jobs.

Data analysis

We started the analysis by first dividing the learners into two groups: male and female learners. There were 77 male learners and 98 female learners.

We then focused our study of in-game behavior on time spent, idle time, distance traveled, area covered, average slope, number of observations, and task completion. Table 2 shows these in-game behaviors and their description.

The time spent and idle time were based on the behaviors in game-based learning that were investigated in prior studies (Hamlen, 2010; Pardos et al., 2014). In WHIMC, learners are expected to explore the environment before making observations, so exploration behaviors were also among the in-game behaviors considered in this study such as distance traveled, area covered, and average slope. The number of observations and task completion were also included since these are important indicators of the performance of the learners in WHIMC.

Based on the interaction logs, we generated the time sequences of learner explorations within WHIMC of each learner. The time sequence indicates the different worlds explored for the whole duration of the module use and the time spent in each world. The total time spent was computed by adding all the time spent in each exploration in the time sequence.

Idle time was defined as disengaging from the learning environment. It was operationalized as the time during which the learner stayed at a location point, with no change in distance traveled and no other activity, i.e., accessing science tools or making an observation.

For the distance traveled, area covered and average slope, we used the logs of location points on a particular WHIMC world – Lunar Base LeGuin which was an assigned world for Module 1 to be explored. We treated successive pairs of location points in the logs as line segments. Using the Euclidean distance, we calculated the total distance traveled by each learner adding all distances of all segments of the learner's path.

Table 2 Description of the in-game behaviors in WHIMC

In-game behavior	Description
Time spent	The total time spent in exploration
Idle time	The time during which the learner stayed at a location point, with no change in distance traveled and no other activity, i.e., accessing science tools or making an observation
Distance traveled	The total distance traveled in a WHIMC world
Area covered	The total area covered during exploration in a WHIMC world
Average slope	The average value of the slopes between the cumulative distance traveled by each learner in one-minute intervals
Number of observations	Total number of observations made for the two modules
Task completion	Whether the missions per module were completed or not

To get the value for the area covered by the learner, we determined the smallest convex polygon that contains all locations that the learner visited, i.e., the convex hull, using the Jarvis March algorithm. Based on these convex hulls, we computed the area that each learner explored.

To get the value for the average slope, we calculated the cumulative distance traveled by each learner in one-minute intervals. We also calculated the slopes between one-minute intervals and then calculated the average slope for each learner to determine how the learner explored the worlds. Lower values of the average slope mean there is a small change in the cumulative distance.

The number of observations made by each learner is the total number of observations made for the two modules.

Task completion refers to whether the missions per module were completed or not. The value of the task completion per module is determined based on the list of conditions to be met for each module (see Table 3) to indicate that the task per module is completed. These conditions were based on the learning modules.

To determine how the in-game behaviors, assessment outcomes, and the change in STEM interests of male and female learners differ, we used two-sample t-tests and the results were used to answer RQ1, RQ2, and RQ3 respectively. To answer RQ4, we then correlated the parameters of in-game behaviors with learners' assessment scores of male and female learners to determine the relationship between students' behavior and how much they learned using Pearson correlation.

The GEQ used Likert scale thus self-reported frustration and boredom were treated as ordinal data. To answer RQ5, the Spearman correlation was used to determine the relationship between students' self-reported frustration and boredom and assessment scores of male and female learners.

Then, we used two-sample t-tests to determine how the in-game behaviors and the change in STEM interests of male and female learners who self-reported frustration and boredom differ from those who did not. The results were used to answer RQ6 and RQ7 respectively. For self-reported frustration, we separated the learners into two groups: the first group was those learners who answered, "not at all" on the GEQ item "I felt frustrated" and the second

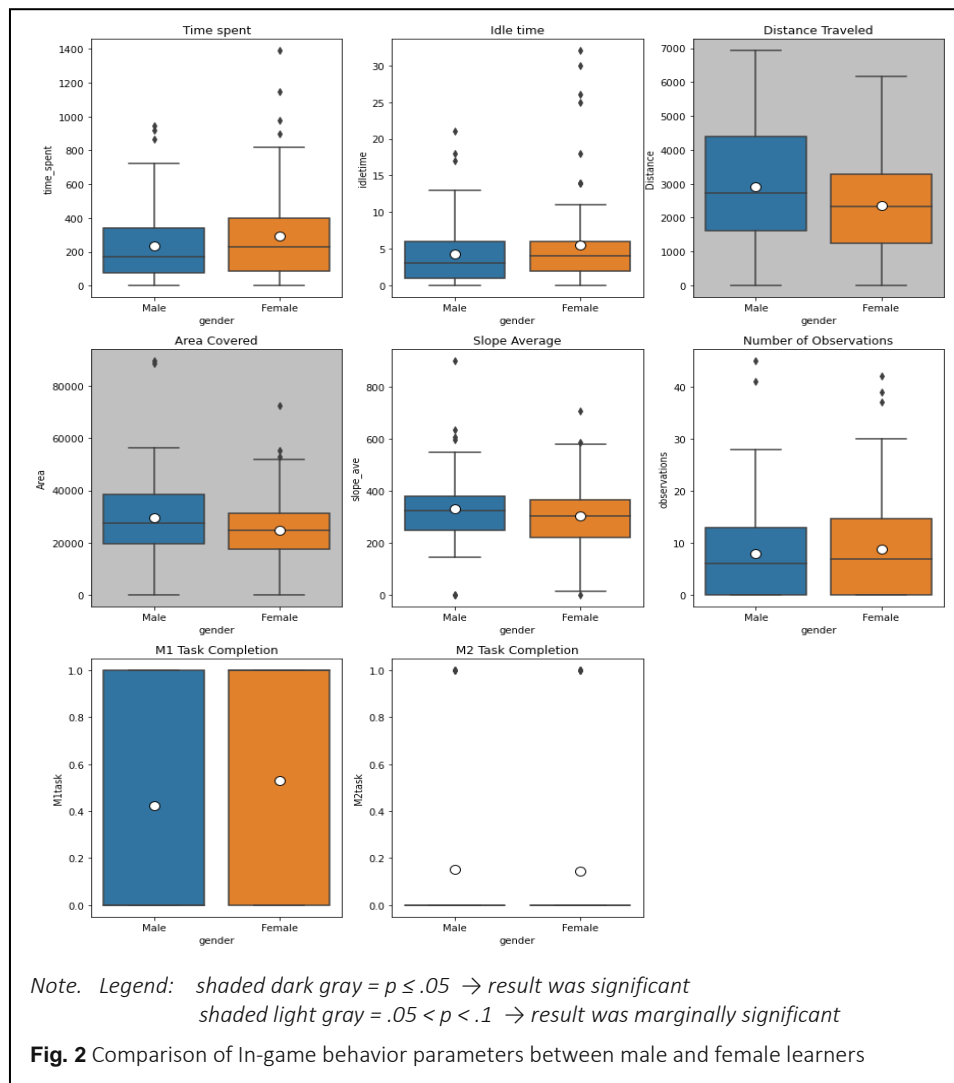**Table 3** List of conditions to be met to indicate that task is completed

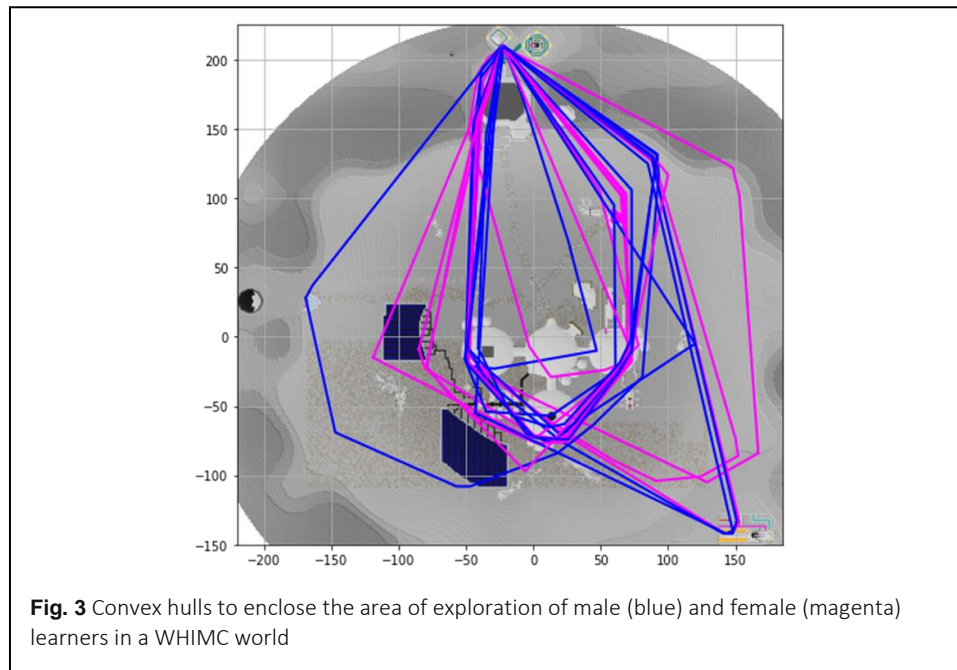| Module	Task completion (Task is completed if all conditions are met)
Module 1	1, if learner explored Lunar Base Le Guin and Observation Training area and made observation
	0, otherwise
Module 2	1, if learner explored No Moon and Colder Sun worlds and used science tools, and explored Tilted Earth worlds and made observation
	0, otherwise

group was those learners who have certain levels of frustration and answered "slightly", "moderately", "fairly", and "extremely". For self-reported boredom, another set of comparisons was also done. The learners were also separated into two groups: the first group was those learners who answered, "not at all" on the GEQ item "I felt bored" and the second group was those learners who have certain levels of boredom and answered "slightly", "moderately", "fairly", and "extremely".

Results

RQ1: How do the in-game behaviors of male and female learners differ?

For in-game behaviors, there were marginally significant differences in the distance traveled ($t(136) = 1.98$, $p = .05$) and area covered ($t(136) = 1.79$, $p = .08$) between the male and female learners. In Figure 2, it is shown that male learners have higher values of

Note. Legend: *shaded dark gray = p ≤ .05 → result was significant*
shaded light gray = .05 < p < .1 → result was marginally significant

Fig. 2 Comparison of In-game behavior parameters between male and female learners

Fig. 3 Convex hulls to enclose the area of exploration of male (blue) and female (magenta) learners in a WHIMC world

distance and area that implies that male learners traveled greater distances and covered a larger area during exploration compared to female learners. Figure 3 illustrates the convex hulls that enclose the area explored by the male (blue) and female (magenta) learners. It shows that most of the male learners explored larger areas compared to the female learners.

RQ2: How do the assessment outcomes of male and female learners differ?

In Figure 4, it is shown that the mean value of the assessment scores of the female learners was higher than the male learners but the difference was not significant ($t(173) = -1.37$, $p = .17$).
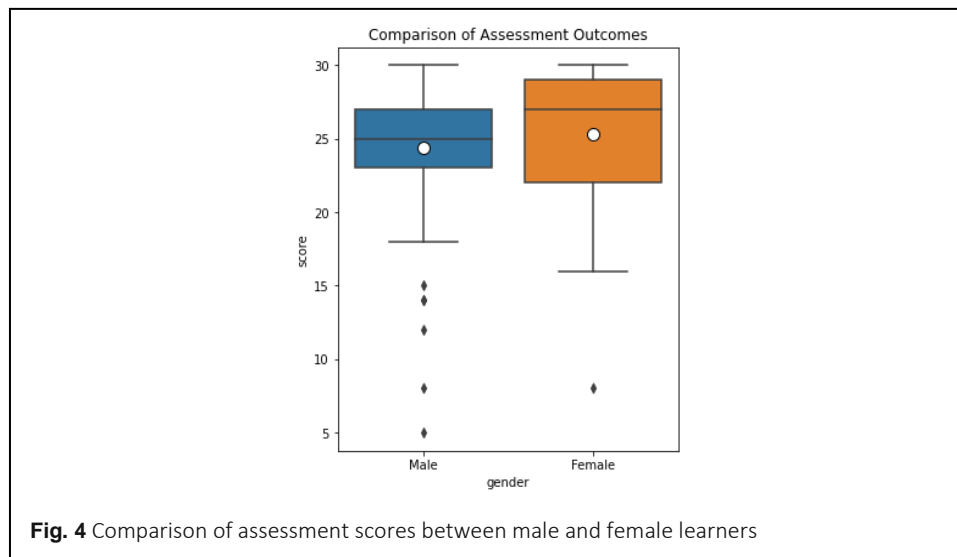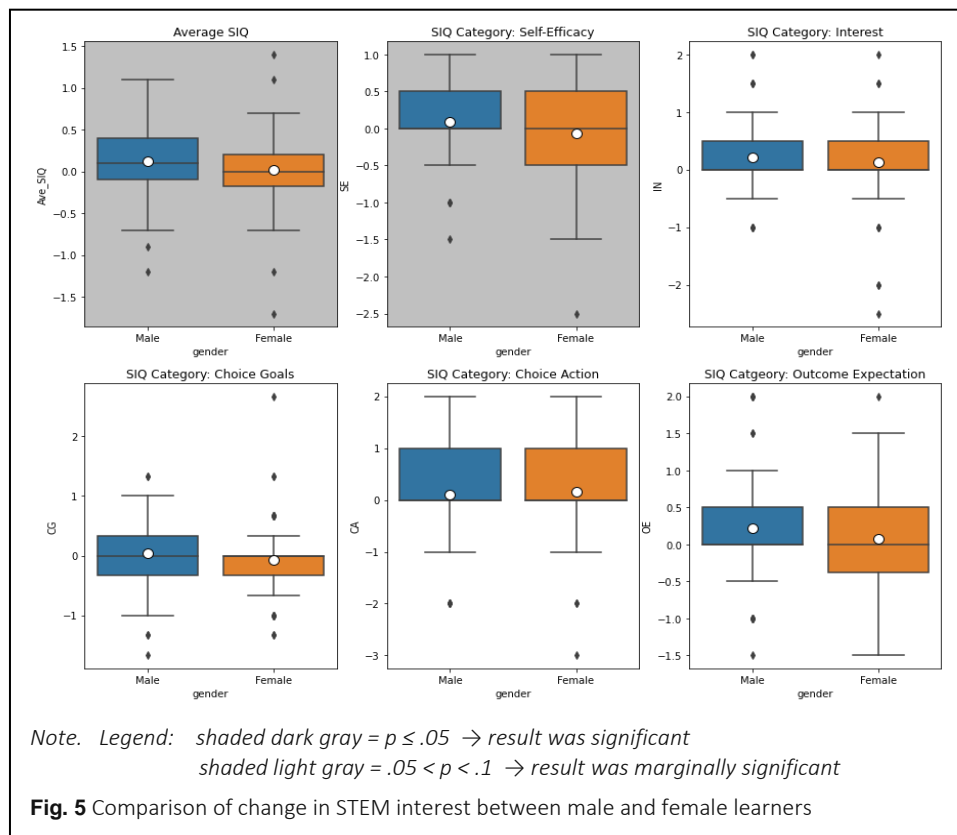

Fig. 4 Comparison of assessment scores between male and female learners

RQ3: How does the change in STEM interest of male and female learners differ after using WHIMC?

Figure 5 shows that the change in the average SIQ of male learners was higher compared to the female learners and the result was marginally significant ($t(173) = 1.7$, $p = .09$). In the category of Self-Efficacy, the change in SIQ of the male learners was higher than the female learners and the result was also marginally significant ($t(173) = 1.65$, $p = .09$). These suggest that male learners had a higher increase in STEM interest after using WHIMC compared to the female learners.

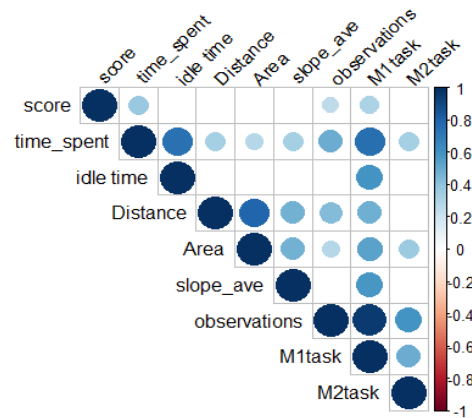**RQ4: What is the relationship between the in-game behaviors and assessment outcomes of male and female learners?**

For all learners, we found significant positive correlations (see Figure 6) between the time spent ($r = .38$, $p = < .001$), idle time ($r = .28$, $p = < .001$), number of observations ($r = .32$, $p = < .001$), and task completion for the two modules and their assessment scores ($r = .4$, $p = < .001$ and $r = .22$, $p = .05$). Similar findings can be seen among female learners since based on the results, there were significant positive correlations (see Figure 8) between the time spent ($r = .38$, $p = < .001$), idle time ($r = .32$, $p = .003$), number of observations ($r = .37$, $p = < .001$), and task completion for module 1 ($r = .45$, $p = .001$) and their

Note. Legend: *shaded dark gray = p ≤ .05 → result was significant*
shaded light gray = .05 < p < .1 → result was marginally significant

Fig. 5 Comparison of change in STEM interest between male and female learners

assessment scores. There was also a marginally significant positive correlation between task completion for module 2 and the assessment score for female learners ($r = .27$, $p = .06$). On the other hand, for male learners, significant positive correlations (see Figure 7) can only be found between time spent ($r = .37$, $p = < .001$) and number of observations ($r = .26$, $p = .025$), and the assessment scores and a marginally significant positive correlation between task completion for module 1 ($r = .32$, $p = .07$) and the assessment scores. These suggest that students learned more if they spent more time in exploration, had more idle time, made more observations, and completed the tasks and these were more common among female learners. Male learners learned more if they spent more time in the exploration, made more observations, and completed tasks. Based on the results, female learners learned more if they have more idle time compared to male learners.
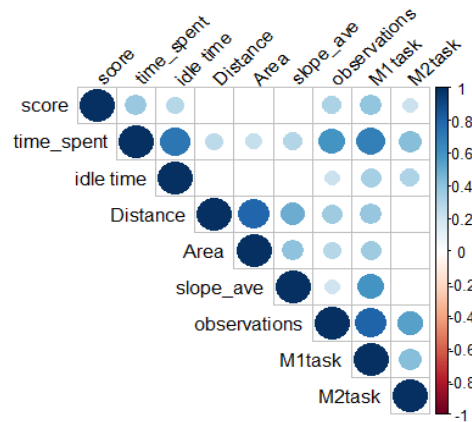
Fig. 6 Results of correlation between parameters of in-game behavior and assessment scores of all learners

Fig. 7 Results of correlation between parameters of in-game behavior and assessment scores of male learners
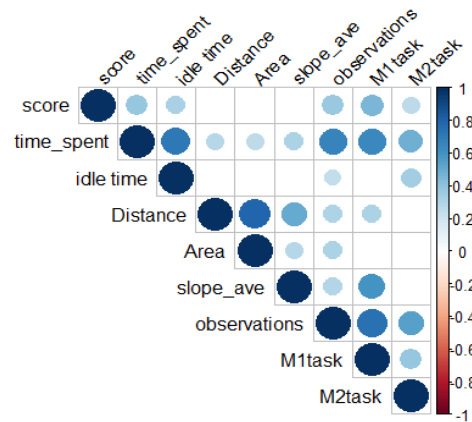
Fig. 8 Results of correlation between parameters of in-game behavior and assessment scores of female learners

RQ5: What is the relationship between the learners' self-reported frustration and boredom and their assessment outcomes of male and female learners?
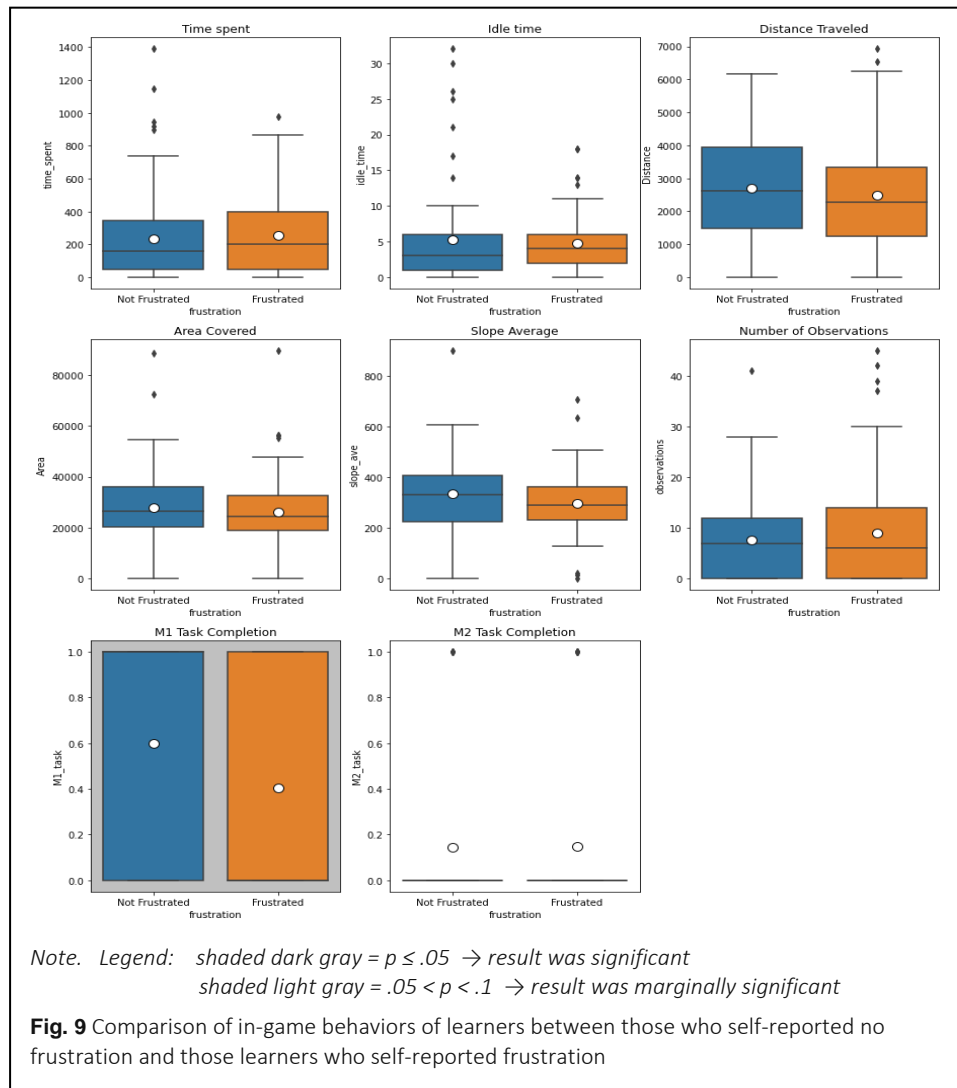
There were no significant correlations between frustration and assessment scores, but we found a significant negative correlation (see Table 4) between boredom and assessment scores for all learners, and this was also marginally significant for the female learners. This suggests that bored learners learned less, and this is more common among female learners.
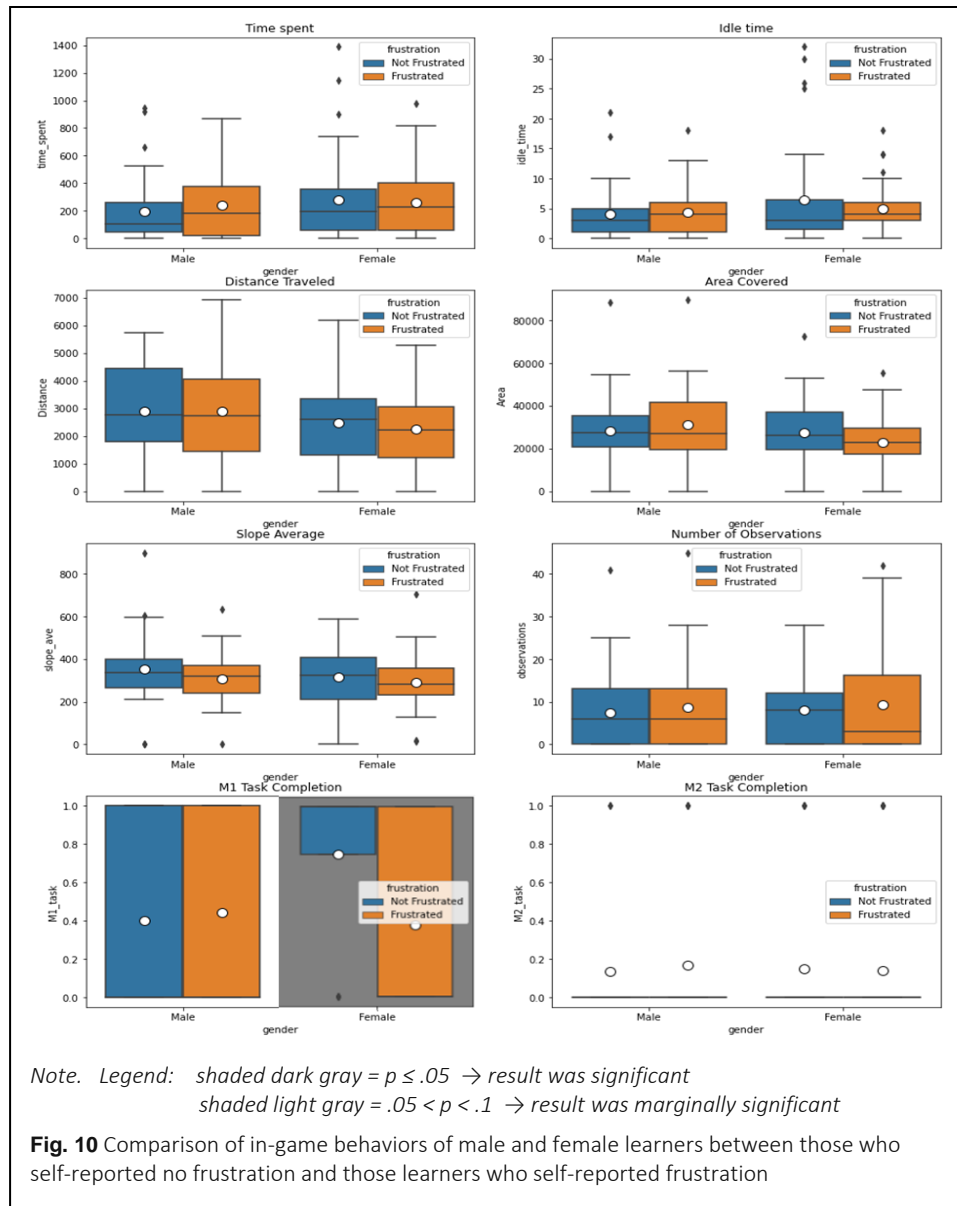
RQ6: How do in-game behaviors of the male and female learners who self-reported frustration and boredom differ from those who did not?

For all learners, the task completion of module 1 of those who self-reported frustration was lower than of those who did not express frustration (see Figure 9) and the difference was marginally significant ($t(80) = -1.77$, $p = .08$). For female learners (see Figure 10), this difference was significant ($t(47) = -2.69$, $p = .01$). On the other hand, for male learners there was no significant difference in the in-game behaviors between those who self-reported frustration and those who did not (see Figure 10). These suggest that frustrated learners, especially female learners, tend to not complete the tasks.

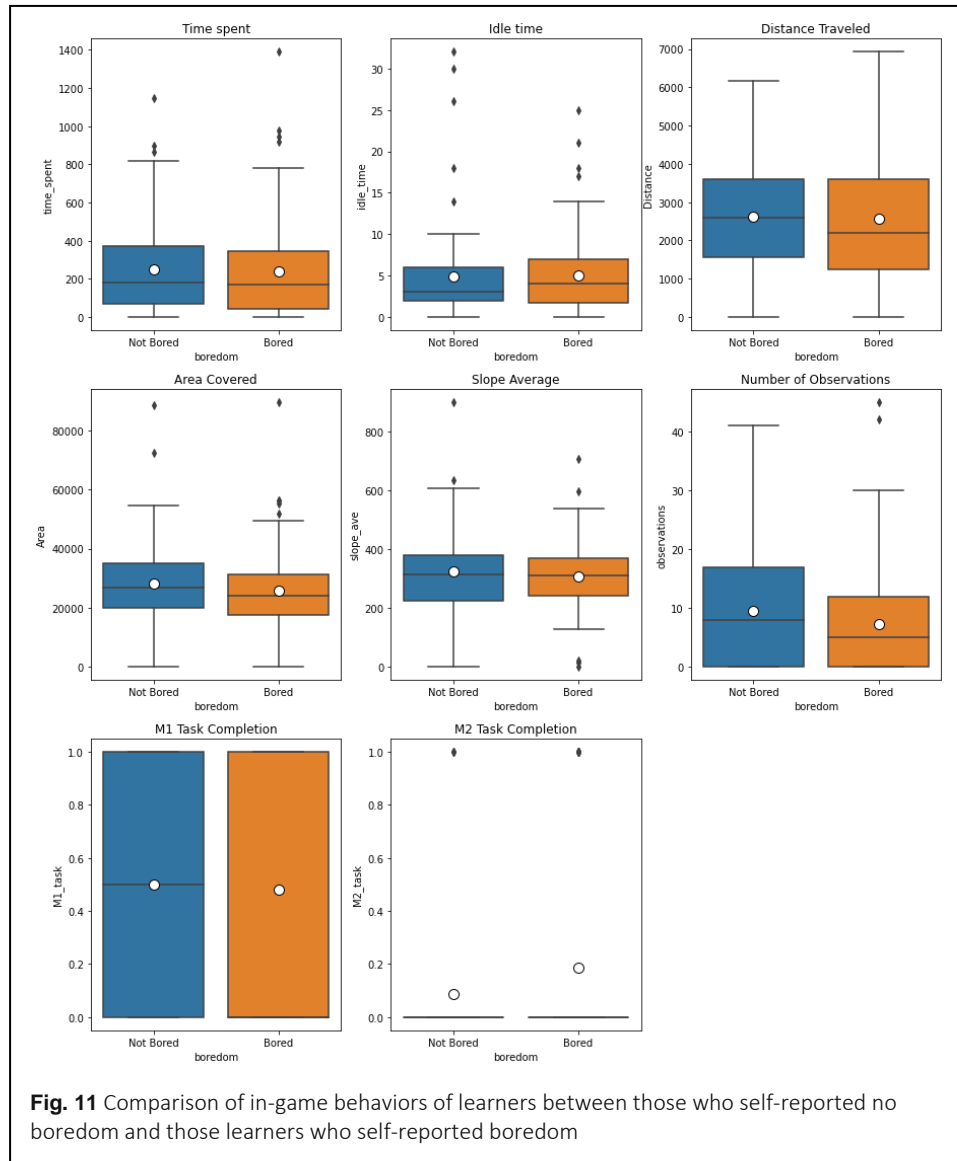**Table 4** Results of correlation between self-reported frustration and boredom and assessment scores of the learners

	All participants (N=175)	Male (N=77)	Female (N=98)
Frustration and assessment score	$r = -.07$, $p = .33$	$r = -.11$, $p = .33$	$r = -.05$, $p = .61$
Boredom and assessment score	$r = -.17$, $p = .03$	$r = -.13$, $p = .28$	$r = -.19$, $p = .06$

Note.   Legend: shaded dark gray = p ≤ .05   → result was significant
shaded light gray = .05 < p < .1   → result was marginally significant

Fig. 9 Comparison of in-game behaviors of learners between those who self-reported no frustration and those learners who self-reported frustration

Note. Legend: *shaded dark gray = p ≤ .05 → result was significant*
shaded light gray = .05 < p < .1 → result was marginally significant

Fig. 10 Comparison of in-game behaviors of male and female learners between those who self-reported no frustration and those learners who self-reported frustration
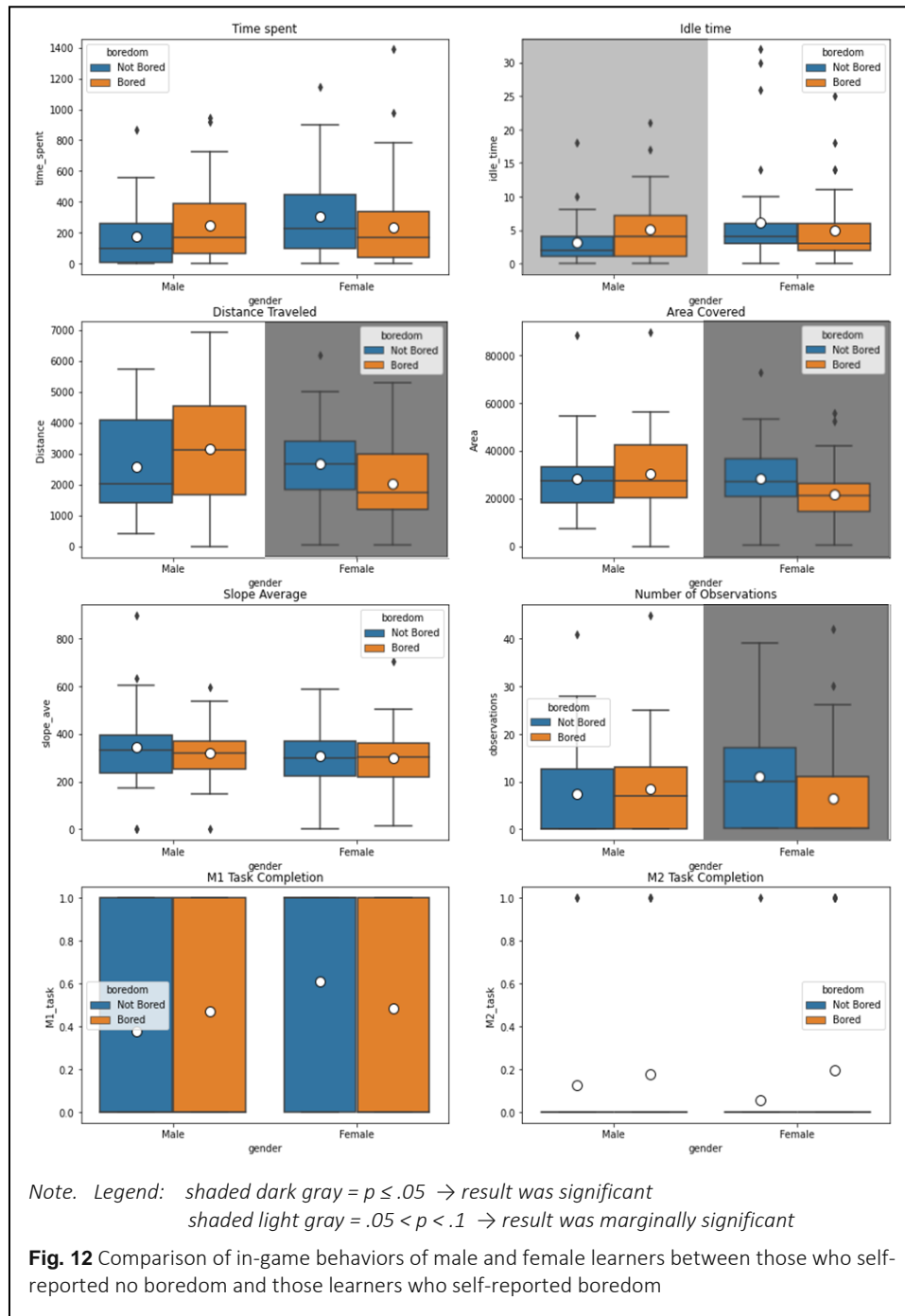
For all learners, there was no significant difference in the in-game behaviors between those who self-reported boredom and those who did not (see Figure 11). For male learners, idle time of those who self-reported boredom was higher than of those who did not (see Figure 12) and the difference was marginally significant ($t(64) = 1.88$, $p = .06$). For female learners, the distance traveled ($t(75) = -1.96$, $p = .05$), area covered ($t(75) = -2.17$, $p = .03$), and number of observations ($t(96) = -2.3$, $p = .02$) of those who self-reported boredom were significantly lower than of those who did not express boredom. These suggest that bored male learners tend to have more idle time compared to those who did not express boredom. On the other hand, bored female learners tend to travel less, cover smaller area, and make fewer observations compared to those who did not express boredom.

Fig. 11 Comparison of in-game behaviors of learners between those who self-reported no boredom and those learners who self-reported boredom
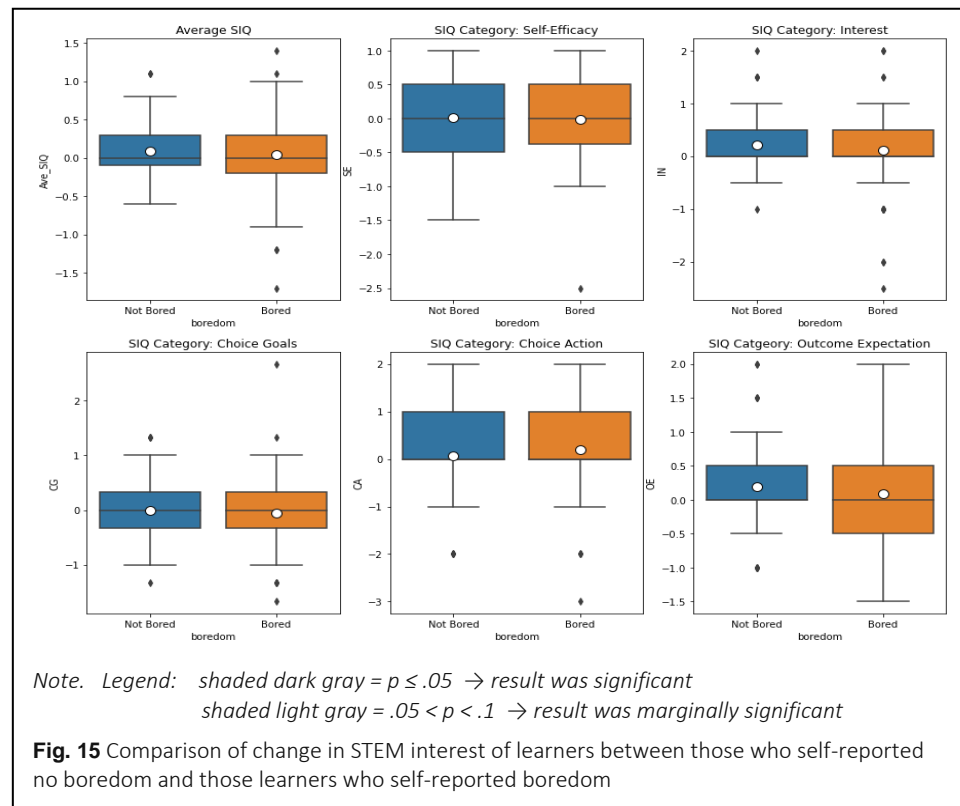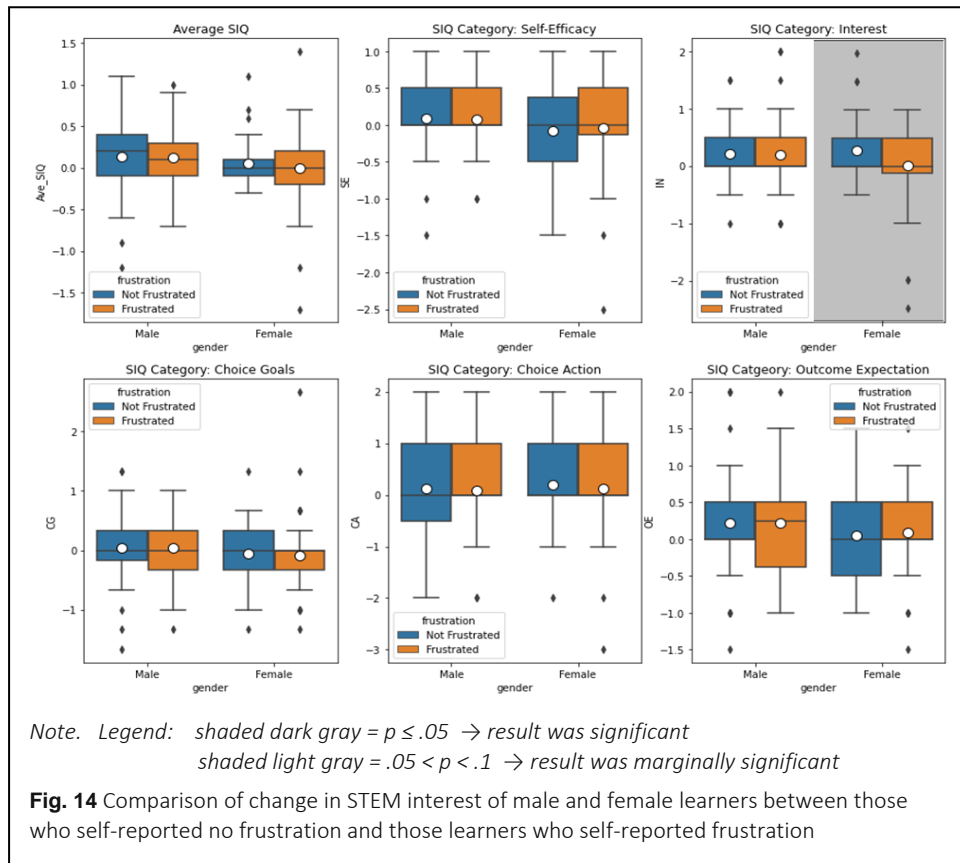
Note. Legend: *shaded dark gray = p ≤ .05 → result was significant*
shaded light gray = .05 < p < .1 → result was marginally significant

Fig. 12 Comparison of in-game behaviors of male and female learners between those who self-reported no boredom and those learners who self-reported boredom
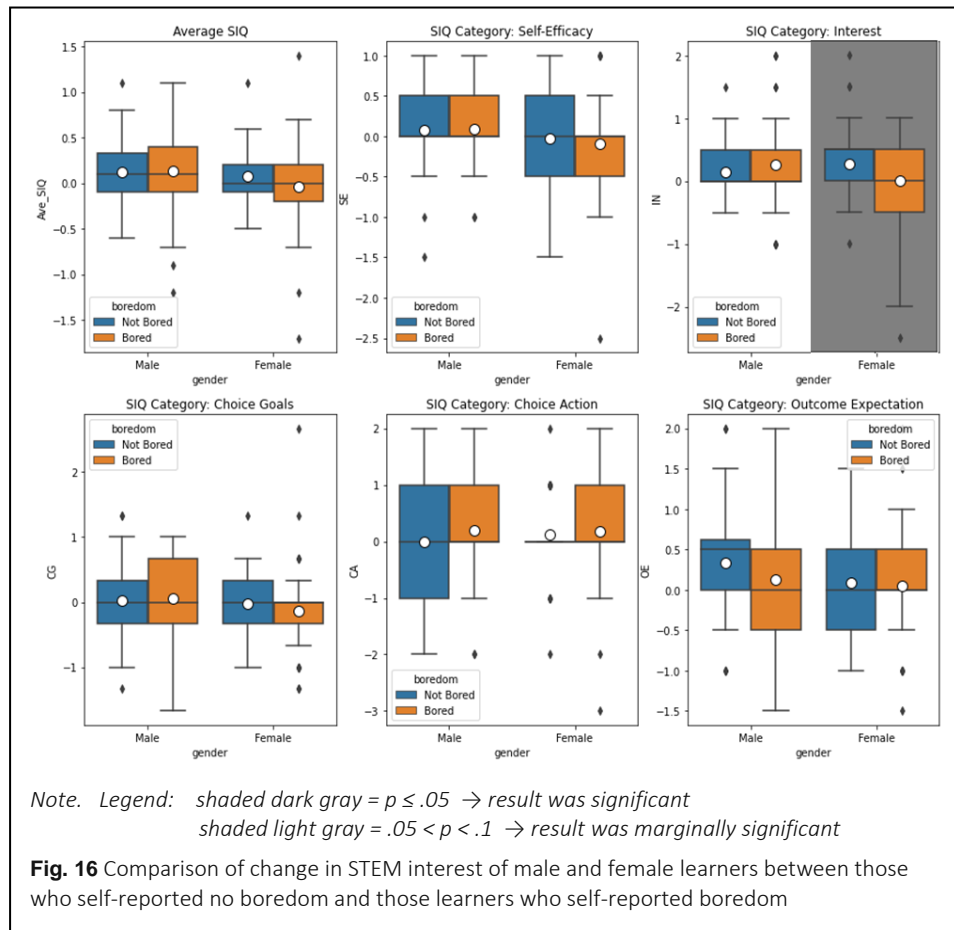
RQ7: How does the change in STEM interest of the male and female learners who self-reported frustration and boredom differ from those who did not?

For all learners, we found no significant difference in the change in SIQ between those who self-reported frustration or boredom and those who did not (see Figures 13 and 15). There was also no significant difference in the change in SIQ between those who self-

Fig. 13 Comparison of change in STEM interest of learners between those who self-reported no frustration and those learners who self-reported frustration
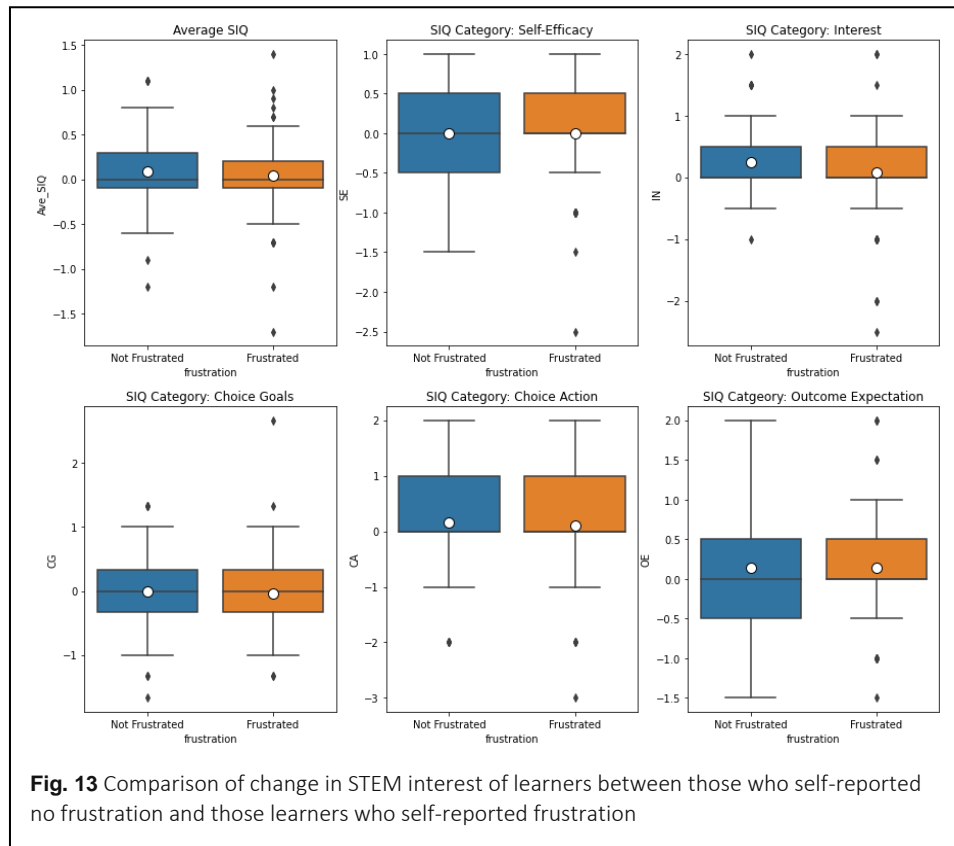
reported frustration or boredom and those who did not among the male learners (see Figures 14 and 16). For the female learners, we found that the change in SIQ in the Interest category of those who self-reported frustration was lower compared to female learners who did not express frustration (see Figure 14) and the result was marginally significant. We also found that the change in SIQ in the Interest category of those who self-reported boredom was significantly lower compared to those who did not express boredom (see Figure 16). This implies that female learners who were frustrated or bored tend to have a decreased interest in STEM after playing WHIMC.

Note. Legend: shaded dark gray = p ≤ .05 → result was significant
shaded light gray = .05 < p < .1 → result was marginally significant

Fig. 14 Comparison of change in STEM interest of male and female learners between those who self-reported no frustration and those learners who self-reported frustration

Note. Legend: shaded dark gray = p ≤ .05 → result was significant
shaded light gray = .05 < p < .1 → result was marginally significant

Fig. 15 Comparison of change in STEM interest of learners between those who self-reported no boredom and those learners who self-reported boredom

Note. Legend: shaded dark gray = p ≤ .05 → result was significant
shaded light gray = .05 < p < .1 → result was marginally significant

Fig. 16 Comparison of change in STEM interest of male and female learners between those who self-reported no boredom and those learners who self-reported boredom

Discussion

Exploratory learning environments such as Minecraft have gained popularity in recent years due to their potential to promote creativity, critical thinking and problem-solving skills. However, assessing student learning and the learning experience in such environments can be challenging. Prior works have examined various assessment tools such as self-reporting (Callaghan, 2016; Melián Díaz et al., 2020), knowledge assessment (Tangkui & Keong, 2021), observations (Callaghan, 2016), and learning analytics to analyze the learning experience of the learners (Horn et al., 2016). Using these tools, we can collect data such as learners' achievement, behavior, emotional responses to the game, and STEM interest which are key factors in assessing students' learning.

Previous studies have investigated gender differences in learners' behavior and outcomes in open-ended learning environments (Chiu, 2020; Khan et al., 2017; Lukosch et al., 2017; Ünlüsoy et al., 2010) and in Minecraft (Beavis et al., 2014; Hughes et al., 2021; Marsh et al. 2015; Mavoa et al., 2018; Peters et al., 2021). Most of these behaviors were identified using out-of-game data. Some studies reported significant differences in achievement, behavior, and learning experience between genders (Chiu, 2020; Khan et al., 2017;

Lukosch et al., 2017; Ünlüsoy et al., 2010). On the other hand, a prior study reported no significant difference in achievement between genders (Hughes et al., 2021). Prior studies on gender difference of learners using Minecraft are limited and no prior studies have investigated gender differences in WHIMC regarding these outcomes.

In this paper, we presented an analysis of the difference in in-game behavior, knowledge assessments, and STEM interest of male and female learners and determined the gender differences between learners' self-reported frustration and boredom with their behavior, knowledge assessments, and STEM interest.

We found that the in-game behaviors and change in STEM interest of male learners differ from the female learners. Male learners tend to travel more and cover a larger area of exploration compared to the female learners (**RQ1**) which resonates with prior research on gender difference that male learners tend to be more engaged in digital game-based learning and spent more time on gameplay (Homer et al., 2012). This may explain why they tend to travel more than female learners. After using WHIMC, male learners had a higher increase in STEM interest compared to the female learners (**RQ3**). Both male and female learners tend to learn more if they spent more time in exploration, made more observations and completed tasks but only the female learners tend to learn more if they have more idle time, which is an indicator of off-task behavior (**RQ4**). Prior research also found that female learners were more likely to study STEM if they are more off-task (Chiu, 2020). The propensity of the learners pursuing STEM may be also related to their performance. This may support why female learners tend to learn more if they have more idle time, an indicator of off-task behavior.

Prior research reported gender differences in performance in game-based learning. Some studies reported that male learners outperformed the female learners in achievement and engagement (Ünlüsoy et al., 2010) while some studies reported that female learners outperformed the male learners (Khan et al., 2017; Lukosch et al, 2017). However, in this paper we found no significant difference in the assessment outcomes between male and female learners (**RQ2**).

When we examined the relationship between learners' frustration and boredom with their knowledge assessment outcomes, bored learners tended to have lower academic outcomes, and this is more common among female learners (**RQ5**). Frustrated learners, especially female learners, tend to not complete the tasks. Bored male learners tend to have more idle time compared to those who did not express boredom. Bored female learners tend to travel less, cover smaller area, and make fewer observations compared to those who did not express boredom (**RQ6**). Female learners who expressed frustration or boredom tend to have a decreased interest in SIQ in the Interest category after playing WHIMC (**RQ7**). These findings aligned with prior research showing that negative emotions such as frustration and boredom can have a detrimental effect on the behavior, achievement, and

STEM interest of male and female learners. Specifically, previous research has found that female learners who experienced boredom and male learners who experienced frustration were less likely to learn and study STEM (Chiu, 2020). The findings build on this by highlighting that frustration and boredom can have a negative impact on the behavior, performance, and STEM interest of learners in game-based learning environments, particularly among female learners. Therefore, it is important to consider how negative emotions may affect different genders in order to optimize learning outcomes.

Conclusion

The findings of this study highlight the importance of considering gender differences in learning behaviors and learning experience when designing game-based learning environments. For instance, teachers can encourage female learners to take more breaks and engage in off-task behavior, which has been found to enhance their performance and interest in STEM subjects. Game-based learning designers can also use these findings to consider the learning behaviors of both male and female learners. Educators should also be aware of the negative impact of negative emotions such as frustration and boredom on the learning outcomes of learners, especially female learners, and consider incorporating strategies to mitigate these negative emotions. Overall, these findings can help create a more effective and engaging game-based learning experience for learners of all genders.

Limitations and future research

The work presented in this paper is subject to several limitations. First, two (2) items from the Game Experience Questionnaire (GEQ): "I felt frustrated" and "I felt bored", were used for the data on student self-reported frustration and boredom respectively. The items in the GEQ are used to measure the factors in a game that contribute to an engaging game experience which includes several dimensions of the player experience. In future work, considering other items from GEQ may provide more insights on gender differences. Second, the affect considered in this paper were frustration and boredom, which are among the most frequent emotional states in open-ended learning. Investigating other emotional states both positive and negative would provide more insights on how the male and female learners differ in terms of behavior, achievement, and STEM interest in these emotional states. Finally, it would also be interesting to perform some cross-cultural analysis on gender differences to compare data from Philippine learners against data of the learners from other countries.

Abbreviations
BKT: Bayesian Knowledge Tracing; GEQ: Game Experience Questionnaire; SIQ: STEM interest questionnaire; STEM: Science, Technology, Engineering, and Mathematics; WHIMC: What-If Hypothetical Implementations using Minecraft.

Authors' contributions

Maricel A. Esclamado was primarily responsible for the quantitative analysis of the data and contributed to the writing of the paper. Maria Mercedes T. Rodrigo conceptualized the study, scaffolded the analysis of the data, and contributed to the writing of the paper.

Authors' information

Maricel A. Esclamado is a graduate student currently taking Ph.D. in Computer Science at Ateneo de Manila University. She has been working on the data analysis for the WHIMC project in Ateneo Laboratory for the Learning Sciences (ALLS) since 2021. She is also an Associate Professor at the University of Science and Technology of Southern Philippines.

Maria Mercedes T. Rodrigo is a professor at the Department of Information Systems and Computer Science of Ateneo de Manila University. She is the head of the Ateneo Laboratory for the Learning Sciences. She has served as local organizer and International Program Coordination Chair of the International Conference in Computers in Education 2018 and 2021 respectively. Most recently, she was program co-chair of the Artificial Intelligence in Education Conference 2022. In 2021, Dr. Rodrigo received the Distinguished Researcher Award from the Asia-Pacific Society for Computers in Education (APSCE) and is President of APSCE.

Availability of data and materials

The dataset(s) supporting the conclusions of this article are available in the repository, https://github.com/CodeDevMarshall/RPTEL_GenderDifference_Esclamado_Rodrigo

For data privacy reasons, please send a request to mrodrigo@ateneo.edu for the actual log files.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details

[1] Ateneo de Manila University, Loyola Heights, Quezon City, Philippines

[2] University of Science and Technology of Southern Philippines, Cagayan de Oro City, Philippines

References

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, *92-93*, 90–103. https://doi.org/10.1016/j.compedu.2015.10.007

Andres, J. M. L., Rodrigo, M. M. T., Baker, R. S., Paquette, L., Shute, V. J., & Ventura, M. (2015). Analyzing student action sequences and affect while playing physics playground. In K. Porayska-Pomsta & K. Vebert (Eds.), *2015 Workshop Proceedings of the 8th International Conference on Educational Data Mining* (pp. 124–133). CEUR-WS.org. https://ceur-ws.org/Vol-1446/edm2015ws_proc.pdf

Baek, Y., Min, E., & Yun, S. (2020). Mining educational implications of Minecraft. *Computers in the Schools*, *37*(1), 1–16. https://doi.org/10.1080/07380569.2020.1719802

Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., Aleven, V., Kusbit, G. W., Ocumpaugh, J., & Rossi, L. M. (2012). Towards Sensor-Free affect detection in cognitive tutor algebra. *Educational Data Mining*, *2012*(1).

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, Ma. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different

computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Beavis, C., Muspratt, S., & Thompson, R. (2014). 'Computer games can get your brain working': Student experience and perceptions of digital games in the classroom. *Learning, Media and Technology*, *40*(1), 21–42. https://doi.org/10.1080/17439884.2014.904339

Bitner, J. (2021, March 15). *What is Minecraft?*. https://www.digitaltrends.com/gaming/what-is-minecraft/

Callaghan, N. (2016). Investigating the role of Minecraft in educational learning environments. *Educational Media International*, *53*(4), 244–260. https://doi.org/10.1080/09523987.2016.1254877

Casano, J., & Rodrigo, M. M. T. (2022, Nov-Dec). An analysis of Filipino primary school learners' game experience and STEM interest within Minecraft. In S. Iyer et al. (Eds.), *Proceedings of the 30th International Conference on Computers in Education*, *I* (pp. 455–460). Asia-Pacific Society for Computers in Education.

Chiu, M. S. (2020). Gender differences in predicting STEM choice by affective states and behaviors in online mathematical problem solving: Positive-affect-to-success hypothesis. *Educational Data Mining*, *12*(2), 48–77. https://doi.org/10.5281/zenodo.4008056

Dorji, U., Panjaburee, P., & Srisawasdi, N. (2015). Gender differences in students' learning achievements and awareness through residence energy saving game-based inquiry playing. *Journal of Computers in Education*, *2*(2), 227–243. https://doi.org/10.1007/s40692-015-0033-2

Esclamado, M. A., & Rodrigo, M. M. (2022). Are all who wander lost? An exploratory analysis of learner traversals of Minecraft Worlds. In M. M. Rodrigo, N. Matsuda, A. I. Cristea & V. Dimitrova (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. AIED 2022. Lecture Notes in Computer Science, vol 13356* (pp. 263–266). Springer, Cham. https://doi.org/10.1007/978-3-031-11647-6_48

Esclamado, M. A., Rodrigo, M. M. T., & Casano, J. A. (2022). Achievement, behaviors, and STEM interest of frustrated and bored learners using Minecraft. In S. Iyer et al. (Eds.), *Proceedings of the 30th International Conference on Computers in Education*, *I* (pp. 477–486). Asia-Pacific Society for Computers in Education.

Felecia, P. (2020). *Using education games in the classroom: Guidelines for successful learning outcomes. A handbook for teachers*. European Schoolnet.

Gadbury, M., & Lane, H. C. (2022). Mining for stem interest behaviors in Minecraft. In M. M. Rodrigo, N. Matsuda, A. I. Cristea & V. Dimitrova (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. AIED 2022. Lecture Notes in Computer Science, vol 13356* (pp. 236–239). Springer, Cham. https://doi.org/10.1007/978-3-031-11647-6_42

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2006). Nonscience majors learning science: A theoretical model of motivation. *Journal of Research in Science Teaching*, *44*(8), 1088–1107. https://doi.org/10.1002/tea.20181

Hamlen, K. R. (2010). Re-examining gender differences in video game play: Time spent and feelings of success. *Journal of Educational Computing Research*, *43*(3), 293–308. https://doi.org/10.2190/ec.43.3.b

Homer, B. D., Hayward, E. O., Frye, J., & Plass, J. L. (2012). Gender and player characteristics in video game play of Preadolescents. *Computers in Human Behavior*, *28*(5), 1782–1789. https://doi.org/10.1016/j.chb.2012.04.018

Horn, B., Hoover, A. K., Barnes, J., Folajimi, Y., Smith, G., & Harteveld, C. (2016). Opening the black box of play. In A. Cox & Phoebe O. Toups Dugas (Eds.), *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 142–153). ACM. https://doi.org/10.1145/2967934.2968109

Hughes, B., Lux, N., Frank, B., Willoughby, S., LaMeres, B., & Weyerbacher, R. (2021). Minecraft learning system for spatial reasoning in middle grades learners. *2020 ASEE Virtual Annual Conference Content Access Proceedings*. ASEE. https://doi.org/10.18260/1-2--34974

Hum, S., Stinar, F., Lee, H., Ginger, J., & Lane, H. C. (2022). Classification of natural language descriptions for Bayesian knowledge tracing in Minecraft. In M. M. Rodrigo, N. Matsuda, A. I. Cristea & V. Dimitrova (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. AIED 2022. Lecture Notes in Computer Science, vol 13356* (pp. 250–253). Springer, Cham. https://doi.org/10.1007/978-3-031-11647-6_45

IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). *The Game Experience Questionnaire*. Technische Universiteit Eindhoven.

Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, *72*, 757–770. https://doi.org/10.1016/j.chb.2016.09.062

Khan, A., Ahmad, F. H., & Malik, M. M. (2017). Use of digital game based learning and gamification in secondary school science: The effect on student engagement, learning and gender difference. *Education and Information Technologies*, *22*(6), 2767–2804. https://doi.org/10.1007/s10639-017-9622-1

Lane, H. C., Gadbury, M., Ginger, J., Yi, S., Comins, N., Henhapl, J., & Rivera-Rogers, A. (2022). Triggering stem interest with Minecraft in a hybrid summer camp. *Technology, Mind, and Behavior*, *3*(4). https://doi.org/10.1037/tmb0000077

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. S. J. d. (2013). Sequences of frustration and confusion, and learning. In S. K. D'Mello, R. A. Calvo & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 114–120). International Educational Data Mining Society.

Lukosch, H., Kurapati, S., Groen, D., & Verbraeck, A. (2017). Gender and cultural differences in game-based learning experiences. *Electronic Journal of e-Learning*, *15*(4), 310–319.

Mahajan, J., Hum, S., Ginger, J., & Lane, H. C. (2022). MineObserver: A deep learning framework for assessing natural language descriptions of Minecraft imagery. In *The International FLAIRS Conference Proceedings*, *35*. https://doi.org/10.32473/flairs.v35i.130729

Manahan, D. M. A. B., & Rodrigo, M. M. T. (2022, Nov-Dec). Support structures and activities for teachers preparing for game-based learning. In S. Iyer et al. (Eds.), *Proceedings of the 30th International Conference on Computers in Education*, *II* (pp. 325–335). Asia-Pacific Society for Computers in Education.

Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J. C., Davenport, A., Davis, S., Robinson, P., & Piras, M. (2015). *Exploring play and creativity in pre-schoolers' use of apps: Final project report*. University of Sheffield. http://www.techandplay.org/reports/TAP_Final_Report.pdf

Mavoa, J., Carter, M., & Gibbs, M. (2018). Children and Minecraft: A survey of children's digital play. *New Media & Society*, *20*(9), 3283–3303. https://doi.org/10.1177/1461444817745320

Mavrikis, M., Gutierrez-Santos, S., & Poulovassilis, A. (2016). Design and evaluation of teacher assistance tools for exploratory learning environments. In D. Gašević & G. Lynch (Eds.), *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16* (pp. 168–172). ACM. https://doi.org/10.1145/2883851.2883909

Melián Díaz, D., Saorín, J. L., Carbonell-Carrera, C., & de la Torre Cantero, J. (2020). Minecraft: Three-dimensional construction workshop for improvement of creativity. *Technology, Pedagogy and Education*, *29*(5), 665–678. https://doi.org/10.1080/1475939x.2020.1814854

Paraskeva, F., Mysirlaki, S., & Papagianni, A. (2010). Multiplayer online games as educational tools: Facing new challenges in learning. *Computers & Education*, *54*(2), 498–505. https://doi.org/10.1016/j.compedu.2009.09.001

Pardos, Z. A., Baker, R. S. J. D., San Pedro, M., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, *1*(1), 107–128. https://doi.org/10.18608/jla.2014.11.6

Peters, H., Kyngdon, A., & Stillwell, D. (2021). Construction and validation of a game-based intelligence assessment in Minecraft. *Computers in Human Behavior*, *119*, 106701. https://doi.org/10.1016/j.chb.2021.106701

Prensky, M. (2003). *Digital game-based learning*. McGraw-Hill.

Rajendran, R., Iyer, S., Murthy, S., Wilson, C., & Sheard, J. (2013). A theory-driven approach to predict frustration in an ITS. *IEEE Transactions on Learning Technologies*, *6*(4), 378–388. https://doi.org/10.1109/tlt.2013.31

Riding, R., & Grimley, M. (1999). Cognitive style, gender and learning from multi-media materials in 11-year-old children. *British Journal of Educational Technology*, *30*(1), 43–56. https://doi.org/10.1111/1467-8535.00089

Sabourin, J. L., & Lester, J. C. (2014). Affect and engagement in game-based learning environments. *IEEE Transactions on Affective Computing*, *5*(1), 45–56. https://doi.org/10.1109/t-affc.2013.27

Sinclair, B. (2021, April 28). *Microsoft game revenues up 50% in Q3*. Gameindustry.biz. https://www.gamesindustry.biz/microsoft-game-revenues-up-50-percent-in-q3

Tangkui, R., & Keong, T. C. (2021). The effects of digital game-based learning using Minecraft towards pupils' achievement in fraction. *International Journal on E-Learning Practices*, *4*, 76–91. https://doi.org/10.51200/ijelp.v4i.3427

Tze, V. M., Daniels, L. M., & Klassen, R. M. (2015). Evaluating the relationship between boredom and academic outcomes: A meta-analysis. *Educational Psychology Review*, *28*(1), 119–144. https://doi.org/10.1007/s10648-015-9301-y

Ünlüsoy, A., de Haan, M., Leseman, P. M., & van Kruistum, C. (2010). Gender differences in adolescents' out-of-school literacy practices: A multifaceted approach. *Computers & Education*, *55*(2), 742–751. https://doi.org/10.1016/j.compedu.2010.03.007

Vogel-Walcutt, J. J., Fiorella, L., Carper, T., & Schatz, S. (2012). The definition, assessment, and mitigation of state boredom within educational settings: A comprehensive review. *Educational Psychology Review*, *24*(1), 89–111. https://doi.org/10.1007/s10648-011-9182-7

WHIMC. (n.d.) What-If Hypothetical Implementations using Minecraft. https://whimcproject.web.illinois.edu/

Yu, Z. (2019). A meta-analysis of use of serious games in education over a decade. *International Journal of Computer Games Technology*, *2019*, 1–8. https://doi.org/10.1155/2019/4797032

Publisher's Note

Research and Practice in Technology Enhanced Learning (RPTEL) **is an open-access journal and free of publication fee.**