Check for updates

"Winning could mean success, yet losing doesn't mean failure"—Using a mobile serious game to facilitate science learning in middle school

Sa Liu[1]*, Brian Grey[2] and Mina Gabriel[2]

[1]Interactive Media, Harrisburg University of Science and Technology, Harrisburg, PA, United States, [2]Department of Computer and Information Sciences, Harrisburg University of Science and Technology, Harrisburg, PA, United States

Researchers have recognized the potential of using Mobile Serious Games (MSG) in teaching various subject matters. However, it is not clear how MSG impacts students differently based on their in-game performance. To fill this gap, the current study examined the MSG "SpacEscape" that teaches middle school students about the solar system through problem-solving activities. To understand whether SpacEscape could facilitate middle school student science learning based on their in-game performance, this study adopted a randomized experimental design and collected pre-test, post-test and game play data from the participants. Independent sample t-test and MANOVA with repeated measures (N = 228) were conducted. The results showed that SpacEscape significantly improved science learning for middle school students. In addition, the game significantly improved students' science knowledge test scores regardless of their in-game performance. Furthermore, students enjoyed playing SpacEscape in the class, and we hope this study will inform the direction of future study in the field.

KEYWORDS

middle school science, problem-solving, learning performance, mobile serious game, mixed methods

1. Introduction

Under the traditional, didactic lecturing approach, students are passively participating during their learning, which has been proven to be far less effective at teaching the desired content to students (Bransford et al., 2000). Based on the constructivist theory (Piaget, 2013), a learner must be an active participant in learning as the process of learning new information necessarily involves integrating the new knowledge into the learner's previous background knowledge. Scholars have suggested that humans are naturally hardwired to find active learning enjoyable (Gee, 2003), which makes learning more effective, and learners can apply what they are learning to real-world situations (Hartikainen et al., 2019).

The concept of "Serious Games" (SG) or teaching pedagogical content via a game (e.g., board game, video game, etc.) is developed based on constructivist ideology. In Abt (1987), Abt first defined SG as games that are created for "non-entertainment purposes" (p. 9). Different from gamification, which has visible learning goals (Kalogiannakis et al., 2021) and

aims to "alter a contextual learner behavior or attitude" (Landers, 2014, p. 759), the goal of SG is to motivate learner to play the game and achieve learning without even knowing the learning goals (Landers, 2014; Kalogiannakis et al., 2021).

In theory, games are inherently motivating by presenting players with challenges and enjoyment, meaning that players could engage with them for no external reward or benefit. Thus, when pedagogical content is embedded within a game, the learner would be exposed to and interact with the desired content in a meaningful and active way for many hours (Zhonggen, 2019; Naul and Liu, 2020). Besides motivation and enjoyment, SG could also potentially provide a powerful learning experience for learners where they can experience failure (Larson, 2020) and "learn by failing" (Charsky, 2010, p. 182). Since games are fail-safe, they give the learner an opportunity to fail, start over, and repeat the learning scenario until successfully finishing the game without real-world consequences.

This theory of using games for learning has been applied to analog games both ancient (e.g., early versions of chess teaching 7th century Indian military strategy) and modern (e.g., Monopoly, or The Landlord's Game, teaching the dangers of laisse-fair capitalism). This potential was also recognized by different industries, such as training/learning in military, health care, corporate, non-profit organization, higher education, and K-12 school system (Michael and Chen, 2005; Annetta et al., 2010; Ma et al., 2011; Stanitsas et al., 2019).

For example, early in 1980, the United States Army commissioned Atari to retrofit their hit arcade game Battlezone for training Bradley tank gunners (Clark, 2008). Known as Army Battlezone or The Bradley Trainer, this early attempt at a SG never made it out of the prototype phase but shows the early interest by the United States military in the potential of this new medium. Since then, there have been notable successes and deployments of educational games and simulations such as with the United Sates Air Force's training application Multi-Domain Command and Control (MDC2) Trading Card Game. This application was used to train USAF members to learn about Multi-Domain Operations. A study by Flack et al. (2020) showed that military members consider the game fun to play and can be used as an effective education and training tool. In the K-12 school system, educators also recognize the potential of SG. Particularly, many researchers have looked at using SG to teach science in middle school (Sánchez and Olivares, 2011; Lester et al., 2013; Liu et al., 2015). One example is CRYSTAL ISLAND (Rowe et al., 2009; Spires et al., 2011; Lester et al., 2013; Taub et al., 2017), which is an intelligent game-based learning environment designed to teach microbiology content to eighth grade students. In the environment, "students are required to gather clues, and create and test hypotheses, to solve a mystery" (Taub et al., 2020, p. 641). Alien Rescue is another SG that teaches science knowledge to 4–6th grade students through an immersive 3D environment (Liu et al., 2015; Kang et al., 2017; Liu and Liu, 2020). Empirical studies have shown that these web-based SG learning environments have facilitated student science learning and have also improved learning performance and motivation (Liu et al., 2011, 2019; Spires et al., 2011).

With the advancement of mobile technologies, whether in the form of a cell phone, tablet or Chromebook, it is possible to render a world of extraordinary detail which a player can interact with in real time using a ubiquitous mobile device. This allows for both levels of immersion and simulation that are otherwise unthinkable

through other means. Further, the ubiquity and portability of these devices allow for creativity and integration in ways that would be impossible with a traditional desktop or computer.

Therefore, a few researchers also studied the impact of mobile serious games (MSG) (Sánchez and Olivares, 2011; Su and Cheng, 2015; Baek and Touati, 2017; Tlili et al., 2020). For example, Sánchez and Olivares (2011) investigated three MSGs to assess whether they can be used to help 8 to 10th grade students in Chile to develop problem solving and collaboration skills. They adopted a quasi-experimental design and collected data from 292 students, which showed that students had a better perception of collaboration skills and higher problem-solving score after using these environments. Therefore, they suggested that these MSG-based learning activities may contribute to student learning improvements and recommended that "future work could be to measure student learning to investigate whether the use of MSGs improves learning" (2011). In addition, Tlili et al. (2020) did a comprehensive review on the impact of MSGs. They reviewed 40 studies regarding computer and mobile educational games, and they found that mobile educational games are more friendly, accessible, immersive, and social compared to computer educational games.

Although researchers have recognized the potential of MSG in teaching problem-solving and collaboration, it is not clear how it impacts students differently based on their game performance. Therefore, using SpacEscape, which is an MSG designed by the research team for teaching middle school space science, this study aims to address the following three research questions:

(1) Can SpacEscape help student learning about space science?
(2) Is there a correlation between the student learning performance (i.e., test score) and their in-game performance (i.e., failed or succeeded in the game)?
(3) What are students' perceptions of the MSG?

2. Materials and methods

This study obtained Institutional Review Board (IRB) approval from author's home institution to conduct a study using mixed methods, which "combine the qualitative and quantitative approaches within different phases of the research process" (Clark, 2008, p. 22). The school and students who participated in the study are anonymized to protect their privacy.

2.1. The MSG environment

The study used an MSG named as SpacEscape, which is an Android application created by the research team consisting of faculty and students.[1] The game was developed using the Google Android Studio Tool,[2] and learners could, at the time of the research, download it directly from the Google Play Store. The

---

1 https://harrisburgu.edu/spacescape/

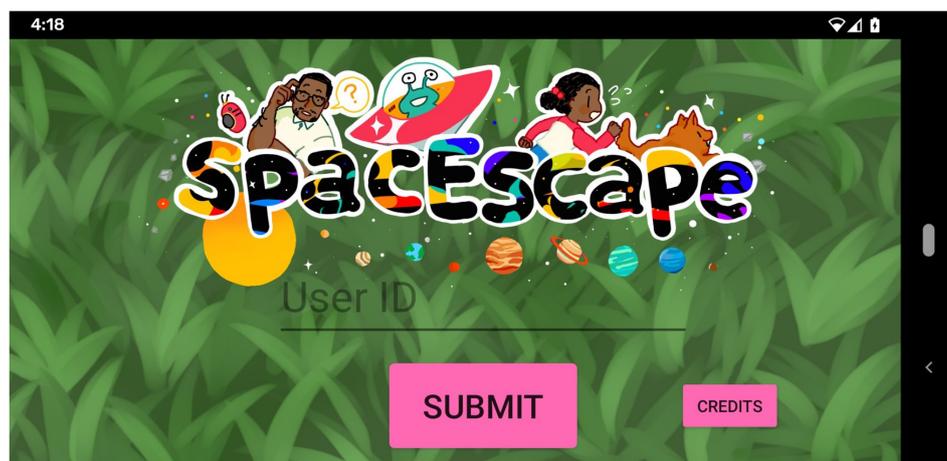2 https://developer.android.com/studio

FIGURE 1
The SpacEscape login screen on a mobile phone.

game is designed based on constructivist theory by first presenting a problem and then requiring the student to be an active participant during the game play to solve the problem. By doing so, SpacEscape aims to fulfill two main objectives: (1) Teach solar system concepts to middle school students; and (2) Increase students' interest in learning about science.

The game starts with an opening video to present the problem the learner will face in the game: a young girl named Lucy is playing with her pet dog Spark in a park; a spaceship passes by and invites Lucy to explore the solar system without informing Spark or her family. Luckily, they leave a walkie-talkie with Spark to communicate with Lucy. After the opening video, students enter the game login screen (see **Figure 1**). They can login using the assigned ID—an anonymous and unique 8-digit numerical number. Once students have logged into the game, they will carry out a search mission for the missing girl Lucy by playing as Spark.

In the game, Lucy is staying on one of the planets or moons in our solar system. While Lucy is playing with aliens in space, she sends out 10 clues to Spark about the characteristics of this astronomical body in a random sequence (i.e., each student might get different clues in the game). Spark needs to locate Lucy based on these clues and conduct research in the game on the different planets and moons. Here are a few example clues: "Spark! It is freezing!!!!," "The gravity here is similar to Earth," and "It is covered with craters." **Figure 2** shows a series of screen-captures highlighting the game environment and interface.

2.2. Participants

Of the 269 sixth grade students who participated in the study (139 boys, 122 girls, and 8 prefer not to tell), only the 228 sixth grade students (119 boys, 101 girls, and 8 prefer not to tell) who finished both pre- and post- test were included in the analysis. These students were all from a public middle school in the northeast region of the United States. The school has 891 students in grades 6–8 with a student-teacher ratio of 14 to 1 and has a good technology infrastructure. Specifically, the school is equipped with
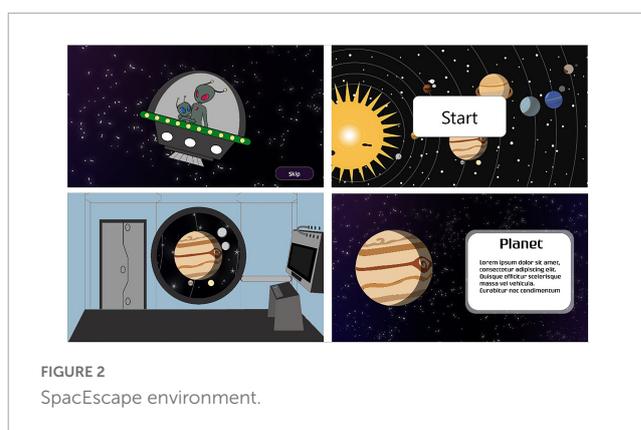

FIGURE 2
SpacEscape environment.

Google Chromebooks (i.e., one Chromebook per student), and has its own technology office to support teaching and learning. The overall rating of this middle school is above average compared to other schools in the United States. In addition, according to state test scores, a majority of the students in this school are proficient in math and reading.

2.3. Procedure and study design

To answer the first research question—whether SpacEscape could facilitate student learning, we adopted a randomized experimental design in this study (Kothari, 2004), as it could "increase the accuracy with which the main effects and interactions can be estimated (p. 40)," and "provide protection, when we conduct an experiment, against the effect of extraneous factors by randomization (p. 40)," while we further "eliminate the variability due to extraneous factor(s) from experimental error" (p. 40).

During the study, students were randomly divided into control and experimental groups based on their student ID. All students had attended their normal science curriculum which covered the solar system before the experiment. They also finished a pre-test on their space science knowledge a day before the experiment. The

experiment was carried out during a supplement science activity class period (40 min). The experimental group students ($n = 107$, 67 boys, 36 girls, and four prefer not to tell) were given a Google Chromebook with SpacEscape pre-installed for them to play during the class period, while the control group students ($n = 121$, 51 boys, 66 girls, and four prefer not to tell) did not have access to the game but were free to conduct other activities related to solar system using their Chromebooks (e.g., reading books, browsing the internet, or watching videos on science topics). All students were given a post-test on their space science knowledge after the supplement science activity class.

To answer the second question, whether there is a significant positive correlation between the student learning performance (i.e., test score) and their in-game performance, we recorded students' play data in the game. Specifically, how many rounds the student play the game and their game performance score—one point for a successful and zero for a failed game play.

To answer the third research question, students' perception of the MSG, researchers added two open-ended questions during the post-test. Qualitative analysis was conducted using the collected data. Specifically, using an open-source, web-based digital texts analysis tool named Voyant (Stéfan and Rockwell, 2016), we examined students' perception on the MSG.

2.4. Instruments

The pre- and post-test on space science knowledge had 10 identical multiple choices questions that were adopted and modified based on previous studies on SG and middle school science about the solar system (Liu et al., 2011, 2014, 2019). The test is designed to measure student factual knowledge on the planets. Each question worth 10 points, and the test is worth 100 points. The Cronbach's alpha for the instrument was 0.76 for pre-test and 0.80 for post-test for this sample, which suggests that the instrument is reliable and consistent in measuring student science knowledge at both time points. One example of a question is:

"Which of these worlds is a gas giant?
A. Saturn.
B. Earth.
C. Pluto.
D. Not Sure."

In addition, as we mentioned above, to understand students' perception on the MSG, the study asked students two open-ended questions during the post-test, "What do you like about the game?" and "How would you improve the game?"

3. Results

Collected data were analyzed using both quantitative and qualitative methods. Specifically, IBM SPSS Statistics software package version 28 was used to conduct descriptive analysis, independent sample t-test and MANOVA with repeated measures.

The results for the above three research questions (RQs) are discussed in this section.

3.1. RQ1: can SpacEscape help student learning about space science?

The independent-sample t-test was used to compare the effects of group differences between the experimental and control groups. For the control group, the average pre-test score was 41.57 (SD = 19.10) points and for the experimental group, the average pre-test score was 44.01 (SD = 20.18) points, see Table 1. There was not a significant difference in the learners' science knowledge with the pre-test mean scores of the two groups at 0.05 alpha level; [$t(226) = -0.941$, $p > 0.05$]. This suggests that the performance of both groups was equivalent at the beginning of this study.

As for the post-test, the average post-test score for the control group was 50.58 (SD = 21.303) points, and the average post-test score for the experimental group was 63.55 (SD = 19.870) points. The independent sample t-test suggested that the difference between these two post-tests means was significant; [$t(226) = -4.736$, $p < 0.001$, see Table 2].

Findings on the science knowledge test scores indicated that both the experimental and control groups were similar in their performance before playing the MSG. After the treatment (playing the MSG SpacEscape during the class period), the experimental group showed significant improvement in their science knowledge test scores compared to the control group.

To examine the main effect of learning using the MSG and different test times as well as the interaction between them, a multivariate analysis of variance (MANOVA) with repeated measures was used in addition to the independent sample T-test. See Table 3.

The Wilks' Lambda confirmed again that there was a statistically significant difference in academic performance based on the science knowledge pre and post-test, [$F(1,226) = 133.755$, $p < 0.001$; Wilk's Lambda = 0.628, partial $\eta^2 = 0.372$]. In addition, the multivariate test indicated that there was a significant interaction between the test time and MSG playing, [$F(1,226) = 18.187$, $p < 0.001$; Wilk's Lambda = 0.926, partial $\eta^2 = 0.074$]. This suggests that the increase in student science

TABLE 1 Independent-sample *t*-test comparing means of pre-test scores between experimental and control groups.

Group	*n*	Mean	SD	*t*	df	Sig. (2-tailed)
Control group	121	41.57	19.10	−0.941	226	0.348
Experimental group	107	44.01	20.18			

TABLE 2 Independent-sample *t*-test comparing means of post-test scores between experimental and control groups.

Group	*n*	Mean	SD	*t*	df	Sig. (2-tailed)
Control group	121	50.58	21.303	−4.736	226	<0.001
Experimental group	107	63.55	19.870			

TABLE 3 MANOVA with repeated measures (main and interaction effects of playing MSG and test-time).

Effect		Value	F	Hypo thesis df	Error df	Sig.	Partial η^2
Time	Wilks' Lambda	0.628	133.755	1,000	226.000	<0.001	0.372
Time* MSG	Wilks' Lambda	0.926	18.187	1,000	226.000	<0.001	0.074

*Means the interaction between two variables in SPSS.

knowledge scores for the experimental group was due to the significant effects of playing the MSG.

3.2. RQ2: is there a correlation between the students' learning performance (i.e., test scores) and their in-game performance (i.e., failed or succeed in the game)?

For the experimental group, 107 students played the game a total of 315 rounds during one class period, resulting in an average of 2.94 game play rounds per student. Two students played the game for more than 10 rounds during the 40-min allotted play session. We considered the problem solved if the student finds Lucy's correct location once. For example, if a student played the game for three rounds, but only found Lucy once (i.e., succeeded once and failed twice), we still consider the student solved the problem in the game.

Therefore, among all the 107 students, 61 of them (57%, including 38 boys, 22 girls, one prefers not to tell) found the correct location for Lucy and solved the problem in the game, while 46 (43%, including 29 boys, 14 girls, and three prefer not to tell) did not find the solution—more students succeeded in completing the game than not, for both gender groups. See **Figure 3**.

We also compared the differences in the science test scores between these two group of students in the experimental group (i.e., succeeded in the game versus failed in the game). For students who succeeded in the game, the average pre-test score is 47. 21 points and the average post-test score is 68.36 points, resulting in an average score increase of 21.15 points. For students who were not able to locate Lucy in the game, the average pre-test score is 39.78 points, and the average post-test score is 57.17 points, resulting in an average score increase of 17.39 points. See **Figure 4** for the test score comparison among all three groups.

In addition to the above simple descriptive analysis, the independent-sample t-test was also used to compare the effects of group learning between students who successfully completed the game (i.e., succeed group) and those who failed to do so (i.e., failed group). For the succeed group, the average pre-test score was 47.21 (SD = 21.06) points and for the failed group, the average pre-test score was 39.78 (SD = 18.31) points, see **Table 4**. There was not a significant difference in science knowledge pre-test mean scores of the two groups at 0.05 alpha level; [$t(105) = 1.909$, $p > 0.05$]. This suggests that performance of both groups was equivalent at the beginning of this study.

As for the post-test, the average post-test score for succeed group was 68.36 (SD = 18.09) points, and the average post-test score for failed group was 57.17 (SD = 20.51) points. The independent sample t-test suggested that the difference between these two post-tests means was significant; [$t(105) = 2.989$, $p < 0.05$, see **Table 5**].

The findings on the science knowledge test scores indicated that students in both the succeed and failed groups were similar in their performance before playing the MSG. After the treatment (playing the MSG SpacEscape during the class period), the succeed group showed a significant improvement in their science knowledge test scores when compared to the failed group. On the other hand, the failed group showed less improvement in the science knowledge post-test mean scores. The results suggest that the increase in student science knowledge scores for the succeed group could either be due to the significant effects of succeeding in the MSG or due to the test effect, which indicates pre-tests can have motivational and teaching functions for learners (Hartley, 1973; Marsden and Torgerson, 2012).

To examine the main effect of in-game performance in the MSG and different test-time as well as the interaction between them, the multivariate analysis of variance (MANOVA) with repeated measures was used in addition to the independent sample t-test, see **Table 6**.

The Wilks' Lambda confirmed again that there was a statistically significant difference in academic performance based on science knowledge pre and post-test scores, [$F(1,105) = 112.026$, $p < 0.001$; Wilk's Lambda = 0.484, partial $\eta^2 = 0.516$]. However, the multivariate test indicated that there was not a significant interaction between the test time and in-game performance, [$F(1,105) = 1.064$, $p = 0.305$; Wilk's Lambda = 0.990, partial $\eta^2 = 0.010$]. These findings imply that succeeding in the MSG had no significant impact on improving students' science knowledge test performance.

Findings on game performance differences indicated that it did not have a significant effect on student learning. The findings indicated that both the succeed and failed groups showed significant improvement in their science knowledge achievement. Although the succeed students performed better than students who failed in the game, the differences were not significant considering the test effect.

3.3. RQ3: what are students' perceptions of the MSG?

To understand students' perception of the MSG, the study asked students two open-ended questions during the post-test, "What do you like about the game?" and "How would you improve the game?"

For the first question, the responses contain 1,728 total words, with an average words per sentence of 6.7. Voyant Tool (Stéfan and Rockwell, 2016) indicated that the most frequent words students mentioned are planets (64), learn (57), like (53), fun (18), and game (18), see **Figure 5**. We further examined these popular words to see how students talked about them, see **Table 7**. For example, when students mentioned planets, they talked about "learn about the planets," "choose different planets you wanted," "click the planets," etc. In addition, analyzing the links between these words, the data
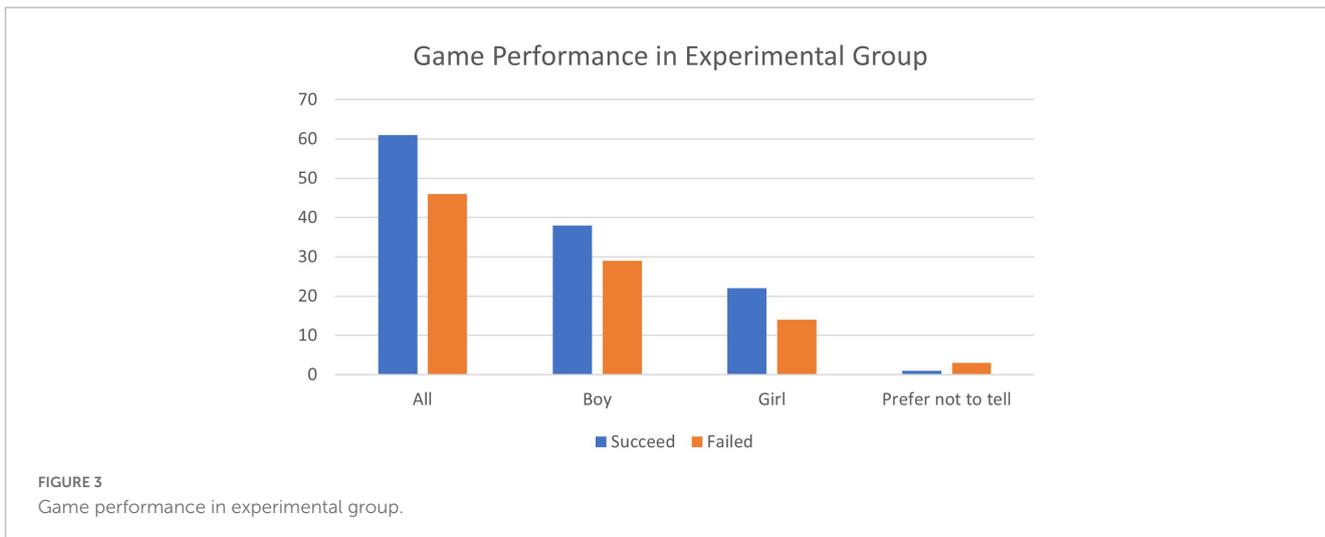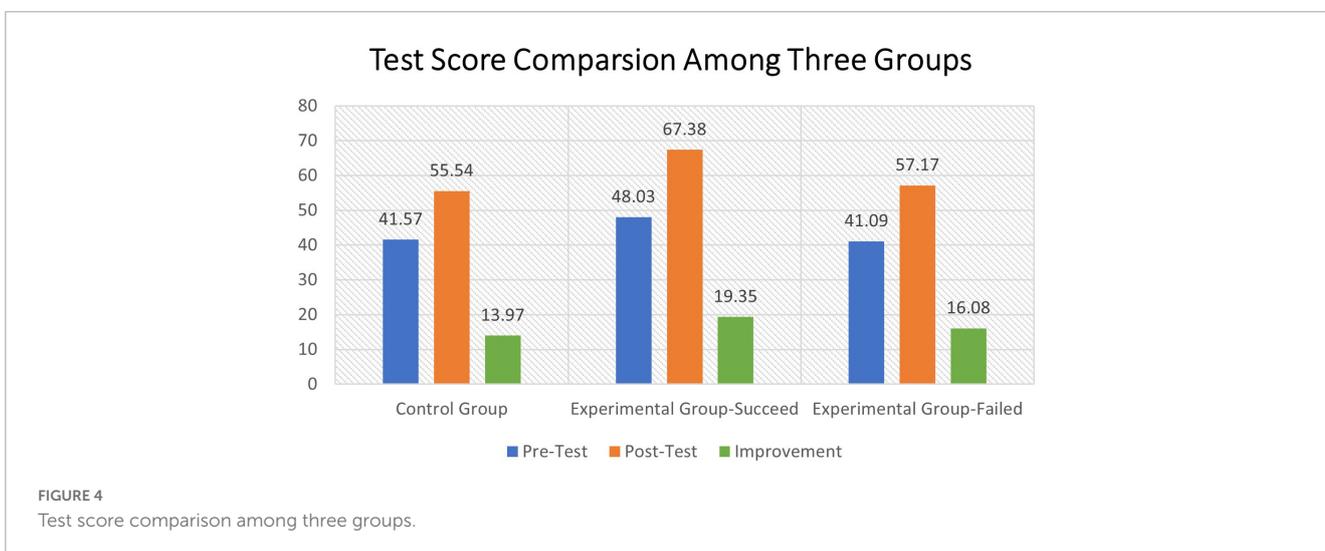
FIGURE 3
Game performance in experimental group.

FIGURE 4
Test score comparison among three groups.

TABLE 4 Independent-sample t-test comparing means of pre-test scores between succeed and failed groups.

Group	n	Mean	SD	t	df	Sig. (2-tailed)
Succeed group	61	47.21	21.06	1.909	105	0.059
Failed group	46	39.78	18.31			

TABLE 5 Independent-sample t-test comparing means of post-test scores between succeed and failed groups.

Group	n	Mean	SD	t	df	Sig. (2-tailed)
Succeed group	61	68.36	18.09	2.989	105	0.003
Failed group	46	57.17	20.51			

TABLE 6 MANOVA with repeated measures (main and interaction effects of in-game performance and test-time).

Effect		Value	F	Hypo thesis df	Error df	Sig.	Partial η^2
Time	Wilks' Lambda	0.484	112.026	1,000	105.00	< 0.001	0.516
Time* Performance	Wilks' Lambda	0.990	1.064	1,000	105.00	0.305	0.010

*Means the interaction between two variables in SPSS.

also showed that students liked the game and liked to learn about the planets and facts in the game, see **Figure 6**.

Admittedly, there were a few students (n = 5; 1.9%) that indicated that they did not like the game. However, based on other the responses from the first question, the feedback was overwhelmingly positive. Students liked the game because they could solve the problem and learn about the solar system while having fun. As for how they would improve the game, students provided various feedback, which can be categorized into the following types: (1) Content; (2) Difficulty level; (3) Interaction; and (4) Glitches; see **Table 8** for the frequency and examples on each feedback type.

Based on the data, students would like to improve the interaction in the game the most—108 students talked about this topic in their feedback, followed by feedback regarding the content of the game, such as feedback on the graphics, including more
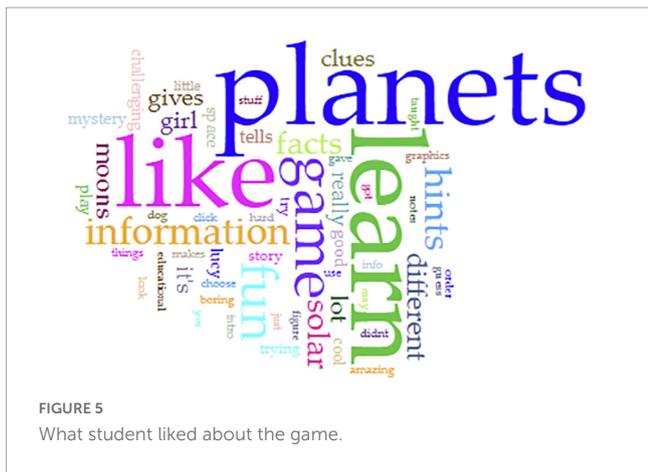
FIGURE 5
What student liked about the game.

TABLE 7 How students used popular words in phrases.

| Popular words (frequency)	Used in phrases
Planets (64)	"You get to learn about each *planet*," "I like learn about the *planets*," "You could choose different *planets* you wanted."
Learn (57)	"I like how we could just click on a planet and *learn* about it," "It taught well and I *learn* a lot"; "It helped me *learn*."
Like (53)	"I *like* the uniqueness of trying to guess the planets," "It is kinda *like* a detective game," "I *like* all the facts".
Fun (18)	"It's *fun* and it gives you info," "it was *fun* to figer out wrer the girl was." "it was *fun* and had a nice story."
Game (18)	"I enjoyed the mystery of the *game* and the clues to solve the puzzle." "I like how the *game* made you read and learn before you could just guess and get clues." "It was a really good *game*. I like how you learn about some of the planets and other kinds of stuff. I might of looked up what Color the planets was, but so far it was amazing *game*!"
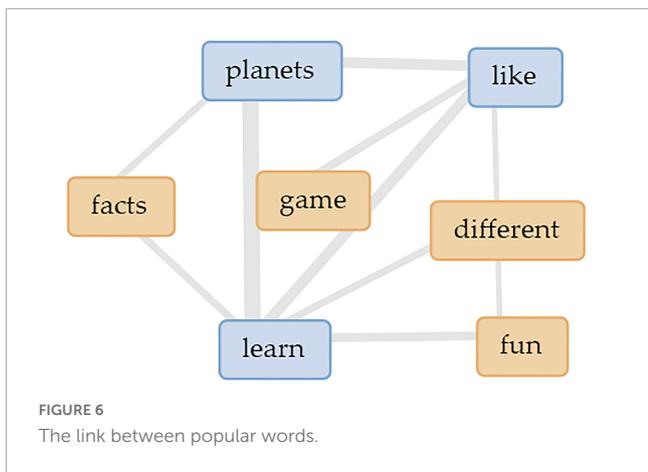
FIGURE 6
The link between popular words.

specific clues, and extra story lines. Students also pointed out one glitch in the game, which automatically logged the players out when they finished a game. A few students also mentioned the difficulty

TABLE 8 Frequency and examples of each feedback type.

Feedback type (frequency)	Examples
Content (85)	"better graphics," "If the clues were more specific," "Add more story."
Difficulty level (9)	"Make the questions easier." "Make it harder and list the clues instead of giving them one at a time"; "To make it easier to find Lucy."
Interaction (108)	"I would improve the game by if you had a life and then if you would die you would have to restart and then follow your same tracks and get farther." "Giving guesses right off the bat," "Make people be able to click on planets multiple times".
Glitches (33)	"Make it less glitchy," "the lag glitches," "Make it that it won't sign you out."

level of the game, although there was no consensus on whether to make it more difficult or easier.

4. Conclusion and discussion

In this study, we found that both the control group and the experimental group students had a significant improvement in their science test scores. The data also showed that the experimental group had a higher improvement compared to the control group despite the test effect. This indicates that the MSG is an effective tool in supplementing and reinforcing in-class instruction. This finding is encouraging, as it suggests that through problem-solving in the MSG, students indeed improved their performance in the science knowledge test. This is consistent with previous studies on MSGs which found that they may contribute to student science learning improvements (Sánchez and Olivares, 2011). It is worth mentioning that these improvements are not only significant, but also higher than students in the control group in this study.

In addition, further examination of students in the experimental group showed that the experimental succeed group had the highest improvement scores among all three groups, followed by the experimental students who failed in the game and finally the control group students who had the lowest improvement scores. This showed that even students who failed in locating Lucy in the game still learned relevant pedagogical content. Their improvement is not as high as the experimental succeed group, but these differences are not significant. In addition, they had higher post-test and improvement scores than students in the control group despite having lower pre-test scores. This finding is significant and consistent with the literature that failures could be valuable learning experiences for players in SG (Charsky, 2010).

Furthermore, the qualitative data suggests that students in the experiment group enjoyed playing the game, as they played this MSG almost 3 times on average during one class session. Two students even played more than 11 times. They liked to "learn about the planets" and "have fun." This finding is consistent with the literature that games are engaging (Gee, 2003), and provide further

evidence that MSG can improve student learning motivation (Liu et al., 2017). In addition, student feedback indicated that the MSG can be improved regarding content and game interactions. More study is also needed to decide whether SpacEscape improved student motivation in learning about other science topics by engaging students in learning about the solar system.

Admittedly, despite the encouraging findings, this study has its limitations. Firstly, the result is based on data that was collected from one middle school in the United States, which might not be generalized to other institutions. Therefore, having different student groups playing the game to see whether this result can be generalized is important for future study. In addition, this study only looked at pre- and post-test conditions, which were only 1 day apart. It is not clear whether the knowledge gained during this short period would be retained long-term. Therefore, future study needs to include a knowledge retention test, which could be conducted at 1–4 weeks after the play-testing period (Chittaro and Buttussi, 2015; van der Spek, 2011). Finally, since the study was conducted in a real-world classroom, the control group students could conduct various learning activities in the class. To improve on this, future study could have the control group students work on one specific activity to better control the experiment condition, which might increase the reliability of the study. Future research could also examine the difference between SG designed using a problem-based learning approach and SG designed using other pedagogical theories.

In conclusion, the data shows that the MSG successfully improved science learning for middle school students. Particularly, this improvement is significantly higher than students who did not play the game. In addition, the game significantly improved student's science knowledge test scores regardless of their performance in the game. Finally, students enjoyed playing this MSG in the class.

For future studies, researchers can investigate why and how students who failed in the game still significantly improved their test scores. With the encouraging results, we hope this study will inform the direction of future study in the MSG field.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board at Harrisburg University of Science and Technology.

Author contributions

SL drafted the work and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. BG provided the substantial contributions to acquisition, analysis, or interpretation of data for the work. MG made substantial contributions to the conception or design of the work. All authors provide approval for publication of the content.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abt, C. C. (1987). *Serious games*. NewYork, NY: The Viking Press.

Annetta, L. A., Cheng, M. T., and Holmes, S. (2010). Assessing twenty-first century skills through a teacher created video game for high school biology students. *Res. Sci. Technol. Educ.* 28, 101–114. doi: 10.1080/02635141003748358

Baek, Y., and Touati, A. (2017). Exploring how individual traits influence enjoyment in a mobile learning game. *Comput. Hum. Behav.* 69, 347–357. doi: 10.1016/j.chb.2016.12.053

Bransford, J. D., Brown, A. L., and Cocking, R. R. (2000). *How people learn*, Vol. 11. Washington, DC: National academy press.

Charsky, D. (2010). From edutainment to serious games: A change in the use of game characteristics. *Games Cult.* 5, 177–198. doi: 10.1177/155541200935 4727

Chittaro, L., and Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Trans. Vis. Comput. Graph.* 21, 529–538. doi: 10.1109/TVCG.2015.2391853

Clark, V. L. P. (2008). *The mixed methods reader*. Thousand Oaks, CA: Sage.

Flack, N., Voltz, C., Dill, R., Lin, A., and Reith, M. (2020). "Leveraging Serious Games in Air Force Multi-Domain Operations Education: A Pilot Study," in *Proceedings of the ICCWS 2020 15th International Conference on Cyber Warfare and Security*, England.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Comput. Entertain.* 1:20. doi: 10.1145/950566.950595

Hartikainen, S., Rintala, H., Pylväs, L., and Nokelainen, P. (2019). The concept of active learning and the measurement of learning outcomes: A review of

research in engineering higher education. *Educ. Sci.* 9:276. doi: 10.3390/educsci9040276

Hartley, J. (1973). The effect of pre-testing on post-test performance. *Instr. Sci.* 2, 193–214. doi: 10.1007/BF00139871

Kalogiannakis, M., Papadakis, S., and Zourmpakis, A. I. (2021). Gamification in science education. A systematic review of the literature. *Educ. Sci.* 11:22. doi: 10.3390/educsci11010022

Kang, J., Liu, S., and Liu, M. (2017). "Tracking students' activities in serious games," in *Learning and Knowledge Analytics in Open Education: Selected Readings from the AECT-LKAOE 2015 Summer International Research Symposium*, eds F. Q. Lai and J. D. Lehman (Berlin: Springer International Publishing). doi: 10.1007/978-3-319-38956-1_10

Kothari, C. R. (2004). *Research methodology: Methods and techniques.* New Delhi: New Age International.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simul. Gaming* 45, 752–768. doi: 10.1177/1046878114563660

Larson, K. (2020). Serious games and gamification in the corporate training environment: A literature review. *TechTrends* 64, 319–328. doi: 10.1007/s11528-019-00446-7

Lester, J. C., Ha, E. Y., Lee, S. Y., Mott, B. W., Rowe, J. P., and Sabourin, J. L. (2013). Serious games get smart: Intelligent game-based learning environments. *AI Magaz.* 34, 31–45. doi: 10.1609/aimag.v34i4.2488

Liu, M., Horton, L., Olmanson, J., and Toprac, P. (2011). A study of learning and motivation in a new media enriched environment for middle school science. *Educ. Technol. Res. Dev.* 59, 249–265. doi: 10.1007/s11423-011-9192-7

Liu, M., Kang, J., Lee, J., Winzeler, E., and Liu, S. (2015). "Examining through visualization what tools learners access as they play a serious game for middle school science," in *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, eds C. S. Loh, Y. Sheng, and D. Ifenthaler (Switzerland: Springer). doi: 10.1007/978-3-319-05834-4_8

Liu, M., Kang, J., Liu, S., Zou, W., and Hodson, J. (2017). "Learning analytics as an assessment tool in serious game: A review of literature," in *Serious games and edutainment applications*, eds M. Ma and A. Oikonomou (New York, NY: Springer), 537–563. doi: 10.1007/978-3-319-51645-5

Liu, M., Liu, S., Pan, Z., Zou, W., and Li, C. (2019). Examining science learning and attitude by at-risk students after they used a multimedia-enriched problem-based learning environment. *Interdisc. J. Prob. Based Learn.* 13:6. doi: 10.7771/1541-5015.1752

Liu, M., Rosenblum, J. A., Horton, L., and Kang, J. (2014). Designing science learning with game-based approaches. *Comput. Sch.* 31, 84–102. doi: 10.1080/07380569.2014.879776

Liu, S., and Liu, M. (2020). The impact of learner metacognition and goal orientation on problem-solving in a serious game environment. *Comput. Hum. Behav.* 102, 151–165. doi: 10.1016/j.chb.2019.08.021

Ma, M., Oikonomou, A., and Jain, L. C. (2011). *Serious games and edutainment applications.* London: Springer. doi: 10.1007/978-1-4471-2161-9

Marsden, E., and Torgerson, C. J. (2012). Single group, pre-and post-test research designs: Some methodological concerns. *Oxford Rev. Educ.* 38, 583–616. doi: 10.1080/03054985.2012.731208

Michael, D. R., and Chen, S. L. (2005). *Serious games: Games that educate, train, and inform.* Boston: Muska & Lipman/Premier-Trade.

Naul, E., and Liu, M. (2020). Why story matters: A review of narrative in serious games. *J. Educ. Comp. Res.* 58, 687–707. doi: 10.1177/0735633119859904

Piaget, J. (2013). *The construction of reality in the child* (Vol. 82). Oxfordshire: Routledge.

Rowe, J., Mott, B., McQuiggan, S., Robison, J., Lee, S., and Lester, J. (2009). "Crystal island: A narrative-centered learning environment for eighth grade microbiology," in *Proceedings of the workshop on intelligent educational games at the 14th international conference on artificial intelligence in education*, London.

Sánchez, J., and Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Comput. Educ.* 57, 1943–1952. doi: 10.1016/j.compedu.2011.04.012

Spires, H. A., Rowe, J. P., Mott, B. W., and Lester, J. C. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *J. Educ. Comput. Res.* 44, 453–472. doi: 10.2190/EC.44.4.e

Stanitsas, M., Kirytopoulos, K., and Vareilles, E. (2019). Facilitating sustainability transition through serious games: A systematic literature review. *J. Clean. Prod.* 208, 924–936. doi: 10.1016/j.jclepro.2018.10.157

Stéfan, S., and Rockwell, G. (2016). *Voyant Tools.* Boston: Muska & Lipman/Premier-Trade.

Su, C. H., and Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *J. Comput. Assist. Learn.* 31, 268–286. doi: 10.1111/jcal.12088

Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., and Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Comput. Hum. Behav.* 76, 641–655. doi: 10.1016/j.chb.2017.01.038

Taub, M., Sawyer, R., Smith, A., Rowe, J., Azevedo, R., and Lester, J. (2020). The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment. *Comput. Educ.* 147:103781. doi: 10.1016/j.compedu.2019.103781

Tlili, A., Essalmi, F., Jemni, M., Chen, N. S., Huang, R., and Burgos, D. (2020). "The evolution of educational game designs from computers to mobile devices: A comprehensive review," in *Radical Solutions and eLearning: Practical Innovations and Online Educational Technology*, ed. D. Burgos (Singapore: Springer). doi: 10.1007/978-981-15-4952-6_6

van der Spek, E. D. (2011). *Experiments in serious game design: a cognitive approach.* Ph.D. thesis. Utrecht: University Utrecht.

Zhonggen, Y. (2019). A meta-analysis of use of serious games in education over a decade. *International J. Comput. Games Technol.* 2019, 4797032. doi: 10.1155/2019/4797032