

Mars, Minecraft, and AI: A Deep Learning Approach to Improve Learning by Building

Samuel Hum, Evan Shipley, Matt Gadbury, H Chad Lane, and Jeffrey Ginger

University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA

{hum3, evanjs3, gadbury2, hclane, ginger}@illinois.edu

Abstract. Middle school students learned about astronomy and STEM concepts while exploring *Minecraft* simulations of hypothetical Earths and exoplanets. Participant groups ($n = 24$) were tasked with building feasible habitats on Mars. In this paper, we designed a new coding scheme for assessment of habitats that was used to build novel multi/mixed-input AI models. Using Spearman's rank correlations, we found that our coding scheme was reliable with regards to team size, face-to-face instruction time, and self-explanation scores. We took an exploratory approach to analyzing image and block data to compare seven different input conditions. Using one-way ANOVA, we found that the means of the conditions were not equal for accuracy, precision, recall, and F1 metrics. A post hoc Tukey HSD test found that models built using images only were statistically significantly worse than conditions that used block data on the metrics. We also report the results of optimized models using block only data on additional Mars bases ($n = 57$). The results demonstrate the nuances of assessing work products, such as the information needed for humans and computers to make similar judgements.

Keywords: Coding Scheme, Artificial Intelligence, Habitat Building, Informal Learning, Minecraft

1 Introduction

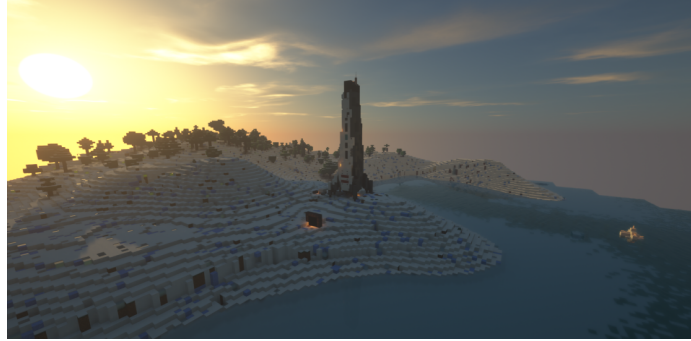
Minecraft has become one of the most popular games in the world. Currently, *Minecraft* has over 140 million monthly users, and 21.21% of daily user traffic in *Minecraft* originates from the U.S. [14]. *Minecraft* is available on tablets, smartphones, consoles, and personal computers. *Minecraft* is a digital environment many adolescents have an established interest in and that can be leveraged in educational settings. Even with little experience, researchers have shown learners master controls quickly and can effectively engage content when *Minecraft* is used in STEM learning environments [7]. Given the ubiquity of the game,

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this contribution is published in the 25th International Conference on Artificial Intelligence in Education, and is available online at https://doi.org/10.1007/978-3-031-64299-9_39

as well as the identified opportunities to conduct research on learning and motivation [8], *Minecraft* deserves attention regarding effectiveness for promoting learning and motivation in formal and informal learning environments.

The data used in this study comes from the What-if Hypothetical Implementations in Minecraft (WHIMC) project that uses *Minecraft* as a vehicle for understanding student interest and motivation in exploring STEM content, primarily Astronomy and Earth Science. Learners are presented with a variety of “what-if” scenarios, such as “What if Earth had a colder sun?” (see Figure 1). Students are tasked with exploring these hypothetical worlds in *Minecraft* and are encouraged to use in-game tools to measure science variables and formulate observations and hypotheses about how and why these environments differ from our own Earth. These “what-if” questions pose novel scenarios that even seasoned *Minecraft* players are unlikely to have experienced in-game previously. Working through counterfactual examples of phenomena in science has shown promise of enhancing learning above and beyond studying strictly factual information [10].

Fig. 1. *Minecraft* simulation demonstrating the impact of being tidally locked on an Earth with a colder sun.



A major part of the experience for learners in this *Minecraft* project is the collaborative build phase, where they work with peers to construct habitats on Mars that humans might inhabit. A *Minecraft* habitat or base is an assigned region under a few hundred blocks in any dimension where participants work in teams to build. Constructing habitats requires integrating the STEM content learners have encountered throughout the exploration phase of the experience.

This paper proposes a novel scoring scheme for work products designed by middle school age students in *Minecraft* and a novel method for work product assessment in educational video games. We propose the use of multi/mixed-input models in *Minecraft* that takes data from different modalities such as in-game images and material data to assess learning. Although multimodal data (MMD) has been used in educational research and video game environments, AI models

using MMD has not been applied to student work product assessment [12]. Thus, our paper is guided by the following questions and hypotheses:

RQ1. What criteria constitute a reliable scoring scheme for Mars habitat builds in *Minecraft*, and to what extent does the scoring scheme hold across differences in group sizes and amount of face-to-face time with instructors? H1: A comprehensive scoring scheme based on inclusion of essential aspects of habitability will contribute to a reliable scoring scheme. Group size and face-to-face time with instructors will not contribute to significant differences between scores.

RQ2. What, if any, relationship do Mars habitat scores have with learning outcomes? H2: Higher scores on Mars habitats will positively correlate with knowledge assessment scores.

RQ3. What type or combination of data input accessible in Minecraft should be used for artificially intelligent detectors of habitat knowledge growth? H3: Incorporating models built using a combination of multiple forms of data concatenated together will outperform models solely built using one form of data. These models can extrapolate information from different aspects of the bases for more accurate predictions that would be impossible for models only built using a single type of data.

2 Background

We are interested in analyzing what are called, “student work products”, referring to specific, task-driven designs and creations, also called “artifacts”. In the end, a student has created a product manifesting their conceptual understanding of the content they interacted with throughout a learning experience. Work products as means of assessing learner knowledge and creativity emerged from Constructionism and the idea that knowledge is produced through students’ creative and collaborative work [5]. This notion has been fully embraced by the Maker Movement and the desired approach to better understand what tools and activities are contributing to learning and other desirable outcomes (*i.e.* creativity) [11]. Another positive outcome from students creative work products is those who do not consider themselves aligned with the given subject (e.g., “science-minded”) may become more engaged due to the empowerment that creating a work product brings them [3]. There is an increasing need to examine digital making as it becomes more integrated into our formal and informal learning environments. With the discussions of AI in education, we seek to understand how models can be designed around digital making assessment.

Artifacts have been used as a means of assessing integrated knowledge a student has acquired throughout a learning experience. In a review of instructional techniques and interventions to promote integration of multiple texts, Barzilai et al. [2] identified collaborative discussion and practice as the most prominent means of instruction for integrating multiple texts and written essays as the most common medium for analyzing text integration. The stark differences in learning environments, techniques, and requirements of tasks make generalizing results difficult, however, the strategies described constitute best practices and

are effective at aiding students in integrating conceptual knowledge. We seek to move beyond written artifacts, though, and contribute to this domain by examining how collaborative discussion and practice impacts integration of knowledge evidenced through collaborative construction of habitats in a customized, STEM-focused *Minecraft* learning environment.

Minecraft has been identified as a medium for learning that offers a lot of potential for educators, researchers, and designers who have an interest in assessing learning and other motivational variables [8]. One *Minecraft* study measured student creativity with the Creative Intelligence Task (CREA) before and after creating “machinima” of their dream house in *Minecraft* [4]. Results showed significant increases in CREA scores regardless of score received on the machinima. An observational study examining learner understanding of urban planning, found that learners from a small town in Brazil incorporated their own interpretations of what matters in a habitat and included additional spaces they deemed important, such as playgrounds [1]. *Minecraft* allows flexibility in design and affords opportunities for learners to combine their knowledge with novel ideas and concepts. A similar study had students identify challenges in their neighborhood and design solutions using *Minecraft*, and findings showed students went from initially considering “loose problems” to address to more structural problems, such as poor lighting, broken roads, and scattered community events [9]. These studies all leverage the affordances of *Minecraft* to allow student autonomy to flourish in addressing problems and designing solutions. To successfully build products in *Minecraft*, creative problem-solving and innovation are key, as students need to plan out details and aspects of their building to fulfill the requirements laid out by the project [6].

3 Methods

3.1 Participants

A total of $n = 48$ middle school age students are included in this study (31% female) with an average age of 11.96 years old. All students participated in 1-week summer camps held in three distinct locations in the West, Midwest, and East United States. Demographic breakdown is as follows: 30% Caucasian, 23.75% African-American, 21.25% PNA, 12.5% Hispanic, 2.5% Asian, 1.25% American Indian, and 7% Other. Consent to participate was obtained from at least one parent/guardian of each participant. Verbal assent was assessed at the beginning of each camp.

A total of $n = 24$ individual bases were made from the participants described above and were analyzed for the coding scheme. The 24 bases include 6 built by groups consisting of 3 students, 12 built by groups consisting of 2 students, and 6 built by individuals. Due to missing block data, 3 bases had to be excluded, resulting in a sample of $n = 21$ bases used for the AI model comparison.

3.2 Materials

Participants were all provided with a laptop, mouse, and an individual account to play *Minecraft: Java Edition*. Participants used the same account for each session of their respective after school program or summer camp. All maps explored by participants were created by our lab and represent simulations of “What if” questions, such as “What if Earth was a moon to a larger planet?”, as well as known exoplanets, planets outside of our solar system (*e.g.* Kepler 186-f). Design of worlds was done in consultation with an astrophysicist and feature extreme conditions, such as high winds, widespread volcanic activity, freezing temperatures, or low gravity, which can all be seen or measured using science tools. As part of the camp curriculum, participants complete pre-written self-explanation questions following their exploration of each in-game world. Each world has three total questions which pertain to the quests, non-player character dialogue, and in-classroom guidance for each world. These questions are scored as either 0, 1, 2, or 3, each showing a level of astronomy explanation and mastery. A score of 0 indicates a nonsense or blank answer, a score of 1 indicates that the answer was either partial, wrong, or low effort, a score of 2 indicates that the answer is correct but has misconceptions getting at something true, and a score of 3 indicates a well-reasoned correct answer with evidence.

Minecraft:Java Edition is an open-world sandbox style digital game, where players can explore vast worlds with varying terrain. Players can interact with every block in the game, collect resources, and build any type of structure. There are different modes of the game, creative and survival, with creative mode providing players with unlimited resources to build, and survival requiring extraction of resources and battling against enemies. Our project relies entirely on creative mode to focus on exploration and building without threat of players’ characters feeling threatened by in-game enemies. The “What if” scenarios and exoplanets were created using “plug-ins”, or coded adjustments to the game that allow for customization and data collection.

3.3 Procedure

The final two sessions of the camp or program, participants were tasked with building a habitat on a Mars map with several features imported in from actual NASA 3D scans, including a 1:2000 scale portion of Valles Marineris and Olympus Mons as well as 1:60 scale adaption of the Jazero Crater. Participant groups were formed based on seating arrangements, existing friendships or by researcher assignment if needed. They were prompted with an introductory video and presentation and then challenged to design a habitat for humans to survive and carry out research on Mars. Participants were told to employ knowledge they learned from exploring all the previous hypothetical worlds and exoplanets to inform how they respond to conditions they can measure on Mars. Participants had around 3 hours to work together and build their habitats. About 40 minutes prior to the end of the last session each group gave a tour of their habitat for the whole class, explaining the problems they addressed and how they solved them as well as what made their habitat special.

3.4 Data Analysis

Habitat Selection Habitats were selected from camps conducted in 2022 to create the habitat coding scheme and experiment with different data inputs for the machine learning models. Selection was a simple random process, where each base, from furthest left on the x-axis to furthest right on the x-axis (west to east on the map) was given a number from one to as many bases as on the map, and three random numbers were generated from that range for each map, providing a total of 24 bases across the 8 camps.

Additional bases were collected in 2023 for a total of 57 bases across 16 camps. The dataset was used to improve existing block data models using feature selection and hyperparameter optimization.

Habitat Coding Scheme An extensive coding scheme was utilized to ensure every part of each group’s habitat was accounted for in a fair and balanced manner, covering all possible aspects of Mars habitat building outlined during the camp process. In total, 11 categories were outlined, with a three-tier system of attaching a score to each category. These tiers are classified from least score to highest score as “Basic”, “Intermediate”, and “Mastered”, each representing a level of application and mastery that the participants of the Mars habitat activity during the camps have shown off. These are awarded starting from nothing, so participants do not lose any points, they are only able to earn them based on the parts of their bases and how each category has been defined. The “Basic” tier awards 0 points in each category and is scored here due to the concept not being present or present but highly unrealistic, such as the component’s presence without stable survival in mind. The “Intermediate” tier awards a number of points halfway between 0 and the maximum for each category and is scored this way due to the concept being present within the habitat, but unfinished, which could be due to time constraints, a smaller team, or a neglect for that concept of survival. The final “Mastered” tier awards the maximum number of points per category, which ranges between 1 point and 1.5 points per category and represents that the team integrated the concept clearly and accurately reflects what would be possible by people establishing a real habitat on Mars.

To ensure that the habitats that are the most accurate and scientifically sound are scored the highest, scoring weights have been applied to multiple categories, with a multiplication factor of 1.5. The weight applies to the categories of concepts that are essential for survival on Mars: atmosphere regulation, protection from radiation, food and water, supply storage, power generation, communications facilities, and rounded structure shape. Categories such as area where the base is built, combating different levels of gravity, health and wellness, and transportation were deemed as important but less essential considerations to the immediate survival of scientists inhabiting a Mars habitat. Six categories describe aspects of the exteriors of student habitats (see Table 1) and five describe interior qualities (see Table 2).

Table 1. Descriptions of the six categories based on the exterior of bases.

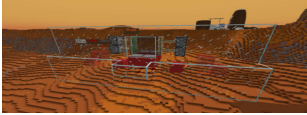
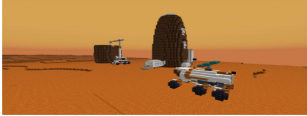
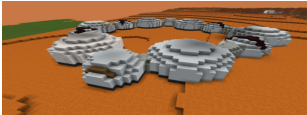
Category	Description	Example
Oxygen Production and Atmosphere Regulation (Category weight: 1.5)	Participants must include a representation of an air regulation system and air locks.	
Radiation Protection (Category weight: 1.5)	Participants must include multiple layers of protection using dense materials specifically to keep radiation from the sun out of the base.	
Power Generation (Category weight: 1.5)	Participants must include nuclear reactors or solar panel arrays as their main source of power.	
Shape of Structure (Category weight: 1.5)	Participants must build their habitats with a rounded or semi-rounded shape to prevent minute toxic particles of dust from penetrating the corners.	
Area Built	Participants will need to build their habitat in what would be considered a “geographically-protected” area to reduce regolith build up and other conditions that would affect the flat lands on Mars.	
Transportation	Participants will need to build transportation for astronauts to traverse Mars or be able to travel to and from Earth. For example, a rover-like vehicle or a launchpad at or near the base.	

Table 2. Descriptions of the five categories based on the interior of bases.

Category	Description	Example
Food and Water (Category weight: 1.5)	Participants must make the supply of food and water easily accessible and not subject to radiation damage.	
Supplies (Category weight: 1.5)	Participants must make their supply storage facilities easily accessible and not subject to radiation damage.	
Health and Wellness	Participants will need to include an infirmary and a psychiatrist's office to deal with injuries, both physical and mental.	
Communications Facilities (Category weight: 1.5)	Participants must include facilities to communicate with Earth.	
Combating Different Level of Gravity	Participants will need to address the lower levels of gravity on the surface of Mars such as including a gym.	

Scoring Process and Reliability To complete the scoring for all 24 habitats, one expert researcher and one novice researcher reviewed each habitat and scored them based on the scheme described above. The novice researcher was trained on the scoring scheme using mock bases, and then both researchers scored the first two bases simultaneously, finding agreement on all aspects. Following this, each researcher scored the next five bases asynchronously and met a week after. Comparing scores, an average agreement of 93% emerged, with a calculated Cohen's kappa of $\kappa = 0.87$, indicating excellent agreement. The remaining 17 habitats were scored independently by the two researchers. Following the scoring process, habitat scores were correlated using a Spearman's rank correlation with team size, face-to-face time, and mean self-explanation score.

Artificial Intelligence Architecture This section describes the seven deep learning models being compared in this paper. The architectures were chosen around the capabilities of the learning environment, the habitat coding categories outlined above, and literature that have demonstrated potential improvements

of mixed and multi-input machines [12,13]. *Minecraft* plug-ins on the server automatically collect instances where students place or remove blocks and can take screenshots of the game. It is impossible, however, for a model to predict all of the categories solely from one input type (food sources cannot be interpreted from aerial images, location cannot be interpreted from underground images, shape cannot be interpreted by block data, etc.). In addition, we predict that models can infer what structures are being built by the block type and frequency of usage, since students often use similar types of blocks when building specific parts of their habitats. Multi/mixed-input models have not been applied to classification of work products in video games [12]. Thus, we are taking an exploratory approach to determine which frameworks and data sources works best for *Minecraft*. The models were designed to take different combinations of the three types of input: aerial images, underground images, and block data for each base.

We designed three baseline models for the three input types. The specific architecture of our baseline models were based off of a study comparing mixed-input models in the medical field that used a similar image and numerical dataset [13]. For both aerial images and underground images, we used a convolutional neural network (CNN). Images in our dataset were resized to 128x128 and each pixel value was normalized. For block data, we used a multi-layered perceptron (MLP). The columns for the dataset were the types of blocks used for all of the groups and for each cell were the number of the block type used by the group normalized.

There were a total of four multi/mixed-input classifiers: aerial and underground image model, aerial and block data model, underground and block data model, and a combination of all three data types. To concatenate the models into a single classifier, we concatenate the output layers together and flatten them as input to another connected layer and then an output layer to get the classification. Before concatenation, we used the same model architectures in the multi/mixed-input classifiers as the ones used in the baseline models.

AI Model Comparison A total of $n = 21$ bases were used to compare different input types for the AI models, 3 were omitted from the dataset due to missing block data. To compare the seven AI frameworks described above, we used 5-fold cross-validation. To handle dataset imbalances we used class weighting and to prevent overfitting we used early stopping. We then ran ANOVA to determine whether the means of the seven models were identical and a post hoc Tukey HSD test to determine which pairwise comparisons between conditions yielded significant differences.

4 Results

4.1 Habitats and Learning

To demonstrate that the scoring process was reliable for all students involved in the 2022 camps, three Spearman’s rank correlations were performed compar-

ing habitat scores to team sizes, face-to-face time, and mean self-explanation score. Assessing the relationship between team size and habitat scores, Spearman’s rank correlation showed a non-significant negative correlation between team sizes and habitat score, $r(22) = -.03$, $p = .89$. This indicates that smaller, or even individual teams, are capable of excelling at the build challenge. The second correlation found a non-significant positive relationship between face-to-face camp instruction time and habitat score, $r(22) = .13$, $p = .54$. This further provides evidence that the build challenge can effectively happen in situations where instructors are not always present. Finally, we found a significant positive correlation between mean self-explanation score and habitat score, $r(22) = 0.51$, $p = .01$. Higher knowledge scores from conceptual knowledge assessments leading up to the build challenge resulted in overall better builds. This could indicate productive engagement with the exploratory phase of the camp is a critical aspect of preparing for the build challenge.

4.2 AI Model Comparison

Model Metrics One-way ANOVAs was conducted to compare the effect of condition on accuracy, precision, recall, and F1 scores. The ANOVAs for all four metrics were significant. For accuracy, $F(6, 378) = 9.82$, $p < .01$. For precision, $F(6, 378) = 6.99$, $p < .01$. For recall, $F(6, 378) = 3.42$, $p < .01$. For F1 score, $F(6, 378) = 6.93$, $p < .01$. Table 3 shows the Tukey HSD test results of the comparisons between conditions on the metrics.

Table 3. Post hoc Tukey HSD results for accuracy, precision, recall, and F1 scores (A = Aerial, U = Underground, B = Blocks). *** $p < .001$, ** $p < .01$, * $p < .05$

Comparison	Accuracy	Precision	Recall	F1
U vs. A	0.04	0.01	0.01	0.01
B vs. A	0.22***	0.18**	0.09	0.16**
A+U vs. A	0.05	0.02	-0.01	0.01
A+B vs. A	0.20***	0.13	0.08	0.13*
U+B vs. A	0.19***	0.15*	0.09	0.14*
A+U+B vs. A	0.25***	0.19***	0.12*	0.18***
B vs. U	0.18**	0.17**	0.08	0.15**
A+U vs. U	0.01	0.01	-0.02	-0.001
A+B vs. U	0.16**	0.12	0.07	0.11
U+B vs. U	0.16*	0.14*	0.08	0.12
A+U+B vs. U	0.21***	0.19***	0.11	0.17**
A+U vs. B	-0.17**	-0.16**	-0.10	-0.15**
A+B vs. B	-0.02	-0.05	-0.01	-0.03
U+B vs. B	-0.03	-0.03	-0.004	-0.02
A+U+B vs. B	0.03	0.02	0.03	0.02
A+B vs. A+U	0.15*	0.11	0.09	0.12
U+B vs. A+U	0.14*	0.13	0.10	0.12
A+U+B vs. A+U	0.20***	0.18**	0.13*	0.17**
U+B vs. A+B	-0.01	0.02	0.01	0.01
A+U+B vs. A+B	0.05	0.06	0.04	0.05
A+U+B vs. U+B	0.06	0.05	0.03	0.05

4.3 Model Optimization

A total of $n = 57$ bases were used for feature selection and finding the optimal amounts of layers and nodes for block only models for each category in the habitat coding scheme. We decided to optimize the block only models, since our comparison found that block only models performed just as well as our best models for the different input types. In addition, using block only data could reduce deployment issues such as lag associated with collecting images while students are on our server and collecting images that potentially do not fully represent the students' bases. For feature selection, we used correlations between each building material used by groups across all camps and a cutoff of $r = .95$ to remove all but one of highly correlated building materials. We removed a total of 267 highly correlated block types leaving a total of 428 building materials that were included in the dataset. Then, using a randomized search with 50 iterations with varying numbers of hidden layers (minimum of 1 and maximum of 4 layers) and nodes in each layer (minimum of 25 and maximum of 500 nodes), we ran a 5-fold cross validation for each habitat coding category (results shown in Table 4).

Table 4. Results of best models regarding accuracy for each category using randomized search 5-fold cross validation.

Feature	Accuracy	Precision	Recall	F1
Oxygen Production/Atmosphere Regulation	.72	.57	.45	.81
Radiation Protection	.62	.57	.41	.64
Power Generation	.77	.79	.58	.77
Communication	.74	.78	.57	.67
Shape of Structure	.62	.54	.45	.72
Area Built	.76	.87	.58	.72
Transportation	.61	.57	.47	.64
Combating Lack of Gravity	.56	.55	.55	.54
Food and Water	.49	.48	.35	.60
Supplies	.67	.63	.47	.65
Health and Wellness	.62	.57	.47	.64
Average	.65	.63	.47	.68

5 Discussion

In this paper we discuss our research to develop a coding scheme to assess student Mars habitats and a habitat classifier based on that coding scheme. We discuss our findings in relation to the research questions hypotheses that guided our work.

Consistent with our first two hypotheses, we conclude that the scoring method for the Mars habitats provides a fair and reliable means of assigning a numerical score to any given base built by camp participants. Because team size and face-to-face time had a non-significant impact on the scores that participants received, we can infer that varied contexts do not substantially impact learners' abilities to construct meaningful habitats. This was further proven by correlating the mean self-explanation score with the habitat scores, where the higher a group was able to score on the questions, the better their habitat scored overall. This indicates that when participants integrate more astronomy knowledge their builds tend to be more comprehensive and accurate habitat. Progressing through the exploration phase of the STEM-focused *Minecraft* experience appears to be a necessary condition for constructing habitats that address core concepts of habitability on extraterrestrial environments.

Contrary to our third hypothesis, a model built with a combination of data sources (aerial and underground images) was significantly outperformed by the model using only block data. This result is surprising, since this approach provides the same visual information that human coders are given when assessing habitats. We believe that although human coders are only given visuals of the student bases, coders connect their understanding of the building activity and prior knowledge of *Minecraft* block types and qualities to make their judgments. Thus, providing models with block data provides necessary context to capture human understanding of the domain and task, which has implications for data

collection and input for future work product classifiers. Similarly, models using a combination of image and block data did not significantly outperform the model only using block data. Our best models used a combination of all three forms of data, which shows potential for this technique to be used in future work product classification work. The non-significant improvement compared to the block only model and deployment considerations, however, informed our decision to optimize the block only models. A possible limitation of using block only data is that *Minecraft* head blocks are not differentiated. These blocks are customized to look like objects (*e.g.* computers, cabbages, etc.) and are often used by students when designing their bases, thus not differentiating these blocks could impact the predictive capabilities of block only models on certain categories.

6 Future Work

Our optimized models have been integrated with a pedagogical agent and will be used in usability studies to provide in-process habitat feedback with students at our future camps. Automated feedback formulated from classifications made by our AI models should improve student building behaviors by giving specific feedback on what they can improve on (see Figure 2). Feedback can also take the burden off of researchers and educators to manually score habitats during our camp sessions, freeing time to work with students and discuss why they are interested in building certain parts of the habitat or how they can continue to improve. Future directions could include providing students with a checklist of features that they have completed for their base, designing a collaborative agent that students can offload building specific parts to, or integrating the system into schools as a stealth assessment for engineering concepts. We are optimistic that students using our pedagogical agent in future studies will find it as a meaningful tool for knowledge growth and developing necessary skills for their future interests.

Fig. 2. Example habitat feedback given by the agent.



7 Acknowledgments

The materials used in this study are based upon work supported by the National Science Foundation and Directorate for Education and Human Resources under Grants 1713609 and 1906873.

References

1. de Andrade, B., Poplin, A., Sousa de Sena, Í.: Minecraft as a tool for engaging children in urban planning: A case study in tirol town, brazil. *ISPRS International Journal of Geo-Information* **9**(3), 170 (2020)
2. Barzilai, S., Zohar, A.R., Mor-Hagani, S.: Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational psychology review* **30**, 973–999 (2018)
3. Burton, E.P.: Student work products as a teaching tool for nature of science pedagogical knowledge: A professional development project with in-service secondary science teachers. *Teaching and Teacher Education* **29**, 156–166 (2013)
4. Checa-Romero, M., Pascual Gómez, I.: Minecraft and machinima in action: Development of creativity in the classroom. *Technology, Pedagogy and Education* **27**(5), 625–637 (2018)
5. Harel, I.E., Papert, S.E.: *Constructionism*. Ablex Publishing (1991)
6. Hewett, K.J., Zeng, G., Pletcher, B.C.: The acquisition of 21st-century skills through video games: Minecraft design process models and their web of class roles. *Simulation & Gaming* **51**(3), 336–364 (2020)
7. Lane, H.C., Gadbury, M., Ginger, J., Yi, S., Comins, N., Henhapl, J., Rivera-Rogers, A.: Triggering stem interest with minecraft in a hybrid summer camp (2022)
8. Lane, H.C., Yi, S.: Playing with virtual blocks: Minecraft as a learning environment for practice and research. In: *Cognitive development in digital contexts*, pp. 145–166. Elsevier (2017)
9. Magnussen, R., Elming, A.L.: Cities at play: Children’s redesign of deprived neighbourhoods in minecraft. In: *European conference on games based learning*. p. 331. Academic Conferences International Limited (2015)
10. Nyhout, A., Ganea, P.A.: Scientific reasoning and counterfactual reasoning in development. *Advances in child development and behavior* **61**, 223–253 (2021)
11. Papavlasopoulou, S., Giannakos, M.N., Jaccheri, L.: Empirical studies on the maker movement, a promising approach to learning: A literature review. *Entertainment Computing* **18**, 57–78 (2017)
12. Sharma, K., Giannakos, M.: Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology* **51**(5), 1450–1484 (2020)
13. Spairani, E., Daniele, B., Signorini, M.G., Magenes, G.: A deep learning mixed-data type approach for the classification of fhr signals. *Frontiers in Bioengineering and Biotechnology* **10**, 887549 (2022)
14. Woodward, M.: Minecraft user statistics: How many people play minecraft in 2023? (2023), <https://www.searchlogistics.com/learn/statistics/minecraft-user-statistics/>