Not All Who Wander Are Lost: Trailblazing Trajectories in a Minecraft-Based Learning Environment

Xiner Liu[1[0009-0004-3796-2251]], Yiqiu Zhou[2[0009-0000-9004-9307]], Jaclyn Ocumpaugh[1[0000-0002-9667-8523]], Amanda Barany[1[0000-0003-2239-2271]], Andres Felipe Zambrano[1[0000-0003-0692-1209]], Zhanlan Wei[1[0009-0002-3931-6398]], Ryan S. Baker[1[0000-0002-3051-3232]], and Camille Giordano[1[0009-0001-2772-3079]]

[1] University of Pennsylvania, Philadelphia, PA 19104, USA
xiner@upenn.edu
[2] University of Illinois, Champaign, IL 61820, USA

Abstract. This study investigates how students explored a Minecraft-based science learning environment by analyzing their in-game movement trajectories. We use GPT-4o to identify recurring trajectory patterns from gameplay visualizations and to automatically label trajectory images, with some constructs labeled and reviewed by humans. These patterns are examined in relation to students' self-reported survey measures. Epistemic Network Analysis is used to compare how different movement behaviors co-occurred across different learner profiles. The findings showed that students with high or improving outcomes engaged in more flexible exploration. They often wandered, changed directions, and alternated between looking around the environment before closely examining specific objects. In contrast, students with low or declining outcomes tended to concentrate on specific areas and frequently backtracked to previously visited locations. These findings highlight the importance of how, not just how much, students explore in open-ended environments as part of the learning process.

Keywords: Player trajectories, Movement patterns, Spatial visualization, Trajectory visualization, Epistemic Network Analysis, Game-based learning

1 Introduction

Over the past two decades, digital games have become a prominent medium for science education to support curiosity [23], persistence [17], and self-directed forms of inquiry [15]. A commonly used format within this space is the open-ended or "sandbox" game, an environment that does not require players to complete predefined tasks or follow a linear path (e.g., [22]). These games are structured to support free exploration; players move through simulated worlds, observe dynamic environmental features, and make independent decisions about what to investigate and how to allocate their attention over time (e.g., [8, 5]). As a result, they are routinely adapted for use in formal and informal teaching settings as tools to support contextualized learning and engagement, particularly in science, technology, engineering, and mathematics (STEM) (e.g., [13]).

The flexible structure of open-ended games introduces distinct challenges for both the students who navigate these systems and the researchers who analyze them. For students, the absence of explicit goals or prescribed tasks requires them to make ongoing decisions about what to investigate and how to allocate their attention [7]. Without targeted support, however, some students may chart courses that limit their opportunities for STEM knowledge and interest development. This also introduces challenges for researchers; because students follow self-determined paths, their behaviors often lack shared structure or reference points, making it difficult to define patterns that can be meaningfully compared across learners [16]. Movement is continuous, context-dependent, and shaped by subtle shifts in attention that are not easily captured by standard behavioral metrics. While existing measures can approximate broad levels of engagement [6], they often miss key distinctions, such as whether students' actions are exploratory or repetitive, or how their behaviors evolve over time [14].

To address these challenges, methods that can characterize students' exploration in open-ended scientific environments in ways that are both interpretable and instructionally meaningful are needed. As part of this effort, we explore how image-based movement data might support qualitative interpretations of player behavior. In doing so, we also examine the reliability of GPT in interpreting image-based data, with the long-term goal of supporting real-time identification of player trajectory patterns during gameplay. In this study, we hope to better understand how students explore in open-ended scientific environments by examining the kinds of movement patterns they exhibit and how those patterns relate to learning and interest outcomes. We conduct our study within WHIMC, a Minecraft-based learning environment that allows player to investigate hypothetical conditions through free exploration. Our research questions are:

- RQ1: What kinds of movement patterns do students display when exploring an open-ended science learning environment like WHIMC?
- RQ2: Do students with different learning or interest profiles show different combinations of movement patterns during gameplay?

2 Related Work

A growing body of research has examined how students' self-directed navigation through open-ended educational games relates to learning, engagement, and persistence. [21] compared three versions of the *Crystal Island* game: high agency (free exploration), low agency (a guided path), and no agency (watching an expert's playthrough). Students in the low-agency condition showed the highest learning gains, even though they displayed more off-task behaviors. In a follow-up study, [20] recorded each student's actions and compared them to an expert solution path. Students who moved more slowly and followed a path closer to the expert's sequence showed higher learning gains and better performance.

This pattern holds across other game-based learning environments. For example, [18] compared each student's in-game actions to an expert model in *Impulse* and *Quantum Spectre*. Students with sequences that more closely matched the model scored higher on learning assessments. Students with disorganized action sequences showed

weaker understanding. [19] created network graphs of student quest sequences in *The Radix Endeavor.* Students with more focused or linear paths showed better self-regulation and needed less support from teachers. Similarly, [6] tracked the order in which students unlocked knowledge components in *Hist Maker*. Students who followed a logical learning path showed the most positive learning gain.

When students are not being guided towards a particular sequence of events in-game, however, the most focused or linear paths may not always be sufficient for improving learning. For example, in more recent work with *Crystal Island,* [12] found that students who learned the most did not take a known most efficient path through the system. However, this path was identified at a fairly large grain size (i.e., any movement from one point of interest to the next, regardless of the details of that path) and exploratory behaviors only correlated with learning when frustration levels were optimized.

Building on prior work, our study examines how students' in-game movement patterns relate to learning and interest outcomes in an open-ended science environment, WHIMC. We focus on movement patterns captured through trajectory data and use path visualizations to identify recurring navigation behaviors at a finer grain size than was previously possible in earlier research. These patterns are then compared across different learner profiles to understand how exploration strategies may reflect or relate to varying outcomes.

3 Method

This study was conducted in WHIMC (What-If Hypothetical Implementations in Minecraft), a Minecraft-based educational environment that supports STEM learning with exploratory simulations. WHIMC presents hypothetical planetary conditions to students who examine environmental variables, propose hypotheses, and compare findings across virtual settings. These activities are designed to support engagement with astronomy and related STEM disciplines, combining exploration with scientific reasoning.

The game is structured as an open-world environment, meaning students are not limited to a fixed sequence of tasks or movement. Although facilitators may recommend which world to explore, these suggestions do not restrict how students interact with the game. Within each world, students can move freely and choose how to explore based on their own interests. Each world includes features intended to prompt scientific investigation. For example, in *[Earth with] No Moon*, several regions include transformed terrain shaped by strong, persistent winds due to the absence of the Moon's gravitational pull. These points of interest act as entry points for inquiry, but students choose what to investigate, how to collect data (using in-game measurement tools), and which comparisons to make across worlds.

3.1 Data Collection

This study draws on data from 32 middle school students recruited from summer camps that took place at two U.S. schools. 19 identified as male, 11 as female, and 2 as a third gender or undisclosed. The sample included 10 White, 14 African American, 2 Native

American/Pacific Islander, and 3 students of other backgrounds. Participation was voluntary, with written assent and consent obtained from students and parents before the camp started. Data collection took place over five days. In Days 1 to 3, students explored Minecraft environments modeled on space-related settings that introduced key astronomy concepts (e.g., wandering on a low-gravity planet surface or on the terrain of a terraformed Mars). They also completed a set of validated measures designed to assess their interests and knowledge development (see Table 1). A total of 22 students completed all scheduled surveys, while others were excluded for logistical reasons.

Table 1. Summary of surveys and data collection schedule

Day	Survey Purpose
1&4	**Science Interest Development** [3] Measures on a 0-7 scale
1&5	**Astronomy Knowledge** Assesses learning gains in astronomy on a 0-20 scale
3	**Minecraft Interest** Measures interest in Minecraft on a 0-5 scale
3	**Astronomy Interest** [11] Measures on a 0-5 scale
4	**Situational Interest** [9] Assesses context-based interest, 0-7
4	**Self-Efficacy** [4] Measures students' confidence in astronomy concepts, 0-6

3.2 Visualizing Student Exploration in Virtual Worlds

To examine how students navigated the virtual worlds in WHIMC, we used the WHIMC Path Generator, a visualization and analysis tool with an accessible API developed as part of the WHIMC project (https://github.com/whimc/Path-Generator). This tool draws on gameplay log data to generate visualizations of each student's movement and interaction history within the game.

These visualizations originated from three Minecraft-based tracking systems: (1) a position tracker that recorded each student's location every three seconds, (2) an observation displayer that captured when and where students made scientific observations, (3) and a block interaction monitor that logged when students placed, removed, or modified blocks in the virtual world. The Path Generator created images by overlaying student movement and observation data onto background maps of each world. These images showed where each student traveled, which areas they visited most, and where they paused to make observations. Each image represents a specific time segment, and within each segment, the player's starting location is marked with a green dot and the ending location with a red dot. If the player teleports during that interval, both the departure and arrival points appear as red dots. The Path Generator also calculated an exploration metric for each student. This metric divided each world into a 10×10 grid and identified how many unique tiles a student had entered or made an observation in. The final output included one image per world per student and a set of exploration scores that quantified the extent of each student's in-game exploration.

We used the Path Generator to produce player path visualizations from three WHIMC worlds, *Lunar Crater*, *No Moon*, and *Mynoa Far*, which students explored in the first three days of the summer camp. Each visualization captured the player's trajectory during a distinct, non-overlapping time segment in which the student remained in a single virtual world. Two sets of visualizations were generated, one using longer

(10 min. intervals) to provide a broader view of exploration behaviors and were used to identify recurring movement patterns. The observed patterns served as the basis for developing a codebook to define and classify common behavioral trajectories (see Section 3.4). The second set was generated using shorter segments (5-minute intervals), which provided a more fine-grained view of student activity while still preserving enough temporal context for patterns to remain identifiable. This choice was also informed by our experience facilitating the camp, where we found that meaningful exploration strategies often unfolded within these time frames. We applied the codebook to this second set to systematically label and compare behaviors across students.

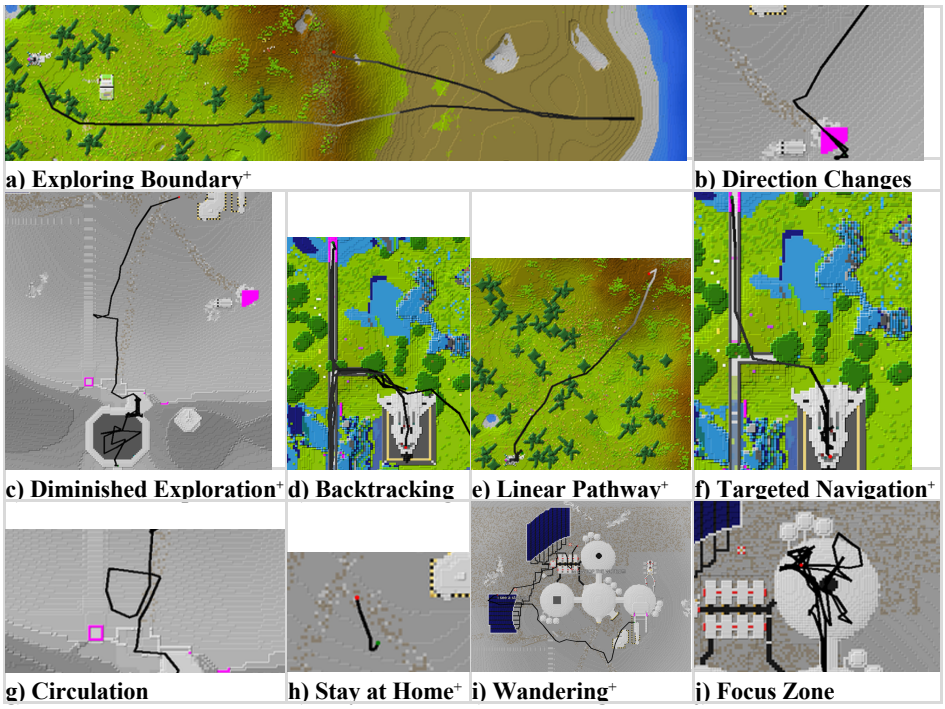
We retained 16 students with (1) at least six visualizations from distinct five-minute intervals (i.e., $\geqslant$ 30 minutes of gameplay) and (2) recorded activity in all three virtual worlds. Students with less data were excluded, as shorter gameplay duration did not provide enough observable movement to support reliable analysis of player trajectories.

3.3 Coding Player Trajectories

Following previous work that has used generative AI to discover categories in textual data with human oversight [1], we adopted an iterative approach to the early phases of using GPT. We used its strength in spotting patterns across large-scale data that may be hard for humans to detect, while closely comparing its categorizations with our own understanding. This process yielded ten constructs, several of which could be automated in a manner that could allow us to apply them at scale in future research.

Use of GPT for generating initial player descriptions. We prepared the path visualizations in our data in a manner that could be processed by GPT-4o's API. Because GPT-4o's ability to handle visual data for this purpose is relatively unstudied, this process began with the first and second authors manually reviewing the data to identify preliminary themes. To remain within the current token limit for GPT-4o's API, we submitted three non-overlapping batches of 30 images (10 from each world), which constituted 5% of the total available trajectory images. We set the temperature parameter to 0 and tested the prompt across multiple iterations.

Human oversight of selecting player trajectory codes. We then consolidated the codes that had emerged across at least two of these batches, reviewing each definition and merging those that were conceptually similar. These resulting six codes served as our initial descriptions for coding player trajectories, but other codes emerged later. For example, *Exploring Boundary* did not appear among the codes initially suggested by GPT, but was consistently observed in the data during human review. Three additional trajectory codes (*Direction Changes*, *Wandering*, and *Backtracking*) were later identified during the initial round of coding by the human coder. Constructs in the codebook were not mutually exclusive; each image could be labeled with multiple patterns. For example, the image of *Exploring Boundary* in Figure 1 is also labeled as a *Linear Pathway* because the player's movement remains consistently linear throughout the 5-minute interval.



+ Denotes the example of this construct shows the complete player trajectory

Fig. 1. Representative images for each construct

Table 2. Code definition, generation process (Gen), and IRR performance (κ) between Human/Human (H/H) and Human/GPT (H/G); codes are not mutually exclusive

Construct Name	Gen	H/H	H/G
Exploring Boundary: Movement along the outer boundaries of the map.	H	**0.85**	**0.68**
Direction Changes: Sudden, sharp shifts in a player's movement trajectory.	H	**0.90**	0.49
Diminished Exploration: Noticeable decline in movement/interaction over time.	H	**0.83**	0.11
Backtracking: Repetition of previous paths or returns to earlier locations.	H	**0.92**	0.53
Linear Pathway: Predominantly straight-line movement covering approximately 1/3 or at least 100 blocks of the total path.	G	**0.87**	**0.84**
Targeted Navigation: Purposefully directing path toward specific point of interest.	G	**0.88**	0.35
Circulation: Movement that forms a distinct circular path on the map.	G	**0.76**	**0.87**
Stay at Home: Minimal movement near the spawn point.	G	**0.78**	**0.92**
Wandering: Movement without a clear focus, often involving the player roaming the map without extensive engagement with points of interest.	G	**0.88**	**0.70**
Focus Zone: Concentrated exploration around the points of interest indicated by densely clustered paths.	G	**0.93**	**0.71**

Interrater reliability (IRR). We implemented partial automation of coding with collaboration between humans and GPT. This was done in three stages. In stage one, re-

searchers established human reliability (Cohen's $\kappa>0.75$) for GPT-generated codes using 50 player trajectories from the *Lunar Crater* world and for human-generated codes using 50 player trajectories from the *No Moon* world.

In stage two, GPT-4o's ability to match human coding was tested on a third validation set (10 images from a third WHIMC world, *Mynoa Far*, and 20 randomly selected from each of the other two worlds). Following [10]'s procedure, we used a few-shot prompting setup with GPT-4o to label trajectory patterns, which were also coded by human researchers. For each construct, GPT-4o was given (1) a definition, (2) a representative example, (3) a description of the study context, (4) guidance on how to interpret the visualizations, and (5) a reference map of the same world without an overlaid path, to help GPT-4o distinguish the player trajectory from other map elements.

Finally, in stage three, the researchers inspected the GPT's reliability against human coders. These researchers manually coded the remaining data set for four constructs (*Backtracking, Direction Changes, Targeted Navigation,* and *Diminished Exploration)* that GPT could not reliably code ($\kappa<0.7$ with humans; values near this threshold, e.g., $\kappa=0.68$, were considered acceptable). Further inspection revealed that GPT failed to accurately interpret the temporal and directional information related to the movement in these four trajectories (see Appendix A). Coding for the other six constructs was automated using the same prompt tested in stage two.

Final Oversight of Segmentation Issues. Finally, we reviewed the coding of two constructs: *Backtracking*, which involves player returning to previously visited locations, and *Diminished Exploration*, which refers to a reduction in movement over time. (Note: GPT did not achieve reliable agreement with human coders on these two constructs.) Specifically, the human oversight process identified a mislabeling risk that could have been introduced by our segmentation process, where the fixed 5-minute visualization intervals selected to satisfy GPT's coding capabilities may have led to coding based on cutoff points that were arbitrary with respect to the students' gameplay patterns rather than on meaningful movement patterns. To address this, we examined any adjacent paths in the student's visualization sequence. For example, some segments might have been initially labeled as *Diminished Exploration* because student activity appeared inactive toward the end, but that inactivity could have also marked the transition into a new, focused exploration in the following image. In these cases, we relabeled the trajectory to better represent the student's path across both segments. This human review process for these two constructs was conducted for all visualizations to strengthen the validity of trajectory interpretations.

3.4 Correlation between Trajectory Pattern and Survey Measures

We then examined how player trajectories relate to external measures of learning and interest. To do so, we calculated the percentage of occurrences for each construct at the student level to account for variation in the number of visualizations per student. As the data was not normally distributed, we applied a Spearman's correlation between these percentages and our external measures. In two cases, pre- and post-measures of the

same scale (i.e., Boeder et al.'s science interest development scale [3] and an internally developed astronomy knowledge measure) allowed us to correlate with gains.

3.5 Epistemic Network Analysis

To examine whether students who differ in interest levels or learning outcomes demonstrated distinct patterns in how they explored the virtual worlds, we applied Epistemic Network Analysis (ENA) to the coded visualization segments. Because the codes were not mutually exclusive, ENA often modeled the structure of behavior by identifying which movement patterns tend to co-occur within lines of data. We generated six ENA visualizations, one for each survey measure, applying difference models that allowed us to identify patterns that appeared more strongly across student groups. For the two measures that captured change over time (i.e., science interest and astronomy knowledge), students who showed positive gains between the pre- and post-tests were compared to those who did not. For the remaining four measures (i.e., Minecraft interest, astronomy interest, situational interest, and self-efficacy), students were categorized into high and low groups using a median split on the corresponding scores. Details about each group are provided in the result section.

Data was segmented by player so that movement patterns were modeled within individual students in the same WHIMC world. Visualizations were divided based on temporal continuity: any two trajectory images separated by more than 10 minutes were treated as different conversations. This threshold allowed for brief disconnections (e.g., temporary logouts or brief moments when the player was not in the same world) without breaking a continuous gameplay trajectory. However, when the time interval between trajectory images exceeded 10 minutes, the visualizations were treated as separate conversations to avoid artificial accumulation of codes across unrelated gameplay sessions (such as when students paused play for in-class activities). In other words, we limited co-occurrence calculations to temporally consistent periods and avoided linking behaviors across disconnected intervals. We used an infinite stanza window in constructing the epistemic networks, which created connections between the codes in each image and all codes that appeared in earlier images within the same segment. This allowed us to analyze how patterns of behavior accumulate over the course of each uninterrupted gameplay session. We excluded connections with line weights below 0.04 ($lw<0.04$) to improve clarity in the visualizations.

4 Results

4.1 Correlation Results for Trajectory Patterns and Survey Data

Table 3 presents correlations between the ten player trajectory patterns and six external survey measures. Three trajectories showed consistently positive/negative correlations across all six surveys: *Wandering* (r=0.33 to 0.68), *Focus* (r=-0.22 to -0.68) and *Targeted Navigation* (r=-0.18 to -0.07). Out of 60 tests (i.e., 10 constructs × 6 measures), we observed 10 significant results. None of these correlations remained significant after applying a Benjamini and Hochberg correction [2]. However, a Monte Carlo analysis

with 10,000 simulations (at significance level of 0.05) shows that 10 or more significant results would occur by chance only 0.06% of the time if all tests were random. This suggests that the overall trends in our results are unlikely to be due to chance.

Table 3. Spearman correlations between construct frequencies and interest measures

	Int. Dev. Gain	Knowledge Gain	Minecraft Int.	Astronomy Int.	Situational Int.	Self-efficacy
Diminished Exploration	-0.65	0.21	-0.35	-0.25	-0.49	-0.4
[+]Focus Zone	-0.45	-0.28	-0.55	-0.22	-0.49	-0.68
Linear Pathway	-0.34	0.56	-0.01	-0.26	-0.06	0.13
[+]Targeted Navigation	-0.15	-0.04	-0.08	-0.07	-0.09	-0.18
Exploring Boundary	-0.02	0.56	-0.16	-0.24	-0.03	0.23
Backtracking	0.03	0.08	-0.25	-0.13	-0.38	-0.12
Circulation	0.22	0.38	-0.03	0.19	0.14	0.16
Direction Changes	0.26	0.45	0.01	-0.08	0.04	0.33
[+]Wandering	0.33	0.34	0.55	0.33	0.68	0.59
Stay at Home	0.39	-0.63	0.09	0.13	0.12	-0.12

[+] Denotes a pattern that shows consistent positive or negative correlations across all measures

4.2 Epistemic Network Visualization of Player Trajectories

We next present Epistemic network comparing the movement patterns of students with different interests and learning profiles. We start with the analyses of the two surveys that were administered twice, comparing high and low gains across both measures (science interest development and astronomy knowledge). We then examine how students with high and low scores on the remaining four surveys differ from one another. For each survey, the number of students in the high/low group varied slightly, as did the average number of images (per student) for each survey (see Table 4).

Table 4. Distribution of students in six difference models

	Pos. Gains or High Scores			Neg./No Gains or Low Scores		
	N	Avg Score	Avg Images	N	Ag Score	Avg Images
Science Interest Dev. - GAINS	5	0.73	13	11	-0.78	15
Astronomy Knowledge - GAINS	7	2.34	16	9	-1.6	13
Minecraft Interest - HI/LO	8	3.1	13	8	1.7	17
Astronomy Interest -HI/LO	7	3.6	11	9	1.7	16
Situational Interest - HI/LO	8	2.8	13	8	1.8	16
Self-Efficacy - HI/LO	8	5.7	13	8	4.1	16

Differences in Movement Patterns Based on Knowledge and Interest Gains. Figure 2 shows the ENA difference models generated using the astronomy knowledge gains (left) and the science interest development gains (right). In both models, line weight differences (lw_{Diff}) that are greater for students with positive gains (shown in red) are compared to all other students (shown in blue). For both measures, Mann-Whitney tests

reveal significant differences between the groups along the X-axis of each network (astronomy knowledge: U=51, p=0.01, r=-0.85; science interest: U=59, p<0.01, r=-0.87).

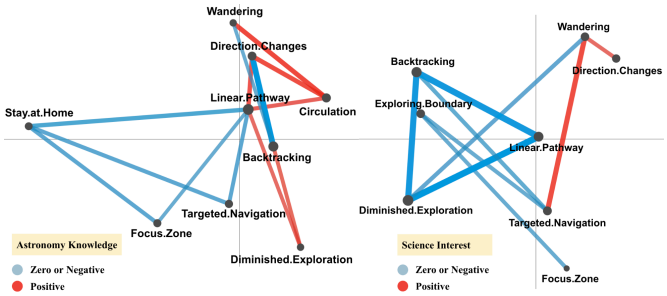


Fig. 2. ENA difference models for astronomy knowledge (left) and science interest development (right), comparing players with positive gains (red) to those with zero or negative gains (blue).

Results related to astronomy knowledge gains (See Fig. 2, left) show that students with low gains have 3 stronger connections to *Stay at Home*: *Linear Pathway*, *Targeted Navigation*, and *Focus Zone* (line weight difference lw_{Diff}=0.05 for each). These findings match the correlation results, where *Stay at Home* was negatively correlated with astronomy knowledge gain (r=-0.63). Although this group included only five players, they accounted for 75% of the trajectory images coded with *Stay at Home*, which suggests limited within game movement is associated lower learning outcomes.

Interestingly, this group also showed stronger connections with *Linear Pathway: Focus Zone* and *Targeted Navigation* (lw_{Diff}=0.04 for each). Inspection of the trajectory images suggests this was a common navigation strategy among these players. For instance, player Rabbit (pseudonym) often moved in a straight line toward a specific object or area of interest and explored that area intensively. Seven of Rabbit's ten trajectory images are coded as *Linear Pathway*, and all of these were also coded as either *Focus Zone* or *Targeted Navigation*. As *Focus Zone* (r=–0.28) and *Targeted Navigation* (r=–0.04) are negatively associated with astronomy knowledge gains, these findings suggest that exploration strategies that are too narrowly focused or overly fixed on a single path may reduce the chance to discover diverse in-game phenomena and ultimately limit knowledge development.

Students with high gains in astronomy knowledge have stronger connections with codes that are less linear, such as *Direction Changes,* a result which can also be seen in the findings for high science interest gains (See Fig. 2, right). Players with positive interest gains had much stronger connections between *Wandering* and *Direction Changes* (lw_{Diff}=0.05). These two constructs were also positively correlated with interest gain (r=0.33 and r=0.26, respectively). A review of trajectories suggests that this pattern often reflects a player moving around a point of interest without immediately entering or closely interacting with it. After looking around the surrounding area, the player would then select a specific location (that might be of interest or worth exploring) and start a more focused exploration. This may reflect a stronger sense of control over the pacing and flow of their exploration, or a more deliberate engagement with the environment, compared to players in the other group.

In contrast, the group with no or negative interest gains showed stronger connections involving *Linear Pathway* and two other constructs: *Backtracking* and *Diminished Explorations* (lw_{Diff}=0.08 for each). While both *Linear Pathway* and *Wandering* may appear as less directed forms of movement, *Wandering* seems to involve open-ended, flexible navigation that responds to features in the environment. *Linear Pathway*, in this context, reflects a more mechanical path with little adjustment based on new information or surroundings. These observations suggest that a more exploratory approach like *Wandering* may better support the development of science interest within the game.

Differences in Movement Patterns Based on Interest Levels. Figure 3 shows ENA difference models comparing the movement patterns of players with high and low scores (split at the median) on the four remaining surveys. As with previous measures, significant differences were found for each of these models (Minecraft: U=63, p<0.01, r=–0.97; astronomy interest: U=7, p=0.01, r=0.78; situational: U=0, p=0, r=1; self-efficacy: U=2, p<0.01, r=0.94).

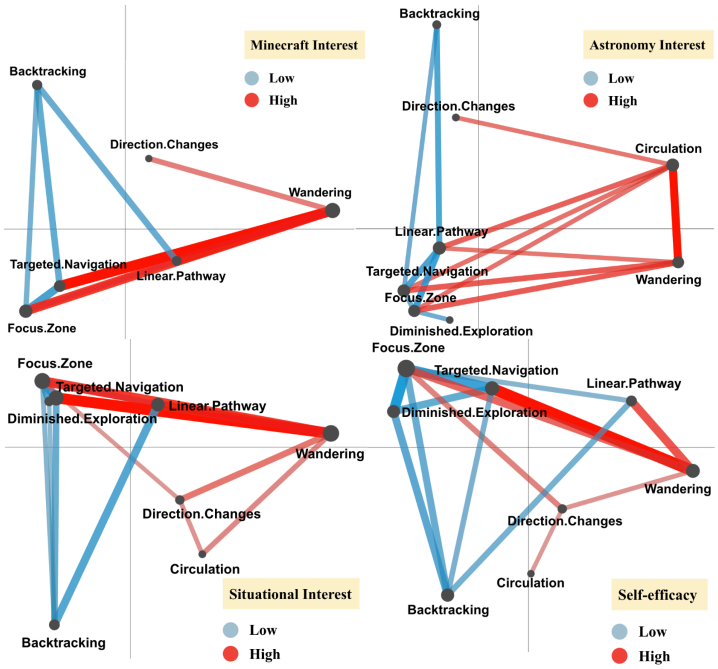


Fig. 3. Difference models for Minecraft interest (top left), astronomy interest (top right), situational interest (top left), and self-efficacy (top right), comparing high (red) and low scores (blue).

These difference plots show patterns similar to those for high/low science interest development gains and high/low astronomy learning gains. In particular, as Table 5 shows, these four plots show that students with high-interest (of any kind) and high-self-efficacy have strong connections involving *Wandering*, while those with lower scores show stronger connections to *Backtracking*.

Table 5. Line weights for prominent connections shown in the four plots; boldface highlights the group with the largest value

	Minecraft Int.		Astronomy Int.		Situational Int.		Self-efficacy	
	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo
Wandering ↔ Targeted Navigation	**0.29**	0.17	**0.26**	0.20	**0.28**	0.17	**0.28**	0.17
Wandering ↔ Focus Zone	**0.29**	0.15	**0.23**	0.18	**0.25**	0.15	**0.24**	0.17
Wandering ↔ Direction Changes	**0.18**	0.12	**0.15**	0.14	**0.18**	0.12	**0.17**	0.13
Circulation ↔ Wandering	**0.11**	0.09	**0.14**	0.07	**0.13**	0.07	0.10	0.10
Focus Zone ↔ Targeted Navigation	0.25	**0.34**	0.27	**0.31**	0.25	**0.34**	0.25	**0.34**
Backtracking ↔ Linear Pathway	0.16	**0.23**	0.16	**0.21**	0.15	**0.23**	0.16	**0.22**
Backtracking ↔ Targeted Navigation	0.15	**0.22**	0.16	**0.20**	0.15	**0.22**	0.16	**0.21**
Backtracking ↔ Focus Zone	0.15	**0.21**	0.17	**0.19**	0.15	**0.20**	0.15	**0.21**

As in the patterns seen in the astronomy knowledge and science interest development gain models, Table 5 shows that *Wandering* often co-occurred with more linear behaviors (i.e., *Targeted Navigation*, *Focus Zone*, and *Direction Changes*) among players with high interest or self-efficacy. In other words, these connections may reflect a flexible exploration strategy—one where players moved through the environment while actively identifying areas for more sustained interaction. In contrast, students with low interest or self-efficacy measures show stronger connections where both codes reflect more constrained trajectories (e.g., those that would reflect less exploration of the WHIMC worlds). These included connections between *Focus Zone* and *Targeted Navigation* as well as *Backtracking* and several other codes that reflect monotonous trajectories *(i.e., Linear Pathway, Targeted Navigation,* and *Focus* Zone).

It is important, however, to distinguish *Backtracking* from *Circulation*, which showed stronger connections in the high-interest or self-efficacy group. *Circulation* reflects quick, deliberate movement around the perimeter of an area, often as a way to assess the surroundings before deciding where to explore more carefully. In contrast, *Backtracking* in the low-interest group showed little evidence of re-orientation and may more likely reflect hesitation, confusion, or disengagement.

Resultant examinations of these trajectories also show that high-interest or self-efficacy players often alternated between major points of interest that were likely discussed by camp facilitators and secondary areas on the map, such as a pond, a crater, or a hill, that they had discovered through their own exploration (i.e., *Wandering* behaviors).

For example, the trajectory of player Tiger (pseudonym), who appeared in all three high-interest and high self-efficacy groups, included 35 visualization segments, with a relatively even distribution across *Wandering*, *Targeted Exploration*, *Focus Zone*, and *Direction Changes*, each appearing 14 to 19 times. Tiger's movement pattern showed repeated transitions between various major points of interest and less obvious areas like a hilltop or a cluster of rocks that were not part of the main facilitation prompts. Rather than returning along the same path, Tiger often explored alternate routes and paused in new locations along the way. This pattern suggests an investigative strategy of discovery, where structured movement between known targets was combined with exploratory detours to self-identified locations that expanded their interaction with the environment.

On the other hand, the low-interest or self-efficacy groups' direct paths seem to reflect movement between locations mentioned in instructions or emphasized during facilitation. Upon inspecting the trajectory images, these players often entered a point of interest directly, interacted with it (sometimes quite extensively), then returned to their previous location, repeating this across multiple areas. These patterns may suggest a gameplay strategy driven by compliance to authority rather than one driven by personal curiosity.

For instance, the trajectory of one player, Cat (pseudonym), who appeared in three low-interest groups and also in the low self-efficacy group, included 18 visualization segments, 11 of which were coded as *Backtracking*, meaning they frequently returned to previously visited locations within even the short (5-min.) intervals. Closer inspection of Cat's overall trajectory shows them shifting between a small number of interest zones—leaving one for another but then returning in a fairly linear path between these points of interest with minimal exploration of surrounding areas. This repetition was common among other players in the low interest group, observed a similar trend: movement centered around revisiting certain areas rather than branching out.

5 Discussion and Conclusions

This study examines how students' movement trajectories unfold within a Minecraft-based learning environment. We used GPT-4o to help identify a set of recurring trajectory patterns and then applied a collaborative coding approach, where both GPT-4o and human annotators labeled the image data. We then correlate these patterns to students' interest, self-efficacy, and knowledge survey measures, and use to compare how these patterns co-occurred across different learner profiles.

The study found that students who show positive gains in science interest or astronomy knowledge, or who report higher astronomy or Minecraft interest and self-efficacy, tend to explore the virtual environment in more flexible and autonomous fashion. Although their movement may appear to drift without a clear destination, their frequent shifts in direction may actually reflect an exploratory strategy where they alternate between looking around and investigating. That is, what appears haphazard may instead reflect the interest and engagement of a student who is receptive to environmental cues and adjusts their actions accordingly. In contrast, students with no gains or lower reported interest and self-efficacy tend to follow straight, fixed paths toward a point of interest. These students often move mechanically from one known object to another, repeat visits to the same areas, and otherwise limit expansion beyond assigned zones, which might reflect confusion, low confidence, or a compliance-driven mindset.

This difference points to one implication: neither science learning outcomes nor higher levels of interest or self-efficacy were linked to the overall amount of concentrated on-task movement. Behaviors that appear unfocused, such as *Wandering*, were often actually part of a more adaptive and self-directed exploration process. More linear or fixed paths sometimes may reflect difficulty exploring or uncertainty about how to move forward, which limits opportunities to notice new features or respond flexibly to the environment.

The findings suggest that interventions in open-ended learning environments should not simply guide students toward specific locations or tasks but instead take full advantage of the benefits of the medium. Educators and practitioners could support student awareness of when and how to shift from exploratory to goal-directed behaviors. For example, prompts or design features might help students reflect on what they have and have not yet explored, or scaffold decisions about when to pursue a new area. Simply increasing movement or encouraging coverage is not enough; what matters is helping students develop productive navigation strategies that combine curiosity with purposeful inquiry.

That said, the strength of such design implications must be interpreted in relation to the scope of the trajectory data. A key limitation of this work lies in its reliance on movement patterns as proxies for cognitive engagement and learning processes. While path trajectories reveal how students navigate the virtual environment, they do not capture students' intentions, reasoning, or interpretation of in-game content. Two students may follow similar paths but differ significantly in what they notice, how they make sense of it, or whether they connect it to scientific ideas. As a result, movement data alone can miss important differences in how learning unfolds.

Future research should triangulate with additional data sources (e.g., in-game observations, post-activity interviews), to better capture students' intentions, interpretations, and reasoning processes during exploration. This would allow researchers to examine not just where students go, but why they move the way they do and how they make sense of what they encounter. Improving temporal modeling is also important; rather than relying on fixed time windows, future work could explore event-based or adaptive segmentation methods that better reflect how students shift between exploration phases. Expanding the sample size and including students from a broader range of backgrounds and learning contexts would also strengthen the generalizability of the findings. Addressing these gaps will allow for more accurate interpretations of player behavior and more effective design of responsive, in-game supports. This work extends the application of QE approaches to image-based data and offers a novel use case for leveraging large language models to automate qualitative coding, broadening the methodological toolkit for analyzing visual traces of learner behavior. Our hope is that applications of visual trace data may one day serve as a complement to discourse-based QE analyses by situating player actions within the spatial and temporal flow of game-based learning.

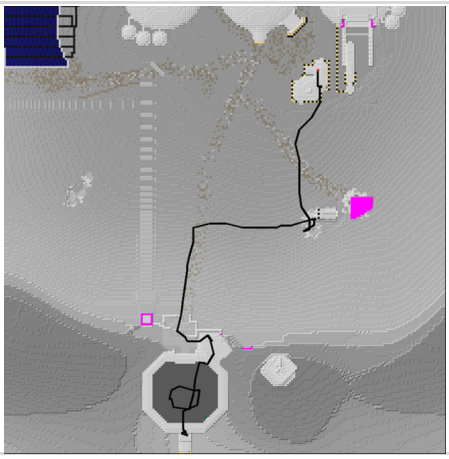
Acknowledgments. This research was supported by National Science Foundation Grants # DRL2301173.

References

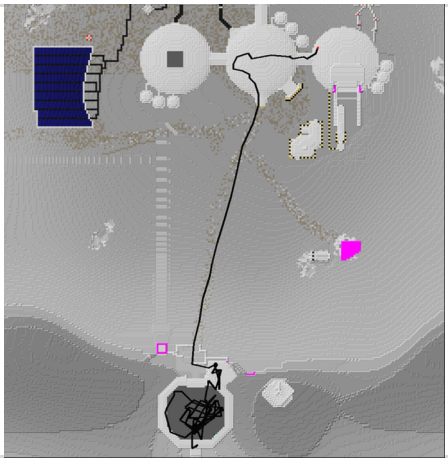
1. Barany, A., Nasiar, N., ... & Baker, R. S.: ChatGPT for education research: Exploring the potential of large language models for qualitative codebook development. Conf. Art. Intell. in Ed., 134–149 (2024)
2. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy. Stat. Soc.: Series B (Method.), 57(1), 289–300 (1995)

3. Boeder, J. D., Postlewaite, E. L., Renninger, K. A., Hidi, S. E.: Construction and validation of the interest development scale. Motivation Sci., 7(1), 68–82 (2021)
4. Britner, S. L., Pajares, F.: Sources of science self-efficacy beliefs of middle school students. J. of Research in Sci. Teaching, 43(5), 485–499 (2006)
5. Chen, C. H., Law, V.: Scaffolding individual and collaborative game-based learning in learning performance and intrinsic motivation. Comp. in Human Behavior, 55, 1201–1212 (2016)
6. Feng, X., Yamada, M.: An analytical approach for detecting and explaining the learning path patterns of an informal learning game. Ed. Technology & Soc., 24(1), 176–190 (2021)
7. Kao, G., Chiang, C, Sun, C: Customizing scaffolds for game-based learning in physics: impacts on knowledge acquisition & game design creativity. Comp. Ed., 113, 294–312 (2017)
8. Ketelhut, D. J.: The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. J. of Sci. Ed. and Technology, 16, 99–111 (2007)
9. Linnenbrink-Garcia, L., Durik, A., Conley, Barron, K., Tauer, J., Karabenick, Harackiewicz, J.: Measuring situational interest in ac. domains. Ed. Psych. Meas., 70(4), 647–671 (2010)
10. Liu, X., Zhang, J., Barany, A., Pankiewicz, M., Baker, R..: Assessing the potential and limits of large language models in qualitative coding. Intl. Conf. on Quant. Eth., 89–103 (2024)
11. LoPresto, M., Murrell, S.: Astronomical Misconceptions Survey (AMS). PhysPort (2022)
12. Nasiar, N., Zambrano, A., Ocumpaugh, J., Hutt, S., Goslen, A., Rowe, J., ... & Mott, B.: It's good to explore: Investigating silver pathways and the role of frustration during game-based learning. Conf. Art. Intell. in Ed., 497–503 (2023)
13. Nebel, S., Schneider, S., Rey, G. D.: Mining learning and crafting scientific experiments: A literature review on the use of Minecraft in education and research. J. of Ed. Technology & Soc., 19(2), 355–366 (2016)
14. Nye, B. D., Core, M. G., Ghosal, A., Walker, P. B.: Metrics for engagement in games and simulations for learning. Using Cognitive and Affective Metrics in Ed. Simulations and Games, 107–130 (2021)
15. Pacheco-Velazquez, E., Rodés Paragarino, V., Glasserman, L. D., Carlos Arroyo, M.: Playing to learn: Developing self-directed learning skills through serious games. J. of Intl. Ed. in Business, 17(3), 416–430 (2024)
16. Podolefsky, N., Moore, E., Perkins, K.: Implicit scaffolding in interactive simulations: Design strategies to support multiple educational goals. arXiv preprint:1306.6544 (2013)
17. Rahimi, S., Shute, V., Zhang, Q.: The effects of game & student charact. on persistence in ed. games: A hierarch. linear modeling approach. Intl. J. Tech. Ed. Sci., 5(2), 141–165 (2021)
18. Rowe, E., Asbell-Clarke, J., ... & Edwards, T.: Assessing implicit science learning in digital games. Comp. in Human Beh., 76, 617–630 (2017)
19. Ruipérez-Valiente, J. A., Rosenheck, L., Kim, Y. J.: What does exploration look like? Painting a picture of learning pathways using learning analytics. Game-Based Assessment Revisited, 281–300 (2019)
20. Sawyer, R., Rowe, J, Azevedo, R, Lester, J: Filtered time series analyses of student problem-solving behaviors in game-based learning. Intl. Conf. Ed. Data Mining Soc., 229–238 (2018)
21. Sawyer, R., Smith, A., Rowe, J., Azevedo, R., Lester, J.: Is more agency better? The impact of student agency on game-based learning. Intl. Conf. Art. Intell. in Ed., 335–346 (2017)
22. Spring, F., Pellegrino, J. W.: The challenge of assessing learning in open games: HORTUS as a case study. Proc. GLS, 8(8), 200–208 (2011)
23. Tu, H.Y., Lee, S.W.: Curiosity, interest, and engagement: Unpacking their roles in students' learning within a virtual game environment. J. Ed. Comp. Res., 62(8), 1995–2019 (2025)

Appendix A. Example images for the four constructs in which GPT achieved insufficient agreement with human coders, each showing a case coded as present by humans but not identified by GPT.



1) Direction Changes[+]



2) Diminished Exploration[+]



3) Backtracking[+]



4) Targeted Navigation[+]

[+] Denotes the example of this construct shows the complete player trajectory